

Contemporary Mathematicians

Gian-Carlo Rota

Editor



Emil L. Post
June 5, 1924



Emil Post with daughter Phyllis (*left*)
and wife Gertrude (*right*) circa 1953

Solvability, Provability, Definability:

The Collected Works of Emil L. Post

Martin Davis

Editor

Birkhäuser
Boston • Basel • Berlin
1994

Martin Davis
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012-1185

Library of Congress Cataloging In-Publication Data

Post, Emil Leon, 1897-1954.

[Works, 1993]

Solvability, provability, definability : the collected works of
Emil L. Post / Martin Davis, editor.

p. cm. -- (contemporary mathematicians)

Includes bibliographical references.

ISBN 0-8176-3579-3 (Boston : alk. paper) -- ISBN 307643-3579-3
(Basel : alk paper)

1. Post, Emil Leon, 1897-1954. 2. Logic, Symbolic and
mathematical. I. Davis, Martin, 1928-

QA3.P78 1993

93-9347

510--dc20

CIP

Printed on acid-free paper
© Birkhäuser Boston 1994

Birkhäuser ®

Copyright is not claimed for works of U.S. Government employees.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the copyright owner.

Permission to photocopy for internal or personal use of specific clients is granted by Birkhäuser Boston for libraries and other users registered with the Copyright Clearance Center (CCC), provided that the base fee of \$6.00 per copy, plus \$0.20 per page is paid directly to CCC, 21 Congress Street, Salem, MA 01970, U.S.A. Special requests should be addressed directly to Birkhäuser Boston, 675 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.

ISBN 0-8176-3579-3

ISBN 3-7643-3579-3

Typeset and camera-ready copy provided by the editor.

Printed and bound by Quinn-Woodbine, Woodbine, NJ.

Printed in the U.S.A.

9 8 7 6 5 4 3 2 1

Contents

Preface	vii
Bibliography of Emil L. Post	viii
Introduction	xi

[1] The generalized gamma functions	1
[2] Discussion of problem 433	17
[5] Introduction to a general theory of elementary propositions	21
[14] Generalized differentiation	44
[17] Finite combinatory processes — Formulation I	103
[18] Polyadic groups	106
[19] The two-valued iterative systems of mathematical logic	249
[20] Absolutely unsolvable problems and relatively undecidable propositions — account of an anticipation	375
[21] Formal reductions of the general combinatorial decision problem	442
[22] Recursively enumerable sets of positive integers and their decision problems	461
[23] A variant of recursively unsolvable problem	495
[24] Note on a conjecture of Skolem	501
[26] Recursive unsolvability of a problem of Thue	503
[34] (with S.C. Kleene) The upper semi-lattice of degrees of recursive unsolvability	514

Abstracts

[3] Introduction to a general theory of elementary propositions	545
[4] Determination of all closed systems of truth tables	545
[6] On a simple class of deductive systems	545
[7] Visual intuition in Lobachevsky space	546
[8] Visual intuition in spherical and elliptic space: Einstein's finite universe	546
[9] A non-Weierstrassian method of analytic prolongation	546
[10] A new method for generalizing e^x in the complex domain	547
[11] A simple geometric proof of the equality of the Brochardt angles of a triangle	547
[12] Theory of generalized differentiation	547
[13] The m th derivative of a function of a function; calculus of m th derivatives	547
[15] Polyadic groups (preliminary report)	548
[16] Finite combinatory processes. Formulation I	548
[25] Recursive unsolvability of a problem of Thue	549
[28] Degrees of recursive unsolvability (preliminary report)	549

[29] (with Samuel Linial) Recursive unsolvability of the deducibility, Tarski's completeness, and independence of axioms problems of propositional calculus	550
[30] Note on a relation recursion calculus	550
[31] Solvability, definability, provability, history of an error	551
[32] A necessary condition for definability for transfinite von Neumann-Gödel set theory sets, with an application to the problem of the existence of a definable well-ordering of the continuum (preliminary report)	551
[33] (with S.C. Kleene) The upper semi-lattice of degrees of recursive unsolvability	552
Permissions	553

Preface

The almost forty years that have passed since the death of Emil L. Post have seen an ever increasing awareness of the importance of his contributions. Although it is sad to think that so much time had to elapse before the publication of his *Collected Works*, I am extremely happy to see it finally taking place while I can still participate in the process.

Post was my teacher when I was an undergraduate student in mathematics at City College in New York during the late 1940s. Although his was a powerful influence on many of his students who went on to successful careers in the mathematical sciences, I am the only one whose own research interests have followed in the directions he pioneered, and so I feel it particularly appropriate that I play the role of his editor.

This volume contains all of Post's published writings as well as his fascinating article, unpublished during his lifetime, *Absolutely Unsolvable Problems and Relatively Undecidable Propositions—Account of an Anticipation*. My introductory biographical essay discusses at length the circumstances leading to the writing of this paper and the reasons why it remained unpublished for so long. It did appear in my anthology "The Undecidable," Raven Press 1965, where however it was marred by extremely crude typography seriously interfering with its readability. For this version, I have made some minor inessential changes in notation to improve readability.

I would like to express my gratitude for the help provided by Post's daughter Phyllis Post Goodman, to David M. Jones for his help with the mysteries of TeX, and to my wife Virginia for her patience and support.

Martin Davis
New York University
May 1993

Bibliography of Emil L. Post

- [1] The generalized gamma functions, *Annals of Math.* **20**(1918–1919), 202–217.
- [2] Discussion of problem 433, *American Mathematical Monthly* **26**(1919), 37–39.
- [3] Introduction to a general theory of elementary propositions, *Bulletin of the American Mathematical Society* **26**(1920), 437.
- [4] Determination of all closed systems of truth tables, *Bulletin of the American Mathematical Society* **26**(July, 1920), 437.
- [5] Introduction to a general theory of elementary propositions, (Doctoral dissertation, Columbia University, 1920), *American Journal of Mathematics* **43**(1921), 163–185; reprinted in van Heijenoort: *From Frege to Gödel: A Source Book in Mathematical Logic*, 1879–1931, Source Book in the History of the Sciences Series, Harvard University Press, Cambridge, Massachusetts, 1967.
- [6] On a simple class of deductive systems, *Bulletin of the American Mathematical Society* **27**(1921), 396–397.
- [7] Visual intuition in Lobachevsky space, *Bulletin of the American Mathematical Society* **29**(1923), 9.
- [8] Visual intuition in spherical and elliptic space: Einstein's finite universe, *Bulletin of the American Mathematical Society* **29**(1923), 113–114.
- [9] A non-Weierstrassian method of analytic prolongation, *Bulletin of the American Mathematical Society* **29**(1923), 114.
- [10] A new method for generalizing e^x in the complex domain, *Bulletin of the American Mathematical Society* **29**(1923), 114.
- [11] A simple geometric proof of the equality of the Brochardt angles of a triangle, *Bulletin of the American Mathematical Society* **29**(1923), 114–115.
- [12] Theory of generalized differentiation, *Bulletin of the American Mathematical Society* **30**(1924), 11.
- [13] The m th derivative of a function of a function; calculus of m th derivatives, *Bulletin of the American Mathematical Society* **30**(1924), 207–208.
- [14] Generalized differentiation, *Transactions of the American Mathematical Society* **32**(1930), 723–781.

- [15] Polyadic groups (preliminary report), *Bulletin of the American Mathematical Society* **41**(1935), 796.
- [16] Finite combinatory processes, Formulation I, *Bulletin of the American Mathematical Society* **42**(1936), 810–811.
- [17] Finite combinatory processes — Formulation I, *The Journal of Symbolic Logic* **1**(1936), 103–105.
- [18] Polyadic groups, *Transactions of the American Mathematical Society* **48**(1940), 208–350.
- [19] The Two-Valued Iterative Systems of Mathematical Logic, *Annals of Mathematics Studies*, No. 5, Princeton University Press, Princeton, 1941; reprinted by Kraus Reprint Company, New York, 1965.
- [20] Absolutely unsolvable problems and relatively undecidable propositions — account of an anticipation, first published in Davis (1965), 340–433.
- [21] Formal reductions of the general combinatorial decision problem, *American Journal of Mathematics* **65**(1943), 197–215.
- [22] Recursively enumerable sets of positive integers and their decision problems, *Bulletin of the American Mathematical Society* **50**(1944), 284–316.
- [23] A variant of a recursively unsolvable problem, *Bulletin of the American Mathematical Society* **52**(1946), 264–268.
- [24] Note on a conjecture of Skolem, *The Journal of Symbolic Logic* **11**(1946), 73–74.
- [25] Recursive unsolvability of a problem of Thue, *Bulletin of the American Mathematical Society* **52**(1946), 1015–1016.
- [26] Recursive unsolvability of a problem of Thue, *The Journal of Symbolic Logic* **12**(1947), 1–11.
- [27] Conjuntos recurrentemente numerables de enteros positivos y sus problemas de decision, Spanish translation of [22] by J.R. Fuentes, *Revista Mathematica Hispano-Americanana* **7**(1947), 187–229.
- [28] Degrees of recursive unsolvability (preliminary report), *Bulletin of the American Mathematical Society* **54**(1948), 641–642.
- [29] [with Samuel Linial (who later changed his name to Samuel Gulden)] Recursive unsolvability of the deducibility, Tarski's completeness, and independence of axioms problems of propositional calculus, *Bulletin of the American Mathematical Society* **55**(1949), 50.
- [30] Note on a relation recursion calculus, *The Journal of Symbolic Logic* **16**(1951), 238.
- [31] Solvability, definability, provability; history of an error, *Bulletin of the American Mathematical Society* **59**(1953), 245–246.
- [32] A necessary condition for definability for transfinite von Neumann–Gödel set theory sets, with an application to the problem of the existence of a definable well-ordering of the continuum (preliminary report), *Bulletin of the American Mathematical Society* **59**(1953), 246.

- [33] (with S.C. Kleene) The upper semi-lattice of degrees of recursive unsolvability, *Bulletin of the American Mathematical Society* **59**(1953), 557.
- [34] (with S.C. Kleene) The upper semi-lattice of degrees of recursive unsolvability, *Annals of Mathematics* **59**(1954), 379–407.

Emil L. Post: His Life and Work

Martin Davis

Courant Institute of Mathematical Sciences
New York University

1.

Emil L. Post was born into a Jewish family in Augustow, Poland on February 11, 1897. With his parents Arnold and Pearl Post, Emil emigrated to New York in May 1904. Post attended Townsend Harris High School, a free secondary school for gifted students on the same City College campus where he was to spend so much of his life. His B.S. from City College was in 1917 and he was on the faculty of that institution from 1935 until his death in 1954. His Ph.D. was from Columbia University in 1920. Although Post attended synagogue regularly, like many Jews he defined his connection with the Jewish religion in his own personal way which did not entail obedience to the detailed rules associated with the faith. In 1929 he was married to Gertrude Singer. They are survived by their daughter Phyllis Goodman.¹

Post's life was a struggle with adversity. He managed well the handicap he suffered in childhood when he lost an arm in an accident. But in his scientific labors, he had to overcome obstacles that would have daunted most. He suffered all his adult life from crippling manic-depressive disease at a time when no drug therapy was available for this malady. Until 1935, he was unable to obtain a regular academic position, making his living, for the most part, by teaching in the New York high school system. At City College he worked under conditions that would seem intolerable nowadays. The standard teaching load was 16 contact hours per week. There were no individual faculty offices (everyone shared one large room with a huge table in the center), so Post did his research sitting at a desk in the living room in his small apartment while his young daughter was required to maintain silence. There was no secretarial help, and Post had to type his own letters

¹ I am grateful to Gertrude Post, Phyllis Goodman, John Dawson, and Michael Scanlon for making various documents available to me.

of recommendation for students unless his wife did it for him.² His research in mathematical logic was ahead of its time and very much out of the mainstream of mathematical research in the United States. Post suffered repeated episodes of mania which required institutionalization. Electroshock therapy was believed by his physicians and his family to be the most efficacious treatment. His tragic death from a sudden heart attack occurred in a mental institution shortly after one of these treatments.

2. Post's Doctoral Dissertation

Although Post is remembered chiefly as a logician, his earliest research was in analysis. As an undergraduate, he considered the question: how shall we understand the differential operator D^n when n is not an integer.³ The resulting paper was presented to the American Mathematical Society in 1923, but not published until 1930, [24]. Included in the paper was an important result about inverting the Laplace transform which became known as the *Post-Widder inversion formula*. While a graduate student Post had published a brief paper [21] on the functional equation of the Gamma function.

In 1917 through 1920, when Post was a graduate student at Columbia, Alfred North Whitehead and Bertrand Russell's monumental three volume opus, [38], *Principia Mathematica*, purporting to demonstrate that all of mathematics could and should be regarded as a branch of logic, was still an exciting new development. Cassius Keyser conducted a seminar at Columbia on *Principia* in which Post participated. Another influence was the also recently published [15] which pointed out that for all their expressive power, systems of logic could be regarded as merely dealing with finite strings of symbols. Post's doctoral dissertation [22] clearly shows these two influences: if *Principia* could be regarded as a system for the finitary manipulation of symbols, why could it not be studied by ordinary

² Phyllis Goodman, Post's daughter, has emphasized the important role that her mother played in her father's accomplishments:

My father was a genius; my mother was a saint . . . Besides typing letters of recommendation, my mother also typed my father's manuscripts and correspondence. . . . My mother was also the one who handled all financial matters . . . she was the buffer in daily life that permitted my father to devote his attention to mathematics (as well as to his varied interests in contemporary world affairs). Would he have accomplished so much without her? I, for one, don't think so.

³ The information to the effect that much of this work was done while Post was still an undergraduate comes from his fellow-student, and later colleague B.P. Gill.

mathematical methods? The contributions of Post's dissertation may be viewed as consisting essentially of three new developments:

- The portion of *Principia* which today is called the *propositional calculus* was isolated, the truth table method was introduced, and the axioms of Whitehead and Russell were shown to be *complete* and *consistent* with respect to this method. Post emphasized that the truth table method provides a solution to the decision problem for the propositional calculus. (Post called it the *finiteness* problem.)
- The truth table method was generalized from two truth values to an arbitrary finite number of truth values. There is an extensive literature on multi-valued logic and on "Post algebras," which emanates from this work.⁴
- Finally, and perhaps most remarkable, Post proposed a general framework for systems of logic regarded as systems for inferences via finitary symbol manipulation. Post referred to such systems as being obtained by "generalization by postulation" and later [35] as being in *canonical form A*; today we recognize the strings produced by such a system as simply being an arbitrary recursively enumerable set of strings on a finite alphabet.

Post's dissertation also mentions the result of his investigation of sets T of two-valued truth functions closed under composition. We may speak of a *basis* of such a set as a subset of T which "generates" all of T by composition. The *order* of a finite basis is then the maximum number of arguments of one of its members. The main theorem is that there are a total of 66 different sets of this kind having a basis of order ≤ 3 , and that for each $n > 3$ there are 8 of these sets (all of them infinite) with a basis of order n . The monograph [27] published so very much later is essentially devoted to this theorem.

3. Post's "Anticipation" of Church and Gödel

It was Post's realization that the entire *Principia* could almost certainly be regarded as a system in the Canonical Form A mentioned above, that led him to undertake his ambitious program to solve the decision problem (*finiteness problem* in Post's terms) for systems in Canonical Form A, and thus to provide a procedure which could decide for any given formula of *Principia* whether it was formally derivable in that system. Since *Principia* was intended to formalize all of existing mathematics, Post was proposing no less than to find a single algorithm for all of mathematics. The abstract

⁴ Post algebras are related to many-valued logic in much the way that Boolean algebras are related to the propositional calculus.

[23] reports a solution to the decision problem for a particular class of systems in Canonical Form A.

The award of the prestigious post-doctoral Procter Fellowship, permitted Post to spend the academic year 1920–21 at Princeton University working on this program. He proved the equivalence of Canonical Form A with an apparently weaker Form B and with an extremely general Form C. It is Form C that has survived as *Post production systems*, and so the definition is given here, but using the contemporary terminology of the theory of formal languages (which indeed eventually evolved from these very Post production systems). Thus let Σ be a finite alphabet. A *canonical production* has the form:

$$\begin{array}{ccccccccc}
 g_{11} & P'_{i_1} & g_{12} & P'_{i_2} & \cdots & g_{1m_1} & P'_{i_{m_1}} & g_{1(m_1+1)} \\
 g_{21} & P''_{i_1} & g_{22} & P''_{i_2} & \cdots & g_{2m_2} & P''_{i_{m_2}} & g_{2(m_2+1)} \\
 \cdots & \cdots \\
 g_{k1} & P_{i_1^{(k)}} & g_{k2} & P_{i_2^{(k)}} & \cdots & g_{km_k} & P_{i_{m_k}^{(k)}} & g_{k(m_k+1)} \\
 & & & & & \downarrow & & \\
 g_1 & P_{i_1} & g_2 & P_{i_2} & \cdots & g_m & P_{i_m} & g_{m+1}
 \end{array}$$

Here the g 's are given strings on the alphabet Σ , the P 's are variable strings, i.e., what are called “non-terminals,” and each of the P 's in the line following the \downarrow also occurs as one of the P 's above the \downarrow . A *system in Canonical Form C* consists of a finite set of strings which Post called *initial assertions* together with a finite set of canonical productions. Such a system *generates* the subset of Σ^* consisting of those strings that can be derived from the initial assertions using the canonical productions. Of course, today it is evident that the sets than can be so generated are exactly the recursively enumerable languages.

Post showed that the part of *Principia Mathematica* corresponding to the first-order predicate calculus could be put into Canonical Form B, and therefore by his equivalence theorem, also into Canonical Form C. It was clear to Post that the same techniques could be used to show that the set of provable formulas (*assertions* in Post's terminology) of all of *Principia* could be regarded as the set of strings generated by some system in Canonical Form C. It was Post's remarkable *normal form* theorem, also obtained during his year at Princeton, that led “to a reversal of [Post's] entire program” [35].

A *normal system* is a system in Canonical Form C of a particularly simple form: there is just one initial assertion and each production has the form

$$gP \implies P\bar{g}.$$

A set of strings $U \subseteq \Sigma^*$ is called *normal* if there is a normal system on an alphabet containing Σ generating a set \mathcal{N} such that

$$U = \mathcal{N} \cap \Sigma^*.$$

Post's Normal Form Theorem.⁵ Let $U \subseteq \Sigma^*$ be the set of strings generated by some system in Canonical form C. Then U is normal.

This remarkable theorem makes it clear that a solution to the decision problem for normal systems, with their appealingly simple form, would have led to a decision procedure for all of *Principia Mathematica*. Post's approach may be compared to the work on Hilbert's Entscheidungsproblem in Europe, via quantificational prefixes. Instead Post showed that all the logical complexity was already contained in a (deceptively) simple combinatorial formulation. Various considerations led Post to undertake the study of a special case of normal systems, the so-called *tag* systems. In effect, tag systems are normal systems in which all of the g 's (but not the \bar{g} 's) are of the same length, and in which each \bar{g} depends only on the first symbol of the corresponding g . It was not only that this was a natural special case of the decision problem for normal systems. It also arose as a special case in Post's efforts to solve what would nowadays be called the *unification problem for the ω order predicate calculus*. He had solved the unification problem for systems in Canonical Form A using what Post called the "L.C.M. process."⁶

Post found his work on the problem of tag extraordinarily frustrating. The simple case:

$$\begin{aligned} agP &\implies Paa \\ bgP &\implies Pbab \end{aligned}$$

where g is any string $\in \{a, b\}^*$ such that $|g| = 2$ "proved intractable." (So far as I know, it is still open!) Post's investigations of even simple special cases of *tag* led to

... an overwhelming confusion of classes of cases, with the solution of the corresponding problem depending more and more on problems of ordinary number theory. Since it had been our hope that the known difficulties of number theory would, as it were, be dissolved ... [in normal systems] ..., the solution of the general problem [of tag] appeared hopeless, ... [35].

Post conjectured (in conversation with me) that *tag* would turn out to be recursively unsolvable, and I later had the opportunity to pass this conjecture on to Marvin Minsky when he told me about his newly developed techniques for dealing with Post production systems. Minsky's elegant proof of the unsolvability of the problem of tag [17] would certainly have pleased Post had he lived a few years longer.

⁵ Proofs of the theorem can be found in [28], [35], and [18].

⁶ I'm indebted to J.A. Robinson for calling my attention to the fact that this L.C.M. process must be just the familiar unification algorithm.

The very generality of *Principia Mathematica* made it seem appropriate to conclude that whatever methods one could imagine for systematically “generating” a set of strings on a finite alphabet could be formalized in *Principia*. But Post’s work during the year in Princeton, which made it seem clear that *Principia* itself could be reduced to one of his canonical forms, led him to the conclusion that *any set so generated would have to be normal*. This conclusion, which today we readily recognize as being equivalent to Church’s thesis, I have elsewhere called *Post’s Thesis* [5]. On the other hand a simple diagonalization on an enumeration of all normal systems leads to a set of “ $\{a\}$ -sequences”, i.e. strings on the one symbol alphabet $\{a\}$, which can not be normal. This seems to contradict Post’s Thesis, for is this diagonal set itself not “generated” by a systematic process? No, for⁷

... we have merely *defined* a set of $\{a\}$ -sequences whereas to yield a true counter-example we must show how to *generate* that set, i.e., set up a system of “combinatory iteration” whose operations would at some time yield each and every $\{a\}$ -sequence not in the set. On the other hand, suppose that the finiteness problem were solved for the class of all normal systems. Then for each of the above normal systems that solution would in a finite number of steps tell whether $aa\dots a$ with m a ’s is or is not in that m -th normal system. An operation could then be set up which in order would pass down the above normal systems, for the m -th apply the test for $aa\dots a$ with m a ’s being or not being in the system, have a production which in the latter case produces $aa\dots a$ with m a ’s for the desired system, and then in any case pass on to the next system. This operation iterated would then actually generate the above defined set of a -sequences. That is, a solution of the finiteness problem for all normal systems would yield a counter-example disproving the correctness of our proposed generalization.⁸

Now we mentioned ... how an ... attempt to solve ... “tag” led to ever increasing difficulties, with all the complexities of number theory in the offing. On the other hand, nothing in the above argument directly weakens the reasoning that led us to our generalization. We therefore hold on to that generalization and conclude that the finiteness problem for the class of all normal systems is unsolvable, ...

⁷ The quotation is from Post’s later account based on his notes of that time [35].

⁸ i.e., of what I have been calling Post’s Thesis.

The correctness of this result is clearly entirely dependent on the trustworthiness of the analysis leading to the above generalization. . . it is fundamentally weak in its reliance on the logic of Principia Mathematica . . for full generality a complete analysis would have to be given of all the possible ways in which the human mind could set up finite processes for generating sequences.

From these conclusions it was a short step to the realization that all "symbolic logics" were necessarily incomplete with respect to what they could prove about membership and non-membership in specific normal sets. Thus, Post had anticipated the major results found only a decade later by Gödel, Church, and Turing. Indeed, he fully recognized that his development was "fragmentary". It would not be possible, he felt, to gain acceptance of these results without the "complete analysis" referred to above. For a discussion of these matters both in the context of Post's work and with respect to the later developments, see [5].

Post reacted to the appearance of the revolutionary [8] in 1931 with some anguish at the thought that results that he had anticipated some years earlier were now definitively to be ascribed to someone else, but also with deep admiration for the way Gödel had cut through the thicket of technical difficulties and produced a clean and utterly rigorous proof of the incompleteness theorem. A postcard addressed to Gödel and dated October 29, 1938 reads as follows:

I am afraid that I took advantage of you on this, I hope but our first meeting. But for fifteen years I had carried around the thought of astounding the mathematical world with my unorthodox ideas, and meeting the man chiefly responsible for the vanishing of that dream rather carried me away.

Since you seemed interested in my way of arriving at these new developments perhaps Church can show you a long letter I wrote to him about them. As for any claims I might make perhaps the best I can say is that I would have *proved* Gödel's Theorem in 1921 – had I been Gödel.

In a letter to Gödel dated October 30, 1938, Post in comparing his "anticipation" with what Gödel had done remarked "... after all it is not ideas but the execution of ideas that constitute a mark of greatness."

With Church's announcement of an unsolvable problem in elementary number theory in 1935, Post could no longer hope to claim a "no finite method" result as his own either. He had waited too long, and scientific progress had caught up with and overtaken him. However, Post had no intention of being left out of the exciting new developments he was certain would come out of the work of Gödel and Church. Believing that the

Herbrand-Gödel notion of general recursiveness and the Church-Kleene notion of λ -definability were both lacking in that neither constituted a “fundamental”⁹ analysis of the notion of algorithmic process. Post proposed as suitably “fundamental” the operations of *marking* an empty “box” or erasing the mark in a marked box. These boxes were to be arranged in a row and a “worker” was to move among them marking and erasing as directed by a pre-assigned list of instructions. Published in 1936 in Volume 1 of the brand new *Journal of Symbolic Logic*, [26] put forth a formulation of computability which was essentially identical to that which Alan Turing had developed in England at about the same time.¹⁰ The difference between the two developments is interesting because Turing’s formulation was in terms of an idealized computing machine whereas Post conceived of what today would be called a computer program, i.e., a list of instructions in a narrowly circumscribed formal language.

In addition, Post tried to tell the story of his own anticipation of the work of Gödel, Church, and Turing. His paper [35] (from which we have been quoting extensively here), only published in 1965 in the anthology [4], was submitted to the American Journal of Mathematics in 1941. Here is the full text of Post’s letter of submission to the editor, Hermann Weyl:

Dear Professor Weyl:

It is with some trepidation that I submit the accompanying paper, “Absolutely Unsolvable Problems and Relatively Undecidable Propositions, Account of an Anticipation,” for publication in the AJM. The reason for my trepidation is in part evident in the title, for my seeking publication at this late day from the opening paragraph of the introduction. As to why I did not attempt publication twenty years ago, when this work was in progress, may I point out that [22] was accepted only on condition that its original length be cut by one-third, that [27] was returned to me by the Annals of Mathematics at the height of the above work without any editorial commitment, and with a very mixed report from the referee. It therefore seemed to me to be hopeless to seek publication of Part One of the present paper. And without it, the then revolutionary Part Two would have seemed but idle chatter. An attempt to obtain a full proof development was interrupted by ill health and led to a constantly receding date of ultimate publication. While at this late date the present paper cannot have the primary importance it might have had twenty years ago, nevertheless,

⁹ This was the word Post used in conversation with me. See also Post’s distinction between “atomic” and “molecular” acts in [29], §11.

¹⁰ Of course, Turing’s work was done in ignorance of Church’s while Post’s was not. See [5].

as I've indicated in the introduction, it may still have a secondary importance at least sufficient to justify the present publication. At any rate, I trust it will receive your serious consideration.

In rejecting the paper as written, Weyl wrote in a letter dated March 2, 1942:

...I have little doubt that twenty years ago your work, partly because of its then revolutionary character, did not find its due recognition. However, we cannot turn the clock back; in the meantime Gödel, Church and others have done what they have done, and the American Journal is no place for historical accounts; ... (Personally, you may be comforted by the certainty that most of the leading logicians, at least in this country, know in a general way of your anticipation.)

Weyl went on to quote the referee as noting that the normal form theorem was both new and important and was in an important direction from which "one might ultimately to obtain proofs of the unsolvability for various ... mathematical problems - e.g. the word problem for groups ..." Post followed the referee's recommendations and the editors eventually accepted a very abbreviated version of the paper, which contained only the proof of Post's Normal Form Theorem [28] with the history very briefly recounted in a long footnote.

4. Polyadic Groups

Post's longest publication was not in mathematical logic at all. [25] considered a notion of *Polyadic group* which generalizes the group concept by basing itself on an operation on n arguments where n is a definite integer ≥ 2 . There are two axioms for an n -group. One is a generalized associative law and the other states the unique solvability of the equation

$$f(x_1, x_2, \dots, x_n) = y$$

in each x_i where f is the n -group operation, and $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, y are given elements of the n -group. Evidently, the case $n = 2$ reduces to ordinary groups.

Some sense of the importance of this work can be gathered from Reinhold Baer's review [1]:

The importance of this concept of polyadic group is due to the fact, discovered by the author, that the theory of polyadic groups is exactly equivalent to the theory of finite cyclic extensions of ordinary groups. More precisely:

If X is a coset of the ordinary group G modulo its normal subgroup N such that X generates G/N and such that $m-1$ is the order of X in G/N , then any product of m elements in X is an element in X and the elements in X form an m -group under this operation. Conversely, it is possible to represent every m -group in one and essentially only one way in this fashion as a generating coset of a cyclic quotient group. ... This theory is concerned partly with a generalization of the facts of the classical theory of ordinary groups which is possible to a really astonishing degree,

5. Recursively Enumerable Sets

In 1943, Post was invited to address the American Mathematical Society. The address was published as [29], perhaps Post's most influential publication. In this paper, for the first time, the theory of recursive unsolvability was presented as an autonomous branch of mathematics, a branch that "...stripped of its formalism, ... admits of an intuitive development" that can readily be followed by mathematicians "laymen though [they] may be" in mathematical logic. Proofs were presented in just such an intuitive form although Post was careful to note that without the "forbidding, diverse and alien formalisms" the subject "would lose most of its cogency," and to emphasize that he had insisted on obtaining "formal proofs" of most of the theorems of the paper. On the other hand, Post insisted that

...the real mathematics involved must lie in the informal development. For in every instance the informal "proof" was first obtained; and once gotten, transforming it into the formal proof turned out to be a routine chore.

Post evidently intended to eventually produce formal proofs for publication, noting that: "Our present formal proofs, while complete, will require drastic systematization and condensation prior to publication."¹¹

The paper begins with the main relationships between recursive and recursively enumerable (r.e.) sets. The term *recursively enumerable* originally meant being enumerated by some recursive function. By "arbitrarily extend[ing]" the r.e. concept to include the empty set, Post could state for the first time the key characterization of recursive sets as r.e. sets with r.e. complements. In addition to this key characterization, Post included among his initial results, the facts that every infinite r.e. set contains an

¹¹ In fact, Post at that time had not worked out for himself some of the basic formal techniques of recursion theory, in particular the $S\text{-}m\text{-}n$ theorem, and for this reason his "formal proofs" were far more verbose and complicated than necessary.

infinite recursive subset and that there exists an r.e. set that is not recursive, as well as a “miniature” form of Gödel’s incompleteness theorem. But it was the concern with various ways in which one r.e. set could be regarded as *reducible* to another that was the truly original contribution of [29].

What had been the “complete” normal set K in [35], now appears as the *complete r.e. set K* , the set of all natural numbers that code pairs (n, B) where n (or rather the string $|^n$) is generated by the normal system represented by the string B . Post showed that for every r.e. set S , there is a recursive function f such that

$$x \in S \iff f(x) \in K.$$

That is, K has maximal degree with respect to what Post called *many-one reducibility*. f could even be made one-one, so K has maximal degree even with respect to 1-1 reducibility. However, Post constructed a *simple* r.e. set (i.e. one whose complement contains no infinite r.e. subset) and which therefore (as is shown) can not be of maximal degree with respect to one-one reducibility. The main question posed in this paper, a question which became known as *Post’s Problem*, was simply: *Is there a non-recursive r.e. set of strictly lower degree of unsolvability than K with respect to arbitrary recursive reducibility* (also called *Turing reducibility*)? The approach was to construct r.e. sets with very “thin” complements in an attempt to obtain lower degrees of unsolvability. Post showed that simple sets were guaranteed to be of lower degree than K not only with respect to one-one reducibility, but also with respect to many-one reducibility and even with respect to a much weaker notion, *reducibility by bounded truth tables*. However, it turned out that a simple set could be constructed to which K is reducible by the still weaker *unbounded truth table reduction*. Finally, using a rather heroic combinatorial argument Post showed how to obtain a simple set which was in addition what he called *hypersimple*. Every hypersimple set then turned out to be of strictly lower degree than K with respect to unbounded truth tables. However, regarding “Post’s Problem” itself, the question of the existence of an r.e. set of strictly lower degree than K with respect to Turing reducibility, “we are left completely on the fence.”

We may cite a number of important long term effects of [29].¹² The informal mode of exposition it introduced has very much become the norm in recursion theory, though it is not clear that Post himself would have been pleased with this consequence of his work. The taxonomic study of the r.e. sets originated in this paper has had a vigorous further development. The various kinds of reducibility Post introduced have been extensively studied, both in their original context, and more recently, as the analogous notions

¹² Post’s remark in this paper that Hilbert’s tenth problem “begs for an unsolvability proof” had a major influence on my own work.

when recursive functions are replaced by polynomial time computable functions. In particular the polynomial time analogue of many-one reducibility has been studied as *polynomial time reducibility*, and the notion of being of maximal degree with respect to polynomial time reducibility for a given class has been studied as *poly-time completeness*, e.g. NP-completeness. Post's Problem itself was the impetus for many investigations. What ultimately turned out to be the right direction was taken by Post himself in the abstract [32]¹³, in which degrees of unsolvability less than that of K were indeed obtained, but not for r.e. sets. The ingenious and extremely fruitful priority method developed independently by Friedberg [7] and Muchnik [19] was needed to render Post's construction sufficiently effective that an r.e. set is obtained of degree less than that of K .

6. Unsolvability of Problems in Combinatorial Mathematics

Church and Turing¹⁴ each used their development of a formal explication of intuitive computability to show that Hilbert's Entscheidungsproblem (the decision problem for ordinary first order logic) had no algorithmic solution.¹⁵ Church [2] emphasized that the unsolvable problems obtained were of the same *logical* character as a host of ordinary mathematical questions. However, for many years there were no actual examples of unsolvable problems of specific mathematical interest not directly involving mathematical logic or one of the formalisms used in defining recursive functions.

In [30] what has become known as the *Post Correspondence Problem* was defined and shown to be unsolvable, beginning with the unsolvability of the decision problem for Post normal systems. A *correspondence system* is a finite set of ordered pairs of strings on some finite alphabet, $(g_1, h_1), (g_2, h_2), \dots, (g_n, h_n)$. A *solution* to this system is a sequence of integers i_1, i_2, \dots, i_k such that $1 \leq i_1, i_2, \dots, i_k \leq n$ and

$$g_{i_1} g_{i_2} \cdots g_{i_k} = h_{i_1} h_{i_2} \cdots h_{i_k}$$

The Post Correspondence Problem is to provide an algorithm for determining of a given correspondence system whether it possesses a solution. Post showed that there is no such algorithm.

The unsolvability of the Post Correspondence Problem turned out to be exactly what was needed to obtain unsolvability results in the theory of formal languages, and it has been cited often.¹⁶

¹³ Fleshed out with formal proofs and with many additional results by Kleene, resulting in the published paper [14]; see §7.

¹⁴ [3] and [36].

¹⁵ Indeed in Turing's case, it was precisely the prospect that this negative result could be obtained that led him to study computability.

¹⁶ See for example [12].

Church suggested to Post that the methods used to show that his Correspondence Problem is unsolvable might be applicable to a particular problem whose credentials as being of genuine and independent mathematical interest were quite unimpeachable. This problem, posed by the Norwegian mathematician Axel Thue in 1914, and often called the *word problem* for *monoids* or *semigroups*, may be stated as follows: An equivalence relation \approx on the set Σ^* of strings on the alphabet Σ is defined by first declaring an initial finite set of pairs of strings for which the relation holds,

$$u_1 \approx v_1, u_2 \approx v_2, \dots, u_m \approx v_m, \quad (1)$$

and then letting \approx be the “least” equivalence relation that contains these pairs and is closed under arbitrary substitutions of substrings v_i for u_i or vice versa. An algorithm is then sought which will determine for an arbitrary pair (u, v) of strings, whether or not $u \approx v$. In [31], it is shown how to construct a set of ordered pairs (1) for which the word problem is unsolvable.

The method of proof used by Post was to show how the theory of Turing machines could be mapped into Thue’s problem in such a way that an algorithm for the latter could be transformed into an algorithm for a problem concerning Turing machines known to be unsolvable. However, Post’s scruples prevented him from simply relying on Turing’s work to obtain his result. Turing’s brilliant, indeed revolutionary, work, when examined in detail, turned out to be riddled with systematic errors.¹⁷ In addition, to establishing what he needed independently, Post also presented a very careful technical critique of [36].

The unsolvability of Thue’s problem was obtained independently by the Russian mathematician A. A. Markov [16]. Interestingly enough, Markov based his proof directly on Post normal systems.

7. Degrees of Unsolvability

Although the notion of degree of unsolvability is perfectly general, in [29] it was only considered for r.e. sets. In the “preliminary report” [32], Post announced his first findings on the general notion. The main result was the existence of a pair of *incomparable* degrees of unsolvability both of lower degree than that of the complete r.e. set K , thereby incidentally proving the existence of such lower degrees. The same abstract announces what Kleene [13] later called “Post’s Theorem”: a set is recursive relative to an “oracle” definable in arithmetic in prenex form with n alternations of quantifiers if and only if the set itself is definable in arithmetic by prenex formulas with $n + 1$ alternations of quantifiers both with \exists first and with \forall first. Thus, the sets of lower degree than that of K found by Post are representable in

¹⁷ really “bugs” in Turing’s programs.

both of the forms $\{u \mid (\forall x)(\exists y)R(x, y, u)\}$ and $\{u \mid (\exists x)(\forall y)S(x, y, u)\}$ and where R and S are both recursive.

These results were obtained by Post in 1947 when I was an undergraduate student at City College. With another student¹⁸, I had begun an “honors” reading course on mathematical logic under Post’s tutelage. We had just reached the “deduction theorem” in propositional calculus, when Post met us full of excitement about his new work. We did not meet with Post again that semester; we had witnessed the beginning of one of his manic episodes. Apparently it was precisely excitement that had the potential to push him over the edge.¹⁹ Between continuing episodes of his illness, and Post’s continuing difficulty with producing succinct “formal” proofs, Post finally sent his work to Kleene with the suggestion that a student could write the work up using Kleene’s formalism and be a joint author. Instead, Kleene worked on the material himself and greatly strengthened Post’s results. In particular, Kleene proved that the ordered set of degrees contains a subset in which the degree ordering is dense. The final result was the very influential [14]. As mentioned above, the priority method of Friedberg and Mucnik which was used in their (independent) solutions to Post’s problem, can be thought of as a constructivization of Post’s method for obtaining incomparable degrees.²⁰

8. Solvability, Definability, Provability

As we have seen, Post’s technical contributions to recursion theory are of great importance. Yet his emphasis was always on the significance of the *absoluteness* and *fundamental* character of the notion of recursive solvability. This same emphasis appears in Gödel’s [9], where he speaks of a “kind of miracle” (see also [5]). Post and Gödel both hoped to find other similarly fundamental and absolute notions in the foundations of mathematics. Post spent many years attempting to elucidate the notion of *proof* of arithmetic propositions. The hoped for concept would do for arithmetic provability what recursiveness did for algorithmic solvability. As the explication of solvability led to recursively unsolvable problems, so a proper explication of arithmetic provability should lead to *absolutely* undecidable arithmetic propositions. Much of Post’s work on “provability” took the form of a study of the provability of arithmetic propositions in the particular system

¹⁸ John Stachel, who is currently Director of the Center for Einstein Studies at Boston University.

¹⁹ Part of Post’s technique for dealing with this problem was to severely limit the length of time spent on each of his research sessions, and to always alternate between two different projects to help avoid undue excitement.

²⁰ In conversation, Post insisted that his technique was just a very easy use of Cantor’s diagonal method. The priority method is certainly much deeper.

P of [8], and in particular of the role of the very strong comprehension principle of P .

Only a year before his death, Post's thinking took a turn. He had always felt that along with "solvability" and "provability", a third notion for which an "absolute" explication was needed was general mathematical "definability". What came to Post was what he thought of as a *necessary* condition for definability. And he then concluded that it had been an error to attempt to explicate provability before definability. It turned out that Post's "necessary condition" was not new. In an equivalent form it had been proposed by Gödel at the Princeton Bicentennial Conference on Problems in Mathematics in 1946 [9]. This notion is now known as *hereditary ordinal definability*; it was rediscovered once more by John Myhill and Dana Scott [20] during the 1960s. Post presented his ideas to the American Mathematical Society in the two brief abstracts [33], [34].

Although none of Post's work on provability or definability has been published, his working notes in bound notebooks remain. They are housed in the library of the American Philosophical Society in Philadelphia.

9. Post As a Teacher

During the 1930s and 1940s, the undergraduate student body at City College in New York included a large group of mathematically talented young people. Today, most of us are professional mathematicians holding positions in various academic institutions. When we meet, we are apt to reminisce about Post's remarkable classes, and their effect on us.

Post's classes were tautly organized tense affairs. Each period would begin with student recitations covering problems and proofs of theorems from the day's assignment. These were handed out apparently at random and had to be put on the blackboard without the aid of textbook or notes. Woe betide the hapless student who was unprepared. He (or rarely she) would have to face Post's "more in sorrow than anger" look. In turn, the students would recite on their work. Afterwards, Post would get out his 3 by 5 cards and explain various fine points. The class would be a success if he completed his last card just as the bell rang. Questions from the class were discouraged: there was no time. Surprisingly, these inelastic pedagogic methods were extremely successful, and Post was a very popular teacher. There were other inspiring mathematics professors at City College, but none were as effective as Post with the weak as well as with the strong students. Questions were discouraged, but students usually found that the answer to the question they would have wanted to ask was there on one of Post's 3 by 5 cards.

Post was a stickler for care and precision in mathematical discourse. One of the enduring lessons from his classes was the frequency of lapses in rigor to be found in the textbooks we were using. In the case of the textbook on real variable theory by E. J. Townsend, the errors were so

ubiquitous that Post had prepared his own extensive list of corrections for distribution to the students. None of us would be likely to make the mistake in our future lives of regarding an assertion as fact simply because it was printed in a book.

10. The Legacy of Emil L. Post

Mathematical logic was once seen as largely concerned with developing formal systems, *within* which ordinary mathematical arguments could be expressed in a purely "symbolic" form. Today, we see it as rather concerned with the study of formal systems and their capabilities *using* ordinary mathematical methods. This change is not to be attributed to any one person, but Post's ideas have certainly played a role in the transformation.

Entire mathematical subjects with their own flourishing literature were initiated by Post:

- multi-valued logic and Post algebras
- the theory of degrees of unsolvability and in particular of r.e. degrees
- the study of various kinds of recursive reducibility
- unsolvability results in combinatorial algebra.

The extent of the influence of Post's ideas on computer science, which has already been mentioned in passing, is particularly remarkable considering Post's apparent lack of any interest in computing machines. For a survey of Post's work from this point of view, see [6].

Post's significance transcends his scientific contributions, important as those were. He remains an inspiration as well, for the manner in which he overcame his potentially crippling mental disability, for his distinctive voice, and for his continued devotion to his science and his students.

References

- [1] Reinhold Baer, Review of [25], *Mathematical Reviews*, vol. 2 (1941), p. 128.
- [2] Alonzo Church, "An Unsolvable Problem of Elementary Number Theory," *American Journal of Mathematics*, vol. 58 (1936), pp. 345–363.
- [3] Alonzo Church, "A Note on the Entscheidungsproblem," *J. Symb. Logic*, vol. 1 (1936), pp. 40–41; Correction, *ibid* pp. 101–102. Edited version including corrections, [4], pp. 110–115.
- [4] Martin Davis, *The Undecidable*, Raven Press, New York 1965.
- [5] Martin Davis, "Why Gödel Didn't Have Church's Thesis," *Information and Control* vol. 54 (1982), pp. 3–24.

- [6] Martin Davis, "Emil Post's Contributions to Computer Science," *Proceedings Fourth Annual Symposium on Logic in Computer Science*, pp. 134–136, IEEE Computer Society Press, Washington, D.C., 1989.
- [7] Richard Friedberg, "Two Recursively Enumerable Sets of Incomparable Degrees of Unsolvability (Solution of Post's Problem)," *Proc. Natl. Acad. Sciences (USA)*, vol. 43(1957), pp. 236–238.
- [8] Kurt Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandte Systeme I," *Monatshefte für Mathematik und Physik*, vol. 38 (1931), pp. 173–198. Reprinted [10], pp. 144–194 (even numbered pages). Translation by Jean van Heijenoort, in [37] pp. 596–616. Reprinted [10], pp. 145–195 (odd numbered pages).
- [9] Kurt Gödel, "Remarks before the Princeton Bicentennial Conference on Problems in Mathematics," [4], pp. 84–88. Reprinted [11], pp. 150–153.
- [10] Kurt Gödel, *Collected Works, volume I, Publications 1929–1936*, edited by Solomon Feferman, et al., Oxford University Press, 1986.
- [11] Kurt Gödel, *Collected Works, volume II, Publications 1938–1974*, edited by Solomon Feferman, et al., Oxford University Press, 1990.
- [12] Michael Harrison, *Introduction to Formal Language Theory*, Addison-Wesley, 1978.
- [13] S.C. Kleene, *Introduction to Metamathematics*, D. Van Nostrand Company, Inc., 1952.
- [14] S.C. Kleene and Emil Post, "The Upper Semi-lattice of Degrees of Recursive Unsolvability," *Annals of Mathematics*, ser. 2, vol. 59 (1954), pp. 379–407.
- [15] Clarence I. Lewis, *A Survey of Symbolic Logic*, University of California Press, Berkeley, 1918.
- [16] A. A. Markov, "On the Impossibility of Certain Algorithms in the Theory of Associative Systems" (Russian), *Doklady Akad. Nauk S.S.R.*, vol. 55 (1947), pp. 587–590. English translation, *Comptes rendus de l'académie des sciences de l'U.R. S.S.*, vol. 55 (1947), pp. 583–586.
- [17] Marvin Minsky, "Recursive Unsolvability of Post's Problem of Tag and Other Topics in the Theory of Turing Machines," *Annals of Math.*, vol. 74 (1961), pp. 437–455.
- [18] Marvin Minsky, *Computation: Finite and Infinite Machines*, Prentice-Hall 1967.
- [19] A.A. Muchnik, "Negative Answer to the Problem of Reducibility of the Theory of Algorithms" (Russian), *Doklady Akad. Nauk S.S.R.*, n.s. vol. 108 (1956), pp. 194–197.
- [20] John Myhill and Dana Scott, "Ordinal Definability," *Proc. Symposia in Pure Math.*, vol. 13 (1971), pp. 271–278.
- [21] Emil Post, "The Generalized Gamma functions," *Annals of Mathematics*, ser. 2, vol. 20 (1918–19), pp. 202–217. Reprinted this volume, pp. 1–16.
- [22] Emil Post, "Introduction to a General Theory of Propositional Functions," *Amer. J. Math.*, vol. 43 (1921), pp. 163–185. Reprinted [37], pp. 264–283; this volume pp. 21–43.

- [23] Emil Post, "On a Simple Case of Deductive Systems" (abstract), *Bull. Amer. Math. Soc.*, vol. 27 (1921), pp. 396–397. Reprinted this volume, p. 545.
- [24] Emil Post, "Generalized Differentiation," *Trans. Amer. Math Soc.*, vol. 32 (1930), pp. 723–781. Reprinted this volume, pp. 44–102.
- [25] Emil Post, "Polyadic Groups," *Trans. Amer. Math Soc.*, vol. 48 (1940), pp. 208–350. Reprinted this volume, p. 548.
- [26] Emil Post, "Finite Combinatory Processes - Formulation I," *J. Symb. Logic*, vol. 1(1936), pp. 103–105. *Reprinted* [4], pp. 289–291; this volume pp. 103–105.
- [27] Emil Post, *The Two-Valued Iterative Systems of Mathematical Logic*, Annals of Mathematics Studies, No. 5, Princeton University Press, 1941. Reprinted this volume, pp. 249–374.
- [28] Emil Post, "Formal Reductions of the General Combinatorial Decision Problem," *Amer. J. Math.*, vol. 65 (1941), pp. 197–215. Reprinted this volume, pp. 442–460.
- [29] Emil Post, "Recursively Enumerable Sets of Positive Integers and Their Decision Problems," *Bull. Amer. Math. Soc.*, vol. 50 (1944), pp. 284–316. *Reprinted* [4], pp. 305–337; this volume pp. 461–494.
- [30] Emil Post, "A Variant of a Recursively Unsolvable Problem," *Bull. Amer. Math. Soc.*, vol. 52 (1946), pp. 264–268. Reprinted this volume, pp. 495–500.
- [31] Emil Post, "Recursive Unsolvability of a Problem of Thue," *J. Symb. Logic*, vol. 12 (1947), pp. 1–11. *Reprinted* [4], pp. 293–303; this volume pp. 503–512.
- [32] Emil Post, "Degrees of Recursive Unsolvability: Preliminary Report" (abstract) *Bull. Amer. Math. Soc.*, vol. 54 (1948), pp. 641–642. Reprinted this volume, pp. 549–550.
- [33] Emil Post, "Solvability, Definability, Provability; History of an Error," *Bull. Amer. Math Soc.*, vol. 59 (1953), pp. 245–246.
- [34] Emil Post, "A Necessary Condition for Definability for Transfinite von Neumann-Gödel Set Theory Sets, with an Application to the Problem of the Existence of a Definable Well-Ordering of the Continuum," *Bull. Amer. Math. Soc.*, vol. 59 (1953), p. 246. Reprinted this volume, pp. 551–552.
- [35] Emil Post, "Absolutely Unsolvable Problems and Relatively Undecidable Propositions: Account of an Anticipation," [4], pp. 340–433. Reprinted this volume, pp. 375–441.
- [36] Alan Turing, "On Computable Numbers with an Application to the Entscheidungsproblem," *Proc. London Math. Soc.*, ser. 2, vol. 42 (1936–37), pp. 230–265. Correction, *ibid*, vol. 43 (1937), pp. 544–546. Reprinted [4], pp. 116–154.
- [37] Jean van Heijenoort, *From Frege to Gödel*, Harvard University Press 1967.
- [38] Alfred Whitehead and Bertrand Russell, *Principia Mathematica*, Cambridge University Press, vol. 1 1910, vol. 2 1912, vol. 3 1913.

THE GENERALIZED GAMMA FUNCTIONS.

BY EMIL L. POST.

Introduction.

The difference equation

$$\phi(z + 1) = f(z)\phi(z) \quad (1)$$

has been studied directly* and indirectly† through

$$\phi(z + 1) - \phi(z) = \psi(z)$$

in the cases where $f(z)$ is a meromorphic function. In the present paper a solution of (1) is obtained under an entirely different assumption with regard to $f(z)$. One class of functions satisfying this condition is of the form

$$g(z)h(z)e^{\psi(z)}$$

where $g(z)$ is regular at infinity, and $h(z)$ and $\psi(z)$ are any algebraic functions, Abelian integrals, or finite combinations of these. It is also attempted to bring out the similarity of the solution and its properties to those of the ordinary Gamma function.

In part I a solution of (1) is obtained as an infinite product. An asymptotic expression is obtained for it, as well as an infinite integral. In part II a number of relations are obtained of which the generalization of the multiplication theorem of the ordinary Gamma function is characteristic.

PART I: FUNDAMENTAL EXPANSIONS.

1. Construction of the Gaussian Form of the Generalized Gamma Functions.
Let $f(z)$ satisfy the following two conditions: (a) that $\log f(z)$ be analytic in a sector enclosing the positive end of the real axis; (b) that for some value of r , a positive real value of ϵ may be found such that

$$\lim_{p \rightarrow +\infty} p^{1+\epsilon} \frac{d^{r+1}}{dz^{r+1}} \log f(z + p) = 0$$

uniformly over any finite region of the z plane. Under these conditions a

* Mellin, Acta Math., vol. 8 (1886), pp. 37–80; Barnes, Proc. London Math. Soc., ser. 2, vol. 2 (1905), pp. 438–469.

† Guichard, Ann. de L'Ecole Norm., ser. 3, vol. 4 (1887), pp. 361–380; Appell, Journ. de Math., ser. 4, vol. 7 (1891), pp. 157–219; Hurwitz, Acta Math., vol. 20 (1896), pp. 285–312.

solution of

$$\phi(z+1) = f(z)\phi(z), \quad (1)$$

denoted by $\Gamma_{f(u)}(z)$, will be obtained which is entirely analogous to the ordinary Gamma function.

In analogy with

$$\Gamma(z) = \lim_{p \rightarrow \infty} \frac{1 \cdot 2 \cdots (p-1)}{z(z+1) \cdots (z+p-1)} p^z$$

set

$$\Gamma_{f(u)}(z) = \lim_{p \rightarrow \infty} \frac{f(1)f(2) \cdots f(p-1)}{f(z)f(z+1) \cdots f(z+p-1)} [f(p)]^z F(p, z) \quad (2)$$

where $F(p, z)$ is to be determined so that (2) converges, and satisfies the fundamental difference equation. For the latter condition

$$\lim_{p \rightarrow \infty} \frac{F(p, z+1)/F(p, z)}{f(z+p)/f(p)} = 1 \quad (3)$$

Taking into account (a) and (b), we may have condition (3) fulfilled by setting

$$\log F(p, z) = \frac{d}{dp} \log f(p) \frac{\phi_2(z)}{2!} + \cdots + \frac{d^r}{dp^r} \log f(p) \frac{\phi_{r+1}(z)}{(r+1)!} \quad (4)$$

where $\phi^2(z), \dots, \phi_{r+1}(z)$ are the Bernouillian polynomials.* Later we shall show that the same value of $F(p, z)$ insures the convergence of (2).

Since $\phi_n(1) = 0$,

$$\Gamma_{f(u)}(1) = 1$$

so that where q is a positive integer

$$\Gamma_{f(u)}(q) = f(1)f(2) \cdots f(q-1) = \underline{|f(q-1)|}$$

in an evident notation.

The general solution of (1) may be written

$$\phi(z) = \Gamma_{f(u)}(z)P(z) \quad (5)$$

where $P(z)$ is any periodic function of period unity. Clearly $\Gamma_{f(u)}(z)$ is the only solution of (1) such that

$$\lim_{p \rightarrow \infty} \frac{\phi(z+p)}{[f(p-1)[f(p)]^z F(p, z)]} = 1 \quad (6)$$

Equation (6) in connection with (1) may therefore be taken as defining $\Gamma_{f(u)}(z)$.

* Whittaker and Watson, Modern Analysis, Second edition, pp. 126, 127.

2. Eulerian Form and Convergence. Equation 2, §1 may be rewritten as follows:

$$\begin{aligned}\Gamma_{f(u)}(z) &= \frac{[f(1)]^z F(1, z)}{f(z)} \prod_{p=1}^{\infty} \frac{f(p)}{f(z+p)} \left[\frac{f(p+1)}{f(p)} \right]^z \frac{F(p+1, z)}{F(p, z)} \quad (1) \\ &= \frac{[f(1)]^z F(1, z)}{f(z)} \prod_{p=1}^{\infty} Q(p).\end{aligned}$$

Since

$$\frac{\phi_2(z)}{(s-1)! 2!} + \frac{\phi_3(z)}{(s-2)! 3!} + \cdots + \frac{\phi_s(z)}{s!} = \frac{z^s - z}{s!},$$

we have

$$\begin{aligned}\log F(p+1, z) &= \log F(p, z) + \sum_{s=1}^r \frac{z^s - z}{s!} \frac{d^s \log f(p)}{dp^s} \\ &\quad + \sum_{s=1}^{r-1} \lambda_s \frac{d^{r+1} \log f(p+\theta_s)}{dp^{r+1}} \frac{\phi_{s+1}(z)}{(r-s+1)! (s+1)!} \\ |\lambda_s| &\leq 1; \quad 0 < \theta_s < 1,\end{aligned}$$

by using Darboux's* form of the remainder in Taylor's series for complex variables in connection with (4 §1). Hence substituting in (1) and reducing we obtain

$$\begin{aligned}\log Q(p) &= \sum_{s=1}^{r-1} \lambda_s \frac{d^{r+1} \log f(p+\theta_s)}{dp^{r+1}} \frac{\phi_{s+1}(z)}{(r-s+1)! (s+1)!} \\ &\quad + \frac{z\lambda_r}{(r+1)!} \frac{d^{r+1}}{dp^{r+1}} \log f(p+\theta_r) - \frac{z^{r+1}\lambda_{r+1}}{(r+1)!} \frac{d^{r+1}}{dz^{r+1}} \log f(p+\theta_{r+1}z).\end{aligned}$$

By means of condition (b) we see that from some value of (p) on, the terms of $\Sigma \log Q(p)$ are less in absolute value than those of the convergent series $\Sigma(1/(p^{1+\epsilon}))$. Hence (1) is absolutely convergent for all finite values of z which are not zeros of some $f(z+p)$. Condition (b) likewise proves (1) to be uniformly convergent over any finite region of the z plane which excludes these points. $\Gamma_{f(u)}(z)$ is therefore an analytic function, except for isolated points, in the entire sector at least over which $f(z)$ is analytic.

3. Weierstrassian Form, and Derivatives. When $r \leq 1$,† equations (2 §1) and (1 §2) become

$$\Gamma_{f(u)}(z) = \lim_{p \rightarrow \infty} \frac{f(1)f(2) \cdots f(p-1)}{f(z)f(z+1) \cdots f(z+p-1)} [f(p)]^z, \quad (1)$$

and

$$\Gamma_{f(u)}(z) = \frac{[f(1)]^z}{f(z)} \prod_{p=1}^{\infty} \left\{ \frac{f(p)}{f(z+p)} \left[\frac{f(p+1)}{f(p)} \right]^z \right\}. \quad (2)$$

Since (1) is uniformly convergent, we may differentiate logarithmically so

* Journ. de Math., series 3, vol. 2 (1876), p. 291.

† Whenever r can be taken ≤ 1 . This is not in general true.

that

$$\frac{\Gamma'_{f(u)}(z)}{\Gamma_{f(u)}(z)} = -\gamma_{f(u)} - \left[\left(\frac{f'(z)}{f(z)} - \frac{f'(1)}{f(1)} \right) + \left(\frac{f'(z+1)}{f(z+1)} - \frac{f'(2)}{f(2)} \right) + \dots \right],$$

where

$$\gamma_{f(u)} = \lim_{p \rightarrow \infty} \left[\frac{f'(1)}{f(1)} + \frac{f'(2)}{f(2)} + \dots + \frac{f'(p)}{f(p)} - \log f(p) \right] \quad (3)$$

convergence of $\gamma_{f(u)}$ being easily established. Clearly

$$\Gamma'_{f(u)}(1) = -\gamma_{f(u)}.$$

The analogy with $\Gamma(z)$ is further brought out by transforming (1) into

$$\frac{1}{\Gamma_{f(u)}(z)} = f(z) e^{\gamma_{f(u)} z} \prod_{p=1}^{\infty} \left\{ \frac{f(p+z)}{f(p)} e^{-\frac{f'(p)}{f(p)} z} \right\}, \quad (4)$$

the generalization of the Weierstrassian form of $\Gamma(z)$.

From (3) we find

$$-\frac{d^2}{dz^2} \log \Gamma_{f(u)}(z) = \frac{d^2}{dz^2} \log f(z) + \frac{d^2}{dz^2} \log f(z+1) + \dots$$

More generally for r unrestricted, we have

$$-\frac{d^{r+1}}{dz^{r+1}} \log \Gamma_{f(u)}(z) = \frac{d^{r+1}}{dz^{r+1}} \log f(z) + \frac{d^{r+1}}{dz^{r+1}} \log f(z+1) + \dots \quad (5)$$

4. Asymptotic Expansions. In the notation we have adopted

$$\log |f(p)| = \log f(1) + \log f(2) + \dots + \log f(p).$$

If $\phi(z)$ is analytic for $R(z) > a$, $-1 \leq I(z) \leq 1$, then

$$\begin{aligned} \phi(1) + \phi(2) + \dots + \phi(p) &= C + \int_a^p \phi(t) dt + \frac{1}{2}\phi(p) + \frac{B_1}{2!}\phi'(p) \\ &\quad - \frac{B_2}{4!}\phi'''(p) + \dots + (-)^q \frac{B_q}{(2q)!}\phi^{(2q-1)}(p) + R_p, \end{aligned}$$

where R_p can be put in either of the forms

$$\sum_{t=1}^{2q} A_t \sum_{s=p}^{\infty} \phi^{(2q+1)}(s + t\theta_{s-p}), \quad \sum_{t=1}^{2q+1} A_t \sum_{s=q}^{\infty} \phi^{(2q+2)}(s + t\theta_{s-p}),$$

provided these forms converge. In the case where $\phi(z) = \log f(z)$, condition (a) insures the fulfilment of the condition of this formula. Let that form of the remainder be chosen which makes the index of differentiation $r+1$. Then by condition (b) a value of p may be chosen such that

for all greater values

$$|R_p| < \frac{1}{p^{1+\epsilon}} + \frac{1}{(p+1)^{1+\epsilon}} + \dots < \frac{1}{\epsilon(p-1)^\epsilon},$$

Hence $\lim_{p \rightarrow \infty} R_p = 0$. Letting $C = \log G_{f(u)}$, we have

$$\begin{aligned} \log f(p) &\sim \log G_{f(u)} + \int_a^p \log f(t) dt + \frac{1}{2} \log f(p) \\ &\quad + \frac{B_1}{2!} \frac{d}{dp} \log f(p) - \dots + (-)^{q-1} \frac{B_q}{(2q)!} \frac{d^{2q-1} \log f(p)}{dp^{2q-1}}. \end{aligned} \quad (1)$$

In general q may be taken to be larger than the value above adopted, since condition (b) will usually be satisfied for larger values of r than the one used in the infinite product.

We shall now show that if for p we substitute z in the above formula, we obtain an asymptotic expansion for $\log \Gamma_{f(u)}(z+1)$, or

$$\begin{aligned} \log \Gamma_{f(u)}(z) &\sim \log G_{f(u)} + \int_a^z \log f(t) dt - \frac{1}{2} \log f(z) \\ &\quad + \frac{B_1}{2!} \frac{d \log f(z)}{dz} - \dots + \frac{(-)^{q-1} B_q}{(2q)!} \frac{d^{2q-1} \log f(z)}{dz^{2q-1}}. \end{aligned} \quad (2)$$

Denote the right hand member by S_z . Using the recurrence formulæ for the Bernoullian numbers, we easily find, provided z is within the analytic sector,

$$S_{z+1} = S_z + \log f(z) + P_z,$$

where P_z can be written in either of the forms

$$\sum_{t=1}^{q+1} A_t \frac{d^{2q+1}}{dz^{2q+1}} \log f(z + \theta_t), \quad \sum_{t=1}^{q+2} A_t \frac{d^{2q+2}}{dz^{2q+2}} \log f(z + \theta_t).$$

Hence

$$S_{z+p} - S_z = \log [f(z)f(z+1)\dots f(z+p-1)] + \sum_{y=z}^{z+p-1} P_y. \quad (3)$$

Substituting this result, and (1) above in the infinite product for $\Gamma_{f(u)}(z)$, we have

$$\begin{aligned} \log \Gamma_{f(u)}(z) &= \lim_{p \rightarrow \infty} \left[S_p + S_z - S_{z+p} + z \log f(p) \right. \\ &\quad \left. + \log F(p, z) + R_p - \sum_{y=z}^{z+p-1} P_y \right]. \end{aligned}$$

Now

$$S_{z+p} - S_p = \int_p^{z+p} \log f(t) dt - \frac{1}{2} \log \frac{f(z+p)}{f(p)} + \dots$$

$$\begin{aligned}
& + \frac{(-)^{q-1} B_q}{(2q)!} \frac{d^{2q-1}}{dp^{2q-1}} \log \frac{f(z+p)}{f(p)} \\
& = z \log f(p) + \frac{d}{dp} \log f(p) \left[\frac{z^2}{2!} - \frac{z}{2} \right] + \frac{d^2}{dp^2} \log f(p) \left[\frac{z^3}{3!} - \frac{z^2}{2 \cdot 2!} \right. \\
& \quad \left. + \frac{B_1 z}{2!} \right] + \cdots + \frac{d^r}{dp^r} \log f(p) \left[\frac{z^{r+1}}{(r+1)!} - \frac{z^r}{2r!} + \frac{z^{r-1} B_1}{2! (r-1)!} \right. \\
& \quad \left. - \frac{z^{r-3} B_2}{4! (r-3)!} + \cdots \right] + Q_p \\
& = z \log f(p) + \log F(p, z) + Q_p,
\end{aligned}$$

where

$$\begin{aligned}
Q_p &= \frac{z^{r+2} \lambda_1}{(r+2)!} \frac{d^{r+1}}{dp^{r+1}} \log f(p + \theta_1 z) - \frac{1}{2} \frac{z^{r+1} \lambda_2}{(r+1)!} \frac{d^{r+1}}{dp^{r+1}} \log f(p + \theta_2 z) \\
&\quad + \frac{B_1 z^r \lambda_3}{2! r!} \frac{d^{r+1}}{dp^{r+1}} \log f(p + \theta_3 z) - \cdots,
\end{aligned}$$

the last term ending in z^2 or z^3 . Hence

$$\lim_{p \rightarrow \infty} Q_p = 0; \quad \lim_{p \rightarrow \infty} R_p = 0; \quad \lim_{p \rightarrow \infty} \sum_{y=s}^{z+p-1} P_y = \sum_{s=0}^{\infty} P_{z+s},$$

and

$$\log \Gamma_{f(u)}(z) = S_z - \sum_{s=0}^{\infty} P_{z+s},$$

S_z will therefore be the asymptotic expansion of $\log \Gamma_{f(u)}(z)$, provided z approaches infinity in such a way that $\sum_{s=0}^{\infty} P_{z+s} \rightarrow 0$. This will clearly be the case if $z \rightarrow \infty$ along a line parallel to or on the real axis in the positive direction. Under this condition (2) holds. We cannot infer more from the conditions we have assumed. If however condition (b) holds when $p \rightarrow \infty$ along any line,* (2) will hold for $z \rightarrow \infty$ along any line not parallel to or on the real axis in the negative direction.

5. An Integral for $\Gamma_{f(u)}(z)$. If x_1 and x_2 are integers, and $\phi(\xi)$ is a function which is analytic and bounded for all values of ξ such that

$$x_1 \leq R(\xi) \leq x_2,$$

then†

$$\begin{aligned}
& \frac{1}{2} \phi(x_1) + \phi(x_1 + 1) + \phi(x_1 + 2) + \cdots + \phi(x_2 - 1) + \frac{1}{2} \phi(x_2) \\
& = \int_{x_1}^{x_2} \phi(\xi) d\xi + \frac{1}{i} \int_0^\infty \frac{\phi(x_2 + iy) - \phi(x_1 + iy) - \phi(x_2 - iy) + \phi(x_1 - iy)}{e^{2\pi y} - 1} dy.
\end{aligned}$$

* This extended condition is satisfied by the class of functions suggested in the introduction.

† Whittaker and Watson, p. 145, Ex. 7.

Let the initial conditions (a) and (b) imposed on $f(z)$ be extended to the following:

(a') $\log f(z)$ is analytic to the right of a line at distance a from the axis of imaginaries;

(b') for some value of r and ϵ , $\epsilon > 0$,

$$\lim_{z \rightarrow \infty} z^{1+\epsilon} \frac{d^{r+1}}{dz^{r+1}} \log f(z) = 0,$$

where $z \rightarrow \infty$ along any line included in the analytic region of (a').

We shall then have, if $R(z) > a$, $x_1 = 0$, $x_2 \rightarrow \infty$,

$$\begin{aligned} -\frac{d^{r+1}}{dz^{r+1}} \log \Gamma_{f(u)}(z) &= \frac{d^{r+1}}{dz^{r+1}} \log f(z) + \frac{d^{r+1}}{dz^{r+1}} \log f(z+1) + \dots \\ &= \frac{1}{2} \frac{d^{r+1}}{dz^{r+1}} \log f(z) + \int_0^\infty \frac{d^{r+1}}{d\xi^{r+1}} \log f(z+\xi) d\xi \\ &\quad - \frac{1}{i} \int_0^\infty \frac{\frac{d^{r+1}}{dz^{r+1}} \log f(z+iy) - \frac{d^{r+1}}{dz^{r+1}} \log f(z-iy)}{e^{2\pi y} - 1} dy + \lim_{x_2 \rightarrow \infty} Rx_2, \end{aligned}$$

where

$$Rx_2 = \frac{1}{i} \int_0^\infty \frac{\frac{d^{r+1}}{dz^{r+1}} \log f(z+x_2+iy) - \frac{d^{r+1}}{dz^{r+1}} \log f(z+x_2-iy)}{e^{2\pi y} - 1} dy.$$

From (b') we easily find

$$\lim_{x_2 \rightarrow \infty} x_2^{\epsilon} Rx_2 = 0, \quad i.e., \lim_{x_2 \rightarrow \infty} Rx_2 = 0.$$

Also

$$\int_0^\infty \frac{d^{r+1}}{d\xi^{r+1}} \log f(z+\xi) d\xi = -\frac{d^r \log f(z)}{dz^r},$$

so that on integration

$$\begin{aligned} \log \Gamma_{f(u)}(z) &= c_0 + c_1 z + \dots + c_r z^r - \frac{1}{2} \log f(z) \\ &\quad + \int_a^z \log f(z) dz + \frac{1}{i} \int_0^\infty \frac{\log f(z+iy) - \log f(z-iy)}{e^{2\pi y} - 1} dy. \end{aligned}$$

If we let $z \rightarrow \infty$, on expanding the latter integral and comparing with the asymptotic expansion, we find

$$c_0 = \log G_{f(u)}; \quad c_1 = c_2 = \dots = c_r = 0,$$

and

$$\begin{aligned} \log \Gamma_{f(u)}(z) &= \log G_{f(u)} - \frac{1}{2} \log f(z) + \int_0^z \log f(z) dz \\ &\quad + 2 \int_0^\infty \frac{\log f(z+iy) - \log f(z-iy)}{2i} \frac{dy}{e^{2\pi y} - 1}. \end{aligned}$$

As a matter of notation, let

$$\frac{\phi(z+iy) - \phi(z-iy)}{2i} = \sin_{\phi(u)}(z, y),$$

$$\frac{\phi(z+iy) + \phi(z-iy)}{2} = \cos_{\phi(u)}(z, y).$$

Then

$$\log \Gamma_{f(u)}(z) = \log G_{f(u)} - \frac{1}{2} \log f(z) + \int_a^z \log f(t) dt + 2 \int_0^\infty \frac{\sin_{\log f(u)}(z, y)}{e^{2\pi y} - 1} dy \quad (1)$$

and by differentiation

$$\frac{\Gamma_{f(u)}'(z)}{\Gamma_{f(u)}(z)} = \log f(z) - \frac{1}{2} \frac{f'(z)}{f(z)} + 2 \int_0^\infty \frac{\cos_{f(u)/f(u)}(z, y)}{e^{2\pi y} - 1} dy. \quad (2)$$

PART II. TRANSFORMATIONS.

6. Integral for the Asymptotic Constant. In the present section and in the one following we shall obtain results which are very useful in establishing particular relations between the generalized Gamma functions.

Let*

$$u = \int_z^{z+1} \log \Gamma_{f(u)}(t) dt;$$

then

$$\frac{du}{dz} = \log \Gamma_{f(u)}(z+1) - \log \Gamma_{f(u)}(z) = \log f(z),$$

and

$$u = \int_a^z \log f(t) dt + C.$$

Since

$$\begin{aligned} \log \Gamma_{f(u)}(z) &\sim \log G_{f(u)} + \int_a^z \log f(t) dt - \frac{1}{2} \log f(z) \\ &\quad + \sum_{s=1}^q \frac{(-)^{s-1} B_s}{(2s)!} \frac{d^{2s-1} \log f(z)}{dz^{2s-1}}, \end{aligned}$$

it easily follows that

$$\begin{aligned} \int_z^{z+1} \log \Gamma_{f(u)}(t) dt &\sim \log G_{f(u)} + \int_z^{z+1} \left[\int_a^t \log f(u) du \right] dt \\ &\quad - \frac{1}{2} \int_z^{z+1} \log f(t) dt + \sum_{s=1}^q \frac{(-)^{s-1} B_s}{(2s)!} \frac{d^{2s-2} \log f(z+1)}{dz^{2s-2}} \log \frac{f(z+1)}{f(z)}. \end{aligned}$$

But, using the Euler-Maclaurin Sum formula,[†] we get

* For the analogue of the ordinary Gamma function see Whittaker and Watson, p. 255, Ex. 21.

† Whittaker and Watson, p. 128.

$$\int_z^{z+1} \left[\int_a^t \log f(u) du \right] dt \sim \int_a^z \log f(t) dt + \frac{1}{2} \int_z^{z+1} \log f(t) dt + \sum_{s=1}^q \frac{(-)^s B_s}{(2s)!} \frac{d^{2s-2}}{dz^{2s-2}} \log \frac{f(z+1)}{f(z)},$$

so that

$$\int_z^{z+1} \log \Gamma_{f(u)}(t) dt \sim \int_a^z \log f(t) dt + \log G_{f(u)}.$$

Comparing with the above, we see that

$$C = \log G_{f(u)}$$

and

$$\int_z^{z+1} \log \Gamma_{f(u)}(t) dt = \int_a^z \log f(t) dt + \log G_{f(u)}. \quad (1)$$

Letting $z = a$, we obtain

$$\log G_{f(u)} = \int_a^{a+1} \log \Gamma_{f(u)}(t) dt, \quad (2)$$

an analytical expression for the constant $G_{f(u)}$.

Since $G_{f(u)}$ depends on the value of a chosen, we shall write it $_a G_{f(u)}$. Then

$$\log_{a_1} G_{f(u)} = \log_{a_2} G_{f(u)} + \int_{a_2}^{a_1} \log f(t) dt. \quad (3)$$

7. The Asymptotic Test. We have seen that $\log \Gamma_{f(u)}(z)$ has the same asymptotic expansion as $\log |f(p-1)|$. Since any other solution than $\Gamma_{f(u)}(z)$ of

$$\phi(z+1) = f(z)\phi(z) \quad (1)$$

must be in the form

$$\phi(z) = \Gamma_{f(u)}(z)P(z), \quad (2)$$

where $P(z)$ is periodic of period unity, it is evident that $\Gamma_{f(u)}(z)$ is the only solution of (1) possessing this property. The following is a more useful expression of the above principle.

From

$$-\frac{d^{r+1}}{dz^{r+1}} \log \Gamma_{f(u)}(z) = \frac{d^{r+1}}{dz^{r+1}} \log f(z) + \frac{d^{r+1}}{dz^{r+1}} \log f(z+1) + \dots,$$

we have

$$\lim_{z \rightarrow +\infty} \frac{d^{r+1}}{dz^{r+1}} \log \Gamma_{f(u)}(z) = 0. \quad (3)$$

Let $\psi(z)$ be some other solution of (1) possessing property (3). Then if $z = p + x$, where p is an integer, we must have for all values of x

$$\lim_{p \rightarrow \infty} \frac{d^{r+1}}{dx^{r+1}} \log \psi(x+p) = \lim_{p \rightarrow \infty} \left[\frac{d^{r+1}}{dx^{r+1}} \log \Gamma_{f(u)}(x+p) + \frac{d^{r+1}}{dx^{r+1}} \log P(x) \right] = 0,$$

so that

$$\frac{d^{r+1}}{dx^{r+1}} \log P(x) = 0.$$

Since $P(z)$ is periodic we can only have

$$P(z) = ce^{2\pi ipz},$$

so that

$$\psi(z) = ce^{2\pi ipz} \Gamma_{f(u)}(z), \quad (4)$$

where p is an integer. If furthermore $\psi(z)$ have an asymptotic expansion of the form

$$b + a_0 \int_a^z \log f(t) dt + a_1 \log f(z) + \cdots + a_{2s} \frac{d^{2s-1}}{dz^{2s-1}} \log f(z),$$

where the a 's are independent of $f(z)$, we find by letting $f(z) = e^{z^{\rho-\epsilon}}$ and using (4) that

$$a_0 = 1, \quad a_1 = -\frac{1}{2}, \quad a_{2s+1} = 0, \quad a_{2s} = \frac{B_s}{(2s)!} (-)^{s-1},$$

and

$$\psi(z) = c \Gamma_{f(u)}(z). \quad (5)$$

We shall refer to condition (3) and the one just given as the asymptotic test. Hence if two solutions $\psi_1(z)$ and $\psi_2(z)$ of (1) satisfy the asymptotic test,

$$\psi_1(z) = c \psi_2(z).$$

The same is clearly true of solutions of

$$\phi(z+n) = f(z) \phi(z),$$

where n is real and positive, since this equation can be transformed into

$$\psi(z+1) = f(nz) \psi(z).$$

8. Elementary Transformations. From the Gaussian form of $\Gamma_{f(u)}(z)$ we see that

$$\Gamma_{[f_1(u)]^m [f_2(u)]^n}(z) = [\Gamma_{f_1(u)}(z)]^m [\Gamma_{f_2(u)}(z)]^n. \quad (1)$$

In particular

$$\Gamma_c(z) = c^{z-1}; \quad \Gamma_u(z) = \Gamma(z); \quad \Gamma_{e(w)}(z) = e^{[\phi_{r+1}(z)]/(r+1)}. *$$

Again, both $\Gamma_{f(u+b)}(z)$ and $\Gamma_{f(u)}(z+b)$ are solutions of

$$\phi(z+1) = f(z+b) \phi(z).$$

They clearly have the same asymptotic expansions, except for a constant

* By means of these results and (2), we easily evaluate $\Gamma_{f_1(u)} f_2(u)(z)$ where $f_1(u)$ is a rational and $f_2(u)$ an integral algebraic function.

factor, so that

$$\Gamma_{f(u+b)}(z) = c\Gamma_{f(u)}(z+b).$$

To determine c , let $z = 1$. Since $\Gamma_{f(u+b)}(1) = 1$,

$$c = \frac{1}{\Gamma_{f(u)}(1+b)},$$

and

$$\Gamma_{f(u+b)}(z) = \frac{\Gamma_{f(u)}(z+b)}{\Gamma_{f(u)}(1+b)}. \quad (2)$$

By means of the integral of §6 these results may be directly applied to ${}_aG_{f(u)}$. We thus find

$$\begin{aligned} {}_aG_{[f_1(u)]^m[f_2(u)]^n} &= [{}_aG_{f_1(u)}]^m[{}_aG_{f_2(u)}]^n, \\ {}_oG_c &= c^{-1/2}, \quad {}_oG_{\epsilon w} = 1, \\ {}_aG_{f(u+b)} &= \frac{{}_a+b\Gamma_{f(u)}}{\Gamma_{f(u)}(1+b)}, \end{aligned} \quad (3)$$

9. Infinite Products in Terms of Generalized Gamma Functions. Consider first

$$\prod_{p=0}^{\infty} \frac{f(a_1 + p)f(a_2 + p) \cdots f(a_k + p)}{f(b_1 + p)f(b_2 + p) \cdots f(b_l + p)} = \prod_{p=0}^{\infty} F(p). \quad (1)'$$

A sufficient condition for convergence is that for some positive value of ϵ

$$\lim_{p \rightarrow \infty} p^{1+\epsilon} \log F(p) = 0. \quad (2)'$$

But

$$\begin{aligned} \log F(p) &= \log f(a_1 + p) + \cdots - \log f(b_1 + p) - \cdots \\ &= (k - l) \log f(p) + \frac{\Sigma a - \Sigma b}{1} \frac{d}{dp} \log f(p) + \cdots \\ &\quad + \frac{\Sigma a^r - \Sigma b^r}{r!} \frac{d^r}{dp^r} \log f(p) + R_p, \end{aligned}$$

where

$$\lim_{p \rightarrow \infty} p^{1+\epsilon} R_p = 0, \quad \lim_{p \rightarrow \infty} p^{1+\epsilon} \frac{d^r}{dp^r} \log f(p) \neq 0.$$

We must therefore have

$$k = l, \quad \Sigma a = \Sigma b, \cdots \Sigma a^r = \Sigma b^r, \quad (3)'$$

if (2)' is to be fulfilled. Now

$$\prod_{p=0}^{\infty} \frac{f(a_1 + p) \cdots f(a_k + p)}{f(b_1 + p) \cdots f(b_l + p)} = \frac{\Gamma_{f(u)}(b_1) \cdots \Gamma_{f(u)}(b_k)}{\Gamma_{f(u)}(a_1) \cdots \Gamma_{f(u)}(a_k)} S_{\infty},$$

where

$$S_p = \frac{[f(p)]^{a_1} F(p, a_1) \cdots [f(p)]^{a_k} F(p, a_k)}{[f(p)]^{b_1} F(p, b_1) \cdots [f(p)]^{b_k} F(p, b_k)}.$$

Using (3)', we see that $S_p = 1$, so that

$$\prod_{p=0}^{\infty} \frac{f(a_1 + p)f(a_2 + p) \cdots f(a_k + p)}{f(b_1 + p)f(b_2 + p) \cdots f(b_k + p)} = \frac{\Gamma_{f(u)}(b_1) \cdots \Gamma_{f(u)}(b_k)}{\Gamma_{f(u)}(a_1) \cdots \Gamma_{f(u)}(a_k)}. \quad (1)$$

Consider now more generally

$$\prod_{p=0}^{\infty} \frac{f_1(a_1 + p) \cdots f_k(a_k + p)}{\phi_1(b_1 + p) \cdots \phi_l(b_l + p)} = \prod_{p=0}^{\infty} F(p), \quad (4)'$$

where (2)' is again a sufficient condition. We may write

$$\begin{aligned} \prod_{p=0}^{\infty} F(p) &= \lim_{q \rightarrow \infty} \prod_{p=1}^q F(p-1) = \lim_{q \rightarrow \infty} \frac{F(q-1)}{F(q)} \\ &= {}_a G_{F(u-1)} e^{\int_a^{\infty} \log F(u-1) du} \end{aligned} \quad (5)'$$

by (1 §4). Furthermore, provided the functions used exist, from §8 it follows that

$$\begin{aligned} {}_a G_{F(u-1)} &= \frac{{}_a G_{f_1(a_1-1+u)} \cdots {}_a G_{f_k(a_k-1+u)}}{{}_a G_{\phi_1(b_1-1+u)} \cdots {}_a G_{\phi_l(b_l-1+u)}} \\ &= \frac{a+a_1-1 G_{f_1(u)} \cdots a+a_k-1 G_{f_k(u)}}{a+b_1-1 G_{\phi_1(u)} \cdots a+b_l-1 G_{\phi_l(u)}} \frac{\Gamma_{\phi_1(u)}(b_1) \cdots \Gamma_{\phi_l(u)}(b_l)}{\Gamma_{f_1(u)}(a_1) \cdots \Gamma_{f_k(u)}(a_k)}, \end{aligned}$$

so that finally

$$\prod_{p=0}^{\infty} \frac{f_1(a_1 + p) \cdots f_k(a_k + p)}{\phi_1(b_1 + p) \cdots \phi_l(b_l + p)} = A \frac{{}_a G_{f_1(u)} \cdots {}_a G_{f_k(u)}}{{}_a G_{\phi_1(u)} \cdots {}_a G_{\phi_l(u)}} \frac{\Gamma_{\phi_1(u)}(b_1) \cdots \Gamma_{\phi_l(u)}(b_l)}{\Gamma_{f_1(u)}(a_1) \cdots \Gamma_{f_k(u)}(a_k)},$$

where

$$\begin{aligned} \log A &= \int_a^{\infty} \log F(u-1) du + \sum_{\lambda=1}^k \int_a^{a+a_{\lambda}-1} \log f_{\lambda}(u) du \\ &\quad - \sum_{\mu=1}^l \int_a^{a+b_{\mu}-1} \log \phi_{\mu}(u) du. \end{aligned} \quad (2)$$

10. The Multiplication Theorem. Both $\Gamma_{f(u/n)}(nz)$ and

$$\Gamma_{f(u)}(z) \Gamma_{f(u)}\left(z + \frac{1}{n}\right) \cdots \Gamma_{f(u)}\left(z + \frac{n-1}{n}\right)$$

are solutions of

$$\phi(z+1) = f(z)f\left(z + \frac{1}{n}\right) \cdots f\left(z + \frac{n-1}{n}\right)\phi(z).$$

Furthermore it is easily shown that they both satisfy the asymptotic test

since $\Gamma_{f(u)}(z)$ does. Hence

$$\Gamma_{f(u)}(z)\Gamma_{f(u)}\left(z + \frac{1}{n}\right) \cdots \Gamma_{f(u)}\left(z + \frac{n-1}{n}\right) = \psi(n)\Gamma_{f(u/n)}(nz).$$

To determine $\psi(n)$, we have

$$\begin{aligned} \log {}_{na}G_{f(u/u)} &= \int_{na}^{na+1} \log \Gamma_{f(u/u)}(z) dz = n \int_a^{a+(1/n)} \log \Gamma_{f(u/u)}(nz) dz \\ &= n \int_a^{a+(1/n)} \left[\log \Gamma_{f(u)}(z) + \cdots + \log \Gamma_{f(u)}\left(z + \frac{n-1}{n}\right) - \log \psi(n) \right] dz \\ &= n \left[\int_a^{a+(1/n)} \log \Gamma_{f(u)}(z) dz + \int_{a+(1/n)}^{a+(2/n)} \log \Gamma_{f(u)}(z) dz + \cdots \right. \\ &\quad \left. + \int_{a+[(n-1)/n]}^{a+1} \log \Gamma_{f(u)}(z) dz \right] - \log \psi(n) \\ &= n \int_a^{a+1} \log \Gamma_{f(u)}(z) dz - \log \psi(n), \end{aligned}$$

so that

$$\psi(n) = \frac{[{}_aG_{f(u)}]^n}{{}_{na}G_{f(u/u)}},$$

and

$$\Gamma_{f(u)}(z)\Gamma_{f(u)}\left(z + \frac{1}{n}\right) \cdots \Gamma_{f(u)}\left(z + \frac{n-1}{n}\right) = \frac{[{}_aG_{f(u)}]^n}{{}_{na}G_{f(u/u)}} \Gamma_{f(u/n)}(nz) \quad (1)$$

When $f(u) = u$, by using the results of §8, the ordinary multiplication theorem results.

Let $z = 1/n$, and we obtain

$$\Gamma_{f(u)}\left(\frac{1}{n}\right)\Gamma_{f(u)}\left(\frac{2}{n}\right) \cdots \Gamma_{f(u)}\left(\frac{n-1}{n}\right) = \frac{[{}_aG_{f(u)}]^n}{{}_{na}G_{f(u/n)}}. \quad (2)$$

If in this $n = 2$, we obtain

$$\Gamma_{f(u)}\left(\frac{1}{2}\right) = \frac{[{}_aG_{f(u)}]^2}{{}_{2a}G_{f(u/2)}}, \quad (3)$$

the analogue of $\Gamma(1/2) = \sqrt{\pi}$.

11. An Integration Theorem Generalized. When n is a positive integer

$$\begin{aligned} \int_a^{a+1} \log f(z) dz &= \int_a^{a+(1/n)} \log f(z) dz + \int_{a+(1/n)}^{a+(2/n)} \log f(z) dz + \cdots \\ &\quad + \int_{a+[(n-1)/n]}^{a+1} \log f(z) dz. \end{aligned}$$

By means of the Generalized Gamma functions the corresponding relation

may be obtained for any positive real value of n .*

$$\begin{aligned} \int_a^{a+(1/n)} \log f(z) dz + \cdots + \int_{a+(n-1)/n}^{a+1} \log f(z) dz \\ = \int_a^{a+(1/n)} \log \left[f(z) f\left(z + \frac{1}{n}\right) \cdots f\left(z + \frac{n-1}{n}\right) \right] dz \\ = \int_a^{a+(1/n)} \log [f(z) \Gamma_{f[z+(u/n)]}(n)] dz. \end{aligned}$$

Let now n have any positive real value. Then

$$\begin{aligned} \int_a^{a+(1/n)} \log [f(z) \Gamma_{f[z+(u/n)]}(n)] dz &= \int_a^{a+(1/n)} \log \frac{\Gamma_{f[u/n]}(n+nz)}{\Gamma_{f[u/n]}(nz)} dz \dagger \\ &= \frac{1}{n} \left[\int_{na+n}^{na+n+1} \log \Gamma_{f[u/n]}(z) dz - \int_{na}^{na+1} \log \Gamma_{f[u/n]}(z) dz \right], \end{aligned}$$

so that, using §6, we find the relation desired

$$\int_a^{a+(1/n)} \log [f(z) \Gamma_{f[z+(u/n)]}(n)] dz = \int_a^{a+1} \log f(z) dz. \quad (1)$$

12. The Multiplication Theorem Generalized. If condition (b) be extended so that for some value of r

$$\lim_{p \rightarrow \infty} p^{2+\epsilon} \frac{d^{r+2} \log f(z+p)}{dz^{r+2}} = 0,$$

the multiplication theorem may be generalized to admit all positive real values of n .‡ We have for positive integral values of n

$$\Gamma_{f(u)}(z) \Gamma_{f(u)}\left(z + \frac{1}{n}\right) \cdots \Gamma_{f(u)}\left(z + \frac{n-1}{n}\right) = \Gamma_{f(u)}(z) \Gamma_{\Gamma_{f(u)}(z+(v/n))}(n).$$

By the condition above imposed we have

$$\lim_{p \rightarrow \infty} p^{1+\epsilon} \frac{d^{r+2}}{dz^{r+2}} \log \Gamma_{f(u)}\left(z + \frac{v+p}{n}\right) = 0.$$

It therefore follows easily that $\Gamma_{\Gamma_{f(u)}(z+(v/n))}(n)$ exists when n is any real and positive number. Furthermore, with the aid of (2 §8)

$$\Gamma_{f(u)}(z) \Gamma_{\Gamma_{f(u)}(z+(v/n))}(n)$$

* The theorem depends only on the existence of the functions involved. That n be real and positive is sufficient for this purpose, under conditions (a) and (b), but not always necessary. For the class of functions suggested in the introduction the theorem holds for all values of n .

† By (2 §8) and (1 §1).

‡ The note to §11 applies here too except that n may not be a negative real number.

is seen to satisfy the equation

$$\phi\left(z + \frac{1}{n}\right) = f(z)\phi(z).$$

Since after transformation by (2 §8) it may be shown to satisfy the asymptotic test, we obtain

$$\Gamma_{f(u)}(z)\Gamma_{\Gamma_{f(u)}(z+(v/n))}(n) = \psi(n)\Gamma_{f(u/n)}(nz).$$

The previous section enables us to determine $\psi(n)$ as in the simpler case.

$$\begin{aligned} \log {}_{na}G_{f(u/n)} &= n \int_a^{a+(1/n)} \log \Gamma_{f(u/n)}(nz) dz \\ &= n \int_a^{a+(1/n)} [\log \Gamma_{f(u)}(z)\Gamma_{\Gamma_{f(u)}(z+(v/n))}(n)] dz - \log \psi(n) \\ &= n \int_a^{a+1} \log \Gamma_{f(u)}(z) dz - \log \psi(n). \end{aligned}$$

Hence, as before

$$\psi(n) = \frac{[{}_aG_{f(u)}]^n}{{}_{na}G_{f(u/n)}},$$

and

$$\Gamma_{f(u)}(z)\Gamma_{\Gamma_{f(u)}(z+(v/n))}(n) = \frac{[{}_aG_{f(u)}]^n}{{}_{na}G_{f(u/n)}} \Gamma_{f(u/n)}(nz) \quad (1)$$

Letting $z = 1/n$, and using (2 §8), we obtain

$$\Gamma_{\Gamma_{f(u)}(v/n)}(n) = \frac{[{}_aG_{f(u)}]^n}{{}_{na}G_{f(u/n)}}. \quad (2)$$

13. The Associated Periodic Functions. Let $f(z)$ be such that both $\Gamma_{f(u)}(z)$ and $\Gamma_{f(-u)}(z)$ exist, and let

$$\Gamma_{f(u)}(z)\Gamma_{f(-u)}(1-z) = F(z),$$

Then

$$F(z+1) = F(z).$$

We shall denote this periodic function by $P_{f(u)}(z)$, i. e.,

$$\Gamma_{f(u)}(z)\Gamma_{f(-u)}(1-z) = P_{f(u)}(z). \quad (1)$$

Let ${}_aG_{f(u)-a}G_{f(-u)} = {}_a\pi_{f(u)}$. Then using the integral of §6, we easily obtain

$$\log {}_a\pi_{f(u)} = \int_a^{a+1} \log P_{f(u)}(z) dz. \quad (2)$$

Clearly

$$P_{f(u)}(0) = \frac{1}{f(0)}, \quad \text{or} \quad \lim_{z \rightarrow 0} P_{f(u)}(z)f(z) = 1,$$

$$P_{f(u)}\left(\frac{1}{2}\right) = \Gamma_{f(u)}\left(\frac{1}{2}\right)\Gamma_{f(-u)}\left(\frac{1}{2}\right) = \frac{[a\pi_{f(u)}]^2}{2a\pi_{f(u/2)}}. \quad (3)$$

Again

$$P_{f(u)}(z) = P_{f(-u)}(-z), \quad (4)$$

$$P_{f(u+b)}(z) = \frac{1}{f(b)} \frac{P_{f(u)}(z+b)}{P_{f(u)}(1+b)}, \quad (5)$$

The multiplication theorem gives

$$\begin{aligned} \Gamma_{f(u)}(z)\Gamma_{f(u)}\left(z + \frac{1}{n}\right) \cdots \Gamma_{f(u)}\left(z + \frac{n-1}{n}\right) &= \frac{[aG_{f(u)}]^n}{naG_{f(u/n)}} \Gamma_{f(u/n)}(nz), \\ \Gamma_{f(-u)}\left(\frac{1}{n} - z\right)\Gamma_{f(-u)}\left(\frac{2}{n} - z\right) \cdots \Gamma_{f(-u)}(1-z) &= \frac{[-aG_{f(-u)}]^n}{-naG_{f(-u/n)}} \Gamma_{f(-u/n)}(1-nz). \end{aligned}$$

Inverting and multiplying,* we finally obtain

$$P_{f(u)}(z)P_{f(u)}\left(z + \frac{1}{n}\right) \cdots P_{f(u)}\left(z + \frac{n-1}{n}\right) = \frac{[a\pi_{f(u)}]^n}{na\pi_{f(u/n)}} P_{f(u/n)}(nz). \quad (6)$$

COLUMBIA UNIVERSITY.

* It will be noticed that whereas in the case of the ordinary Γ function it is usual to obtain the multiplication theorem from that of the sine function, the reverse method is used here.

Discussion of Problem 433

In the April number of the Monthly appeared two solutions of the fractional differential equation

$$\frac{d^{1/2}y}{dx^{1/2}} = \frac{y}{x}. \quad (1)$$

The first of these, $y = Cx^{-1/2}e^{-1/x}$, was obtained by reducing the given equation to an ordinary differential equation of the first order; while the second, $y = A_0(1 - i\sqrt{\pi}x^{-1/2} - 2x^{-1} + i\sqrt{\pi}x^{-3/2} + \dots)$ was found by equating coefficients in an assumed expansion in series. Now each of the two methods used would seem to indicate that only one solution was possible, i.e., the solution found by the corresponding method, yet the two solutions are clearly irreducible. A discussion of this difficulty might serve as a beginning of that more thorough discussion of the subject which the proposer of the problem desires.

Now first of all what does $d^{1/2}y/dx^{1/2}$ mean? More generally, what does $d^\mu y/dx^\mu$ mean where μ is any number? Before 1860, certainly, the method followed in answering that question was that due to Liouville. Starting from the known fact that $(d^\mu/dx^\mu)e^{ax} = a^\mu e^{ax}$, when μ was a positive integer, and also when μ was a negative integer, if by $d^{-p}y/dx^{-p}$ we mean the p th indefinite integral of y , he then assumed the relation to hold for all values of μ , and proceeded from that to obtain everything else. The most important formula that thus resulted was

$$\frac{d^\mu x^n}{dx^\mu} = (-1)^\mu \frac{\Gamma(-n + \mu)}{\Gamma(-n)} x^{n-\mu}. \quad (2)$$

It was this formula that the proposer of the problem used when he solved it by expansion in a series of powers of x .

But this method was too narrow, when compared with the general and rigorous methods of analysis then in use, to last. It was Riemann who first gave a definition of $d^\mu y/dx^\mu$ essentially like those used now. He gave it in the form of a definite integral. More recent writers express it as a contour integral. One of the most significant features of this newer development is the introduction of "differentiation" thereby making fractional derivatives more like definite integrals than ordinary derivatives, while the ordinary derivative appears as a peculiar (though singularly important!) degeneration. The analogy of the general binomial expansion expressed as an infinite series, with its particular terminating form for a positive integral power is complete.

A useful form applicable only when the real part of the index of differentiation is negative is

$$\left\{ \frac{d^{-\mu}}{dx^{-\mu}} \right\}_{x_0}^X f(x) = \frac{1}{\Gamma(\mu)} \int_{x_0}^X (X-x)^{\mu-1} f(x) dx. \quad (3)$$

It is to be noticed that when $x_0 = 0$, it becomes Riemann's form, while for $x_0 = -\infty$, it gives all the consistent results obtained by Liouville's method. As it stands it is more general than either.

The theorem that justifies the definition is that

$$\left\{ \frac{d^{\mu_1}}{dx^{\mu_1}} \right\}_{x_0}^X \left\{ \frac{d^{\mu_2}}{dy^{\mu_2}} \right\}_{x_0}^X f(y) = \left\{ \frac{d^{\mu_1+\mu_2}}{dx^{\mu_1+\mu_2}} \right\}_{x_0}^X f(x), \quad (4)$$

where for positive integral indices we get the ordinary derivatives. The generalization of Leibnitz's theorem also follows, i.e.,

$$\begin{aligned} \left\{ \frac{d^\mu}{dx^\mu} \right\}_{x_0}^X u \cdot v &= u \left\{ \frac{d\mu}{dx^\mu} \right\}_{x_0}^X v + \mu \frac{d}{dx} \left\{ \frac{d^{\mu-1}}{dx^{\mu-1}} \right\}_{x_0}^X v \\ &\quad + \frac{m(\mu-1)}{2^3} \frac{d^2u}{dx^2} \left\{ \frac{d^{\mu-2}}{dx^{\mu-2}} \right\}_{x_0}^X v + \dots \end{aligned} \quad (5)$$

We are now in a position to reconcile the two solutions of equation (1). In reducing (1) to an equation of the first order, we used equation (5), and then equation (4) in the form

$$\frac{d^{1/2}}{dx^{1/2}} \frac{d^{1/2}y}{dx^{1/2}} = \frac{dy}{dx},$$

and

$$\frac{d^{-1/2}}{dx^{-1/2}} \frac{d^{1/2}y}{dx^{1/2}} = y.$$

Now although this procedure is valid when the lower limit, x_0 is finite, it is not valid when $x_0 = -\infty$, for we then have

$$\frac{d^{-1/2}}{dx^{-1/2}} \frac{d^{1/2}y}{dx^{1/2}} = y - B \quad (6)$$

where B is a constant depending on y .¹ The second solution can be now obtained in the same way as the first.

Clearing (1) of fractions, operating through by $d^{1/2}/dx^{1/2}$ and using (5) we find

$$x \frac{d^{1/2}}{dx^{1/2}} \frac{d^{1/2}y}{dx^{1/2}} + \frac{1}{2} \frac{d^{-1/2}}{dx^{-1/2}} \frac{d^{1/2}y}{dx^{1/2}} = \frac{d^{1/2}y}{dx^{1/2}}$$

¹ The following explains the failure of the general theorem. From (3), when extended for all values, we find

$$\left\{ \frac{d^\mu}{dx^\mu} \right\}_{x_0}^X B = B \frac{(X-x_0)^{-\mu}}{\Gamma(1-\mu)}.$$

If x_0 is not infinite on taking the μ th derivative of this result, we get back $C_1 B$. If x_0 is infinite, the above result is zero, the $-\mu$ th derivative leaves it zero, and so the constant is lost, as (6) indicates.

Using (4), (6), and (1) this reduces to

$$x \frac{dy}{dx} + \frac{1}{2}(y - B) = \frac{y}{x}$$

whose solution is

$$y = x^{-1/2} e^{-1/x} \left[\frac{B}{2} \int x^{-1/2} e^{1/x} dx + C \right]. \quad (7)$$

Since the order of the equation solved was raised, the best we can say is that if a solution of (1) exists for $x_0 = -\infty$ it is contained in (7). As a check, we expand (7) in series obtaining

$$y = B + Cx^{-1/2} - 2Bx^{-1} - Cx^{-3/2} + \dots,$$

which, on comparison with the known solution for $x_0 = -\infty$, gives $B = A_0$; $C = -i\sqrt{\pi}A_0$. Since C is no longer arbitrary, we must change the indefinite integral of equation (7) to a definite integral. We then find,

$$y = A_0 \left[1 - i\sqrt{\pi}x^{-1/2}e^{-1/x} + x^{-1/2}e^{-1/x} \int_{-\infty}^x t^{-3/2}e^{1/t} dt \right], \quad (8)$$

which is valid for all except positive real values of x .

The derivation of the solution

$$y = Cx^{-1/2}e^{-1/x}, \quad (9)$$

which was obtained in the same manner as (7) except that $B = 0$, was not rigorous since it was not checked. If equation (1) be written

$$y = \frac{d^{-1/2}}{dx^{-1/2}} \left(\frac{y}{x} \right)$$

we can check directly by the use of (3). In fact,

$$\begin{aligned} \frac{d^{-1/2}}{dx^{-1/2}} \left(\frac{y}{x} \right) &= \frac{c}{\Gamma(\frac{1}{2})} \int_{x_0}^X (X-x)^{-1/2} x^{-3/2} e^{-1/x} dx \\ &= \frac{c}{\Gamma(\frac{1}{2})} X^{-1/2} e^{-1/X} \int_0^{(1/x_0 - 1/X)} t^{-1/2} e^{-t} dt \end{aligned}$$

which is obtained by letting $x = X/(1+tX)$.

Since $\int_0^\infty t^{-1/2} e^{-t} dt = \Gamma(\frac{1}{2})$, if we let $x_0 = 0$ we find

$$\frac{d^{-1/2}}{dx^{-1/2}} \left(\frac{y}{x} \right) = CX^{-1/2}e^{-1/X} = y$$

which checks the solution. Clearly

$$\left\{ \frac{d^{-1/2}}{dx^{-1/2}} \right\}_0^X \left(\frac{y}{x} \right)$$

exists only when the real part of X is positive. The two solutions thus supplement each other.

Suppose now that $x_0 \neq 0$, and yet is finite. Is there no solution? Clearly we always have the trivial solution $y = 0$. If however we restrict ourselves to real values of the variable and let $x_0 < 0$, the function $y = g(x)$ where $g(x) = 0$ for $x_0 \leq x \leq 0$, and $g(x) = cx^{-1/2}e^{-1/x}$ for $x > 0$ clearly satisfies the auxiliary differential equation, and also the original one; since

$$\begin{aligned}\left\{\frac{d^{-1/2}}{dx^{-1/2}}\right\}_{x_0}^x \left(\frac{g(x)}{x}\right) &= \left\{\frac{d^{-1/2}}{dx^{-1/2}}\right\}_{x_0}^x cx^{-3/2}e^{-1/x} = g(x); x > 0, \\ \left\{\frac{d^{-1/2}}{dx^{-1/2}}\right\}_{x_0}^x \left(\frac{g(x)}{x}\right) &= 0 = g(x); x \leq 0.\end{aligned}$$

However, (8) and (9) are the only analytic solutions here found. It is noteworthy that they correspond to the definition of the “generalized derivative” given by Liouville and Riemann respectively.

INTRODUCTION TO A GENERAL THEORY OF ELEMENTARY PROPOSITIONS.

BY EMIL L. POST.

INTRODUCTION.

In the general theory of logic built up by Whitehead and Russell* to furnish a basis for all mathematics there is a certain subtheory† which is unique in its simplicity and precision; and though all other portions of the work have their roots in this subtheory, it itself is completely independent of them. Whereas the complete theory requires for the enunciation of its propositions real and apparent variables, which represent both individuals and propositional functions of different kinds, and as a result necessitates the introduction of the cumbersome theory of types, this subtheory uses only real variables, and these real variables represent but one kind of entity which the authors have chosen to call elementary propositions. The most general statements are formed by merely combining these variables by means of the two primitive propositional functions of propositions Negation and Disjunction; and the entire theory is concerned with the process of asserting those combinations which it regards as true propositions, employing for this purpose a few general rules which tell how to assert new combinations from old, and a certain number of primitive assertions from which to begin.

This theory in a somewhat different form has long been the subject matter of symbolic logic.‡ However, although it had reached a high state of development as a theory of classes, it had this incurable defect as a logic of propositions, that it used informally in its proofs the very propositions whose formal statements it tried to prove. This defect appears to be entirely overcome in the development of 'Principia.' But owing to the particular purpose the authors had in view they decided not to burden their work with more than was absolutely necessary for its achievement, and so gave up the generality of outlook which characterized symbolic logic.

It is with the recovery of this generality that the first portion of our paper deals. We here wish to emphasize that the theorems of this paper

* A. N. Whitehead and B. Russell, *Principia Mathematica*, Vol. 1, 1910; Vol. 2, 1912; Vol. 3, 1913. Camb. Univ. Press.

† *Ibid.*, Vol. 1, part 1, section A.

‡ See C. I. Lewis, "A Survey of Symbolic Logic," University of California Press, 1918. An extensive bibliography is given there.

are *about* the logic of propositions but are *not included* therein. More particularly, whereas the propositions of 'Principia' are *particular* assertions introduced for their interest and usefulness in later portions of the work, those of the present paper are about the set of *all* such possible assertions. Our most important theorem gives a uniform method for testing the truth of any proposition of the system; and by means of this theorem it becomes possible to exhibit certain general relations which exist between these propositions. These relations definitely show that the postulates of 'Principia' are capable of developing the complete system of the logic of propositions without ever introducing results extraneous to that system—a conclusion that could hardly have been arrived at by the particular processes used in that work.

Further development suggests itself in two directions. On the one hand this general procedure might be extended to other portions of 'Principia,' and we hope at some future time to present the beginning of such an attempt. On the other hand we might take cognizance of the fact that the system of 'Principia' is but one particular development of the theory—particular in the primitive functions it employs and in the postulates it imposes on those functions—and so might construct a general theory of such developments. This we have tried to do in the other portions of the paper. Our first generalization leads to systems which are essentially equivalent to that of 'Principia' and connects up with the work of Sheffer* and Nicod† in reducing the number of primitive functions and of primitive propositions respectively. The second generalization, on the other hand, while including the first also seems to introduce essentially new systems. One class of such systems, and we study these in detail, seems to have the same relation to ordinary logic that geometry in a space of an arbitrary number of dimensions has to the geometry of Euclid. Whether these "non-Aristotelian" logics and the general development which includes them will have a direct application we do not know; but we believe that inasmuch as the theory of elementary propositions is at the base of the complete system of 'Principia,' this broadened outlook upon the theory will serve to prepare us for a similar analysis of that complete system, and so ultimately of mathematics.

Finally a word must be said about the viewpoint that is adopted in this paper and the method that is used. We have consistently regarded the system of 'Principia' and the generalizations thereof as purely *formal de-*

* H. M. Sheffer, "A Set of Five Independent Postulates for Boolean Algebras, with Applications to Logical Constants," *Trans. Amer. Math. Soc.*, 14 (1913), pp. 481-88.

† J. G. P. Nicod, "A Reduction in the Number of the Primitive Propositions of Logic," *Proc. Camb. Phil. Soc.*, Vol. XIX, Jan., 1917.

*developments,** and so have used whatever instruments of logic or mathematics we found useful for a study of these developments. The fact that one of the interpretations of the system of ‘Principia’ is part of the informal logic we have used in this study makes the full significance of this *interpretation*, at least with regard to proofs of consistency, uncertain, but it in no way affects the actual content of the paper which is in connection with the *formal systems*.

I welcome this opportunity to thank Prof. Keyser for the many suggestions he has offered in connection with this paper, as well as for the labor he assumed in reading and correcting it.

THE SYSTEM OF PRINCIPIA MATHEMATICA.

1. Description of the System.—Let $p, p_1, p_2, \dots, q, q_1, q_2, \dots, r, r_1, r_2, \dots$ arbitrarily represent the variable elementary propositions mentioned in the introduction. Then by means of the two primitive functions $\sim p$ (read not p —the function of Negation) and $p \vee q$ (p or q —the function of Disjunction) with the aid of the primitive propositions

- I. If p is an elementary proposition $\sim p$ is an elementary proposition,
 If p and q are elementary propositions $p \vee q$ is an elementary proposi-
 tion,

we combine these variables to form the various propositions or rather ambiguous values of propositional functions of the system. It is desirable in what follows to have before us the vision of the totality of these functions streaming out from the unmodified variable p through forms of ever growing complexity to form the infinite triangular array

$$\begin{array}{c}
 p \\
 p \vee p, \quad p_1 \vee p_2, \quad \sim p \\
 p \vee \sim p, \quad \dots, \quad \sim p_1 \vee \sim p_2, \quad \dots, \quad (p_1 \vee p_2) \vee (p_3 \vee p_4), \\
 \sim (p_1 \vee p_2), \quad \sim (p \vee p), \quad \sim \sim p
 \end{array}$$

and to note and remember that this array of functions formed merely through combining p 's by \sim 's and \vee 's constitutes the entire set of enunciations it is possible to make in the theory of elementary propositions of 'Principia.'

But the actual theory is concerned with the assertion of a certain subset of these functions. We denote the assertion of a function by writing \vdash before it. Then the motive power for the resulting process of deduction is furnished by the two rules of operation:

* For a general statement of this viewpoint see C. I. Lewis, *Loc. Cit.*, Chapter VI, section III.

II. The assertion of a function involving a variable p produces the assertion of any function found from the given one by substituting for p any other variable q , or $\sim q$, or $(q \vee r)$.*

III. " $\vdash P$ " and " $\vdash : \sim P \cdot \vee . Q$ " produce " $\vdash Q$."

These enable us to assert new functions from old, or rather in the form in which we have put them, generate new assertions from old. And the complete set of assertions is produced by applying II and III both to the following assertions which give us the start, and to all derived assertions that may result:

- IV. $\vdash : \sim (p \vee p) \cdot \vee . p$, $\vdash : \sim [p \vee (q \vee r)] \cdot \vee . q \vee (p \vee r)$,
 $\vdash : \sim q \cdot \vee . p \vee q$, $\vdash : \sim (\sim q \vee r) \cdot \vee : \sim (p \vee q) \cdot \vee . p \vee r$,
 $\vdash : \sim (p \vee q) \cdot \vee . q \vee p$.

We here again point out what was emphasized in the introduction that this theory concerns itself exclusively with the production of particular assertions through the detailed use of the rules of operation upon the primitive assertions, and as a consequence the set of theorems of this portion of 'Principia' consists of the assertions of a certain number of particular functions of the above infinite set.†

2. Truth-Table Development.‡—Let us denote the truth-value of any proposition p by + if it is true and by - if it is false. This meaning of + and - is convenient to bear in mind as a guide to thought, but in the actual development that follows they are to be considered merely as symbols which we manipulate in a certain way. Then if we attach these two primitive truth-tables to \sim and \vee

* This operation is not explicitly stated in 'Principia' but is pointed out to be necessary by B. Russell in his "Introduction to Mathematical Philosophy," London, 1919, p. 151. Its particular form was suggested to us by the first portion of the operation of "Substitution" given by Lewis, *loc. cit.*, p. 295. It will be noticed that the effect of II is to enable us to substitute any function of the system for a variable of an asserted function.

† We have consistently ignored the idea of definition in this description. We here rigorously follow the authors in saying that definition is a convenience but not a necessity and so need not be considered part of the theoretical development. And so although we too shall at times use its shorthand, we do not encumber our theoretical survey with it.

‡ Truth-values, truth-functions and our primitive truth-tables are described in 'Principia,' Vol. 1, p. 8 and p. 120, but the general notion of truth-table is not introduced. This notion is quite precise with Jevons and Venn (see Lewis, *loc. cit.*, p. 74 and pp. 175 et seq. respectively) and has its foundation in the formula for the expansion of logical functions first given by Boole. (G. Boole, "An Investigation of the Laws of Thought," London, Walton, 1854, especially pp. 72-76.) For the relation to Schröder see the footnote to section 3.

p	$\sim p$	p, q	$p \vee q$
+	-	++	+
-	+	+-	+
		-+	+
		--	-

we have a means of calculating the truth-values of $\sim p$ and $p \vee q$ from those of their arguments. Now consider any function $f(p_1, p_2, \dots, p_n)$ in our system of functions, which we will designate by F . Then since f is built up of combinations of \sim 's and \vee 's, if we assign any particular set of truth-values to the p 's, successive application of the above two primitive tables will enable us to calculate the corresponding truth-value of f . So corresponding to each of the 2^n possible truth-configurations of the p 's a definite truth-value of f is determined. The relation thus effected we shall call the truth-table of f .

For example consider the function

$$\sim(\sim(\sim p \vee q) \vee \sim(\sim q \vee p))$$

which is the ultimate definition of the function $p \equiv q$ of Principia. We have when p is + and q is + the following truth-values of the successive components of the function and so finally of the function:

$$\begin{aligned} p : +, \quad \sim p : -, \quad \sim p \vee q : +, \quad \sim(\sim p \vee q) : - \\ q : +, \quad \sim q : -, \quad \sim q \vee p : +, \quad \sim(\sim q \vee p) : - \\ \sim(\sim p \vee q) \vee \sim(\sim q \vee p) : -, \quad \sim(\sim(\sim p \vee q) \vee \sim(\sim q \vee p)) : + \end{aligned}$$

the successive truth-values being found by direct application of the primitive tables. In the same way the truth-values for $p +, q -$ etc. can be calculated and so we finally get the truth-table of $p \equiv q$, i.e.,

p, q	$p \equiv q$
++	+
+-	-
-+	-
--	+

It is needless to say that in actual work this amount of detail is quite unnecessary.

We shall call the number of variables which appear in a function the order of that function as well as that of its truth-table. It is evident that there are 2^{2^n} tables of the n th order. We now prove the

THEOREM. *To every truth-table of whatever order there corresponds at least one function of F which has it for its truth-table.*

For first corresponding to the four tables of the first order $\pm|\pm, \pm|\mp, \pm|\mp, \mp|\mp$ we have the functions $p \vee p, p \vee \sim p, \sim(p \vee \sim p), \sim p$. Now assume there is a function for each m th order table. Then in any table of order $m+1$ the configurations for which p_{m+1} is + constitute an m th order table for which there is some function $f_1(p_1, p_2, \dots, p_m)$. Likewise corresponding to $p_{m+1} = -$ we obtain $f_2(p_1, p_2, \dots, p_m)$. Let $p \cdot q$ stand for $\sim(\sim p \vee \sim q)$ a function which has the truth-table

p, q	$p \cdot q$
++	+
+-	-
-+	-
--	-

Then it easily follows that the function

$$p_{m+1} \cdot f_1(p_1, p_2, \dots, p_m) \cdot \vee \cdot \sim p_{m+1} \cdot f_2(p_1, p_2, \dots, p_m)$$

has for its truth-table the given $m+1$ st order table.

The functions of F can then be classified according to their tables as follows: those which have all their truth-values +, all -, or some + and some -. We shall call these functions respectively positive, negative, and mixed. This classification is of great importance in connection with the process of substitution which is so fundamental in the postulational development. We shall say that any function obtained from another by the process of substitution is contained in that function. We then have the

THEOREM. *Every function contained in a positive function is positive; every function contained in a negative function is negative; every mixed function contains at least one function for every possible truth-table.*

The first two results are immediate. In the third case note that any mixed function $f(p_1, p_2, \dots, p_n)$ has at least one configuration which yields + and one which yields -. Let the truth-value of p_i in the positive configuration be denoted by t_i and in the negative by t'_i , and construct a function $\phi_i(p)$ with the truth-table

p	$\phi_i(p)$
+	t_i
-	t'_i

Then $\psi(p) = f(\phi_1(p), \phi_2(p), \dots, \phi_n(p))$ will be + when p is + and - when p is -. But by our first theorem there is at least one function $g(q_1, q_2, \dots, q_m)$ corresponding to any table of order m . Hence $\psi[g(q_1, q_2, \dots, q_m)]$ is a function contained in $f(p_1, p_2, \dots, p_n)$ corresponding to that table.

COROLLARY. *Every mixed function contains at least one positive function and one negative function.*

3. The Fundamental Theorem.*—A necessary and sufficient condition that a function of F be asserted as a result of the postulates II, III, IV is that all its truth-values be +.

Note first that each of the primitive assertions of IV is a positive function. Furthermore from the assertion of positive functions we can only get positive functions. For the only method we have of producing new assertions from old is through the use of II and III. Now II can only produce positive functions since every function contained in a positive function is positive. As for III, if P is + and Q is -, $\sim P \vee Q$ is -, so that so long as P is a positive function and $\sim P \vee Q$ is a positive function Q must be positive, so that III can only produce positive functions. Hence every asserted function is positive and we have proved the condition necessary.

In order to prove it also sufficient we give a method for deriving the assertion of any positive function. It will simplify the exposition to introduce the other two defined functions of 'Principia' besides $p \cdot q$ (p and q) given above, viz.,

$$p \supset q . = . \sim p \vee q \quad Df†; \quad p \equiv q . = . p \supset q \cdot q \supset p \quad Df$$

read "p implies q" and "p is equivalent to q" respectively, and having the tables

p, q	$p \supset q$	p, q	$p \equiv q$
++	+	++	+
+-	-	+-	-
-+	+	-+	-
--	+	--	+

* The method for testing propositions embodied in this theorem is essentially the same as that given by Schröder for the logical system he has developed. (Ernst Schröder, Vorlesungen über die Algebra Der Logik, Leipzig, Teubner; 2. Bd. 1. Abth, 1891; §32.) But we believe the range of significance of the proof we have given to be quite different from that of the work of Schröder. For first, as has been emphasized by Lewis (*Loc. cit.*, Chap. IV), formal and informal logic are inextricably bound together in Schröder's development to an extent that prevents the system as a whole from being completely determined. As a result the necessity of the condition of the theorem, which evidently requires such a complete determination if it is to be proved, remains unproved. As for the sufficiency, parts *B* and *C* of our proof appear in the proof for the expression of functions given by Schröder. (1. Bd, 1890). Part *A*, however, seems not to have been given explicitly, while corresponding to part *D* are all the theoretical difficulties met with in passing from the theory of classes to that of propositions when the development is not strictly formal. Hence the sufficiency of the condition is only incompletely proved. The theorem as given by Schröder is therefore of only partial significance even in his own system; and when transplanted to the system of *Principia* requires independent proof. Finally we may mention that the applications we have made of the theorem depend for their significance on those parts of the proof which do not appear, and could not appear, in Schröder.

† III can now be written " $\vdash P$ " and " $\vdash P \supset Q$ " produce " $\vdash Q$ ".

It will be noticed that if we have " $\vdash f_1(p_1, \dots, p_n) \equiv f_2(p_1, \dots, p_n)$ " this asserted equivalence must have a positive table by the first part of our theorem, and so f_1 and f_2 must have the same truth-values for the same configurations, i.e., they must have the same truth-table.

The proof is most conveniently given in four stages.

A. We prove the theorem $p \equiv q . \triangleright . f(p) \equiv f(q)$ where the function f may involve other arguments besides the one indicated and need not involve that. By means of this theorem we shall be able to replace a constituent of a given function by any equivalent function, and have the result equivalent to the given function.

It becomes necessary for the first time to introduce the notion of the rank of a function which we define inductively as follows: the unmodified variable p will be said to be of rank zero, the negative of a function of rank m will be of rank $m + 1$; the logical sum of two functions the rank of one of which equals and the other does not exceed m will be of rank $m + 1$. Each function of F then is of finite rank as well as of finite order.* Returning now to the theorem we notice that it is true for a function of rank zero since it reduces either to $p \equiv q . \triangleright . p \equiv q$ which follows from $p \triangleright p \dagger$ by II, or to $p \equiv q . \triangleright . r \equiv r$ which follows from $p \triangleright q \triangleright p, r \equiv r$, III and II. Assume now that the theorem holds for functions of rank m and lower. Then it also holds for functions of rank $m + 1$. For if f is of rank $m + 1$ it can be written in the form $\sim f_1(p)$, or, $f_2(p) \vee f_3(p)$ where f_1, f_2 and f_3 are at most of rank m ; and then the theorem follows by using $p \equiv q . \triangleright . \sim p \equiv \sim q, p \equiv q . \triangleright : . r \equiv s : \triangleright : p \vee r . \equiv . q \vee s$ along with $p \triangleright q : \triangleright : q \triangleright r . \triangleright . p \triangleright r$, III and II.

B. Consider now any function $f(p_1, p_2, \dots, p_n)$. Using $\sim (p \vee q) \equiv . \sim p . \sim q$ and $\sim . \sim p \equiv p$ with the aid of the equivalence theorem of A and $p \equiv q : \triangleright : q \equiv r . \triangleright . p \equiv r$ we finally obtain $f(p_1, p_2, \dots, p_n)$ equivalent to a function $f'(p_1, p_2, \dots, p_n)$ which is expressed merely through combinations of p 's and $\sim p$'s by \cdot 's and \vee 's.

C.† If we then apply the distributive law of logical multiplication to f' , it will be reduced to an equivalent function consisting of successive logical sums of successive logical products of the p 's and $\sim p$'s. If any of these products has neither p_n nor $\sim p_n$ as a factor we can introduce them through the propositions $p \vee \sim p$, and $p : \triangleright : q \equiv . p . q$, whence $q : \equiv : (p \vee \sim p) . q : \equiv : p . q . \vee . \sim p . q$. Now apply the commutative and associative laws

* But whereas the number of functions of given order is infinite those of given rank are finite.

† This as well as all other particular assertions that we use without an indication of proof appear in *Principia*, Vol. I, Part A.

‡ This portion of the proof is essentially that given by A. N. Whitehead in his "Universal Algebra," p. 46. Camb. Univ. Press, 1898.

of logical multiplication along with $p \cdot p \equiv p$ so that each product has at most one p_i and one $\sim p_i$. Again using the distributive law for purposes of factorization along with the commutative and associative laws of addition we finally obtain f equivalent to

$$f_1(p_1, p_2, \dots, p_{n-1}) \cdot p_n \cdot \sim p_n : \vee : f_2(p_1, \dots, p_{n-1}) \cdot p_n \cdot \vee \cdot f_3(p_1, \dots, p_{n-1}) \cdot \sim p_n$$

where one or more of the terms and arguments may not appear.

D. Suppose now that the original function is positive; then this equivalent function will be positive. If in particular it be of first order, it can only be $p \vee \sim p$ or $p \cdot \sim p$, $\vee \cdot p \vee \sim p$. The first is an asserted function; likewise the second through $p \supset q \vee p$. Hence also $f(p)$ will be asserted through $p \equiv q \supset q \supset p$; and so every positive first order function is asserted. Assume now that this is true for all m th and lower ordered functions and let f be any positive $(m+1)$ st order function. The reduced function being then positive, both f_2 and f_3 will be positive, and hence will be asserted. From the use of $p : \supset : q \supset p \equiv q$, $p \cdot r \cdot \vee \cdot p \cdot \sim r : \equiv : p \cdot (r \vee \sim r)$, $p : \supset : S \supset p \cdot S$, and $p \supset q \vee p$, the reduced function will be asserted and so finally f . Hence every positive function can be asserted and so the proof is complete.

We thus see that given any function the theorem gives a direct method for testing whether that function can or cannot be asserted; and if the test shows that the function can be asserted the above proof will give us an actual method for immediately writing down a formal derivation of its assertion by means of the postulates of 'Principia'.

Before we pass on to theorems about the system itself irrespective of truth-tables we give the following definitions which apply directly to the system: a true function is one that can be asserted as a result of the postulates, any other is false; a completely false function is a false function such that every function therein contained is false—otherwise we call it incompletely false. We then have the

COROLLARY. *The set of true, completely false, and incompletely false functions is identical with the set of positive, negative, and mixed functions respectively.*

4. Consequences of the Fundamental Theorem.—In the above development the truth-values $+$, $-$ were arbitrary symbols which were found related in certain suggestive ways through the fundamental theorem. We are now in a position to give direct definitions of these truth-values in terms of the postulational development. In fact we shall define $+$ to be the set of true functions, $-$ the set of completely false functions. The truth-value of a function will then exist when and only when it is true or completely false, and it will be defined as that class $(+, -)$ of which it is a member. The content of the fundamental theorem consists now of these two theorems:

1. The truth-value of $\sim p$ and $q \vee r$ exists whenever the truth-values of p , q and r exist, and depends only on those truth-values as given by the primitive tables. It therefore follows that the same is true of any function of F , and that the truth-table of such a function can be directly calculated from the primitive tables.

2. The fundamental theorem as stated, or else in the form: if f_1 and f_2 is any pair of positive and negative functions respectively, then a necessary and sufficient condition that a function $f(p_1, p_2, \dots, p_n)$ be asserted is that each of the 2^n contained functions found by substituting f_1 and f_2 for the p 's is asserted. It will be noticed that theorem (1) tells us how to determine whether these latter are asserted.

We now pass on to several theorems about the system.

THEOREM. *It is possible to find 2^{2^n} functions of order n such that no two of them are equivalent and such that every other function of order n is equivalent to one of these.*

For we can find 2^{2^n} functions corresponding to the 2^{2^n} different tables of order n . The equivalence of any two of these will then not have a positive table and so will not be asserted. On the other hand any other n th order function will have the same table as one of the 2^{2^n} possible tables, and so the corresponding equivalence will be positive and hence asserted.

THEOREM. *An incompletely false function contains at least one function for each given function which is equivalent to that given function.*

COROLLARY. *An incompletely false function contains at least one true function and one completely false function.*

THEOREM. *The negative of a completely false function is true.*

For a completely false function has a negative truth-table, and so its negative will have a positive table and hence be asserted. It is worth noticing that although this theorem is immediate once we have the fundamental theorem it would be quite difficult without it.

COROLLARY. *Every function of F is either true, or its negative is true, or it contains both a true function and one whose negative is true.*

THEOREM. *The system of elementary propositions of 'Principia' is consistent.*

For if it were inconsistent we would have both a function and its negative asserted. But then both the function and its negative would have to have positive tables whereas if a function has a positive table its negative has a negative table.*

THEOREM. *Every function of the system can either be asserted by means of the postulates or else is inconsistent with them.*

* This argument requires merely the first part of the fundamental theorem which was proved quite simply.

For if a function be not asserted as a result of the postulates it will contain a function whose negative can be so asserted. If then we assert the original function, the contained function will be asserted so that we have asserted both a function and its negative, i.e., we have a contradiction.

COROLLARY. *A function is either asserted as a result of the postulates or else its assertion will bring about the assertion of every possible elementary proposition.*

For by the theorem we would obtain the assertion of both a function and its negative and so by $\sim p \supset p \supset q$ the assertion of the unmodified variable q . But q then represents any elementary proposition.

In conclusion let us note that while the fundamental theorem shows that the postulates bring about the assertion of those and only those theorems which should belong to the system, this last theorem enables us to say that they also automatically exclude the very possibility of any added assertions.

GENERALIZATION BY TRUTH-TABLES.

5. General Survey of the Systems Generated.—The system we have studied in the preceding sections is a particular system depending upon the two primitive functions $\sim p$ and $p \vee q$. Two modes of attack have presented themselves. On the one hand we have the original postulational method, on the other the truth-table development. In passing to a general study of systems of the kind discussed these two methods present themselves as instruments of generalization. We reserve the postulational generalization for the next portion of our paper and now take up the truth-table generalization.

To gain complete generality let us assume for our primitives μ arbitrary functions with an arbitrary number of arguments which we will designate by

$$f_1(p_1, p_2, \dots, p_{m_1}), f_2(p_1, p_2, \dots, p_{m_2}), \dots, f_\mu(p_1, p_2, \dots, p_{m_\mu})$$

and let us attach an arbitrary truth-table to each. By successive combinations of these functions with different or repeated arguments we generate the set of derived functions which as before we designate by F . Again each function of F will possess a truth-table in virtue of the tables of the primitive functions of which each is composed. Denote the set of truth-tables thus generated by T . Then whereas in the system of 'Principia' T consists of all possible truth-tables, this will not necessarily be the case here.

In another paper we completely determine all the possible systems T and show that there are 66 *systems that can be generated by tables of third and lower order, and 8 infinite families of systems that are generated by the introduction of fourth and higher ordered tables,*

If two systems have the same truth-tables the primitives of each can evidently be expressed in terms of those of the other so that truth-tables are preserved. We can then say that each system has a representation in the other and the two are equivalent. In particular *every truth-system has a representation in the system of Principia while every complete system, i.e., having all possible truth-tables, is equivalent to it.* In the aforementioned paper we also determine the ways in which a complete system may be generated, and it turns out that one table alone is sufficient to generate it, and it can be either of these two

+	+	-	+	+	-
+	-	+	+	-	-
-	+	+	-	+	-
-	-	+	-	-	+

a result first given by Sheffer as stated in the introduction.

The truth-table development for complete systems is essentially the same as that given in section 2. It is easy to prove for all systems the

THEOREM. *Every function contained in a positive function is positive; every function contained in a negative function is negative; every mixed function contains a function for every table of the system.*

6. Postulates for a Complete System.—We now show how to construct a set of postulates for any complete system such that: *the set of asserted functions is identical with the set of positive functions, while the assertion of any other function brings about the assertion of every elementary proposition* a property which also characterized the system of 'Principia.'

Let $\sim' p$ and $p \vee' q$ be functions in the given complete system with the tables of \sim and \vee . Out of \sim' and \vee' we then construct $p \supset' q$ and $p \equiv' q$ as $p \supset q$ and $p \equiv q$ are found from \sim and \vee , and also $f_1(p_1, \dots, p_{m_1}), \dots, f_\mu(p_1, \dots, p_{m_\mu})$ with the same tables as $f_1(p, \dots, p_{m_1}), \dots, f_\mu(p, \dots, p_{m_\mu})$. This is possible since \sim and \vee , and so \sim' and \vee' can generate a complete system. All the functions $\sim', \vee', \supset', \equiv', f'_1, \dots, f'_\mu$ are ultimately expressed in terms of the f 's and so belong to the system. Construct now the following set of postulates:

I. If p_1, \dots, p_{m_1} are elementary propositions, $f_1(p_1, \dots, p_{m_1})$ is.

If p_1, \dots, p_{m_μ} are elementary propositions, $f_\mu(p_1, \dots, p_{m_\mu})$ is.

II. The assertion of a function involving a variable p produces the assertion of any function found from the given one by substituting for p any other variable q , or $f_1(q_1, \dots, q_{m_1}), \dots$ or $f_\mu(q_1, \dots, q_{m_\mu})$.

III. " $\vdash P$ " and " $\vdash P \supset' Q$ " produces " $\vdash Q$ ".

IV. (1) $\vdash : p \vee' p \supset' p$ (a) $\vdash .f_1(p_1, p_2, \dots, p_{m_1}) \equiv' f'_1(p_1, p_2, \dots, p_{m_1})$,

(5) $\vdash \dots \quad (\mu) \vdash .f_\mu(p_1, p_2, \dots, p_{m_\mu}) \equiv' f'_\mu(p_1, p_2, \dots, p_{m_\mu})$.

where (1)–(5) are the assertions of IV in sec. 1 with \sim' and \vee' in place of \sim and \vee .

That all asserted functions are positive can be verified as in the proof of sec. 4. As for the converse, note that III and IV (1)–(5) being of the same form as III and IV of sec. 4 will yield the assertion of all positive functions expressed in terms of \sim' and \vee' . By the use of (a)–(μ) every function can be shown to be equivalent (\equiv') to some function expressed by \sim' and \vee' and so every positive function will be asserted. In the same way the assertion of any non-positive function will bring about the assertion of a non-positive function in \sim' and \vee' alone, and so of any proposition.

We thus see that complete systems are equivalent to the system of 'Principia' not only in the truth table development but also postulationally. As other systems are in a sense degenerate forms of complete systems we can conclude that no new logical systems are introduced.

7. Application to Nicod's Postulate Set.—Although, as in most existence theorems, the above set of postulates may not be the simplest in any one case, it can be used to advantage in showing that a given set has the same property as it possesses. For this purpose we show directly that all asserted functions are positive, and then that by means of the given postulates (a) each of our formal postulates may be derived (b) that the results derivable by our informal postulates can also be derived by the given ones.*

As an example we consider the set of postulates given by Nicod for the theory of elementary propositions in terms of the single primitive function of Sheffer's which Nicod denotes by $p|q$ and is termed incompatibility by Russell.† It is the first of the two functions given in section 5 as generating a complete system. Nicod gives the definitions

$$\sim p . = . p | p \quad Df, \quad p \vee q . = . p / p | q / q \quad Df$$

which we take to be our $\sim'p$ and $p \vee'q$ respectively. His $p \supset q . = . p | q / q Df$ however is not our $p \supset'q$ which is $\sim'p \vee'q$. The primary distinction of his system is that he uses but one formal primitive proposition.

In carrying out the proof suggested we merely note that by means of his informal proposition " $\vdash P$ " and " $\vdash P|R/Q$ " produce " $\vdash Q$ " we get the effect of " $\vdash P$ " and " $\vdash P|Q/Q$ " i.e., " $\vdash P \supset Q$ " produce " $\vdash Q$ " when $R = Q$. Since he has $p \supset'q \supset p \supset q$ we thus get the effect of " $\vdash P$ "

* That the informal postulates of a system must be proved effectively replaced by others in another system is a precaution rarely taken in discussions of equivalence or dependence of logical systems. Such a discussion is unnecessary in ordinary mathematical systems since their distinctive postulates are all formal, the informal ones being those of a common logic. But in comparing logical systems, which usually do contain different informal postulates, such a discussion is fundamental.

† B. Russell, *loc. cit.*, chap. XIV.

and " $\vdash P \supset' Q$ " produce " $\vdash Q$ " our III. Likewise each function IV is proved with however \supset in place of \supset' . But by means of $p \supset q, \supset p \supset' q$ this too is remedied. We then easily complete the proof of the

THEOREM. *If in Nicod's system we give to $p|q$ the table*

p, q	$p q$
++	-
+-	+
-+	+
--	+

then the set of asserted functions is identical with the resulting set of positive functions; and the assertion of any other function would bring about the assertion of every elementary proposition.

GENERALIZATION BY POSTULATION.

8. The Generalized Set of Postulates.—As in the truth-table development we assume arbitrary primitive functions of propositions

$$f_1(p_1, p_2, \dots, p_{m_1}), \dots, f_\mu(p_1, p_2, \dots, p_{m_\mu});$$

but in place of the arbitrary associated truth-tables we have a set of postulates of the following form. We have tried to preserve all the informal properties of the postulates of 'Principia' (and of sec. 5) but generalize the formal properties completely.

I. (As in sec. 5.)

II. (As in sec. 5.)

III. " $\vdash g_{11}(P_1, P_2, \dots, P_{k_1})$ " ... " $\vdash g_{\kappa_1}(P_1, P_2, \dots, P_{k_\kappa})$ "

" $\vdash g_{1k_1}(P_1, P_2, \dots, P_{k_1})$ " ... " $\vdash g_{\kappa k_\kappa}(P_1, P_2, \dots, P_{k_\kappa})$ "

produce ... produce

" $\vdash g_1(P_1, P_2, \dots, P_{k_1})$ " ... " $\vdash g_\kappa(P_1, P_2, \dots, P_{k_\kappa})$ "

where the P 's are any combinations of f 's including the special case of the unmodified variable, while the g 's are particular combinations of this kind which need not have all the indicated arguments.

IV. $\vdash h_1(p_1, p_2, \dots, p_{l_1})$

$\vdash h_2(p_1, p_2, \dots, p_{l_2})$

... ...

$\vdash h_\lambda(p_1, p_2, \dots, p_{l_\lambda})$

where the h 's are particular combinations of the f 's.

The retention of I and II which are characteristic of the theory of

elementary propositions is our justification for giving that name to the systems that may be generated by the above set of postulates. In what follows we give what we consider to be merely an introduction to the general theory.

9. Definition of Consistency and Related Concepts.—The prime requisite of a set of postulates is that it be consistent. Since the ordinary notion of consistency involves that of contradiction which again involves negation, and since this function does not appear in general as a primitive in the above system a new definition must be given.

Now an inconsistent system in the ordinary sense will involve the assertion of a pair of contradictory propositions which as we have seen will bring about the assertion of every elementary proposition through the assertion of the unmodified variable p . Conversely since p stands for any elementary proposition its assertion would yield the assertion of contradictory propositions and so render the system inconsistent. The two notions are thus equivalent in ordinary systems; and since one retains significance in the general case we are led to the

DEFINITION.—*A system will be said to be inconsistent if it yields the assertion of the unmodified variable p .*

In a consistent system we may then define a true function as one that can be asserted as a result of the postulates. Instead of defining a false function as one not true, we give the following

DEFINITION. *A false function is one such that if its assertion be added to the postulates the system is rendered inconsistent.*

We can then state that in the system of 'Principia' every function is true or false. This suggests the

DEFINITION. *If every function of a consistent system is true or false the system will be said to be closed.**

As a justification of this name we may note that the postulates of such a system automatically exclude the possibility of any added assertions—a state of affairs we believe to be highly desirable in the final form of a logical theory.

10. Properties of Consistent Systems.—In all that follows we assume that the system discussed is consistent. If it be inconsistent one could hardly say anything more about it.

We turn to a theorem which will give us most of the results of this section. But first we must state two lemmas which we do not further prove.

LEMMA 1.—If a given set of functions gives rise to some other function in accordance with II and III, and if these functions involve certain letters

* Had the name not been in use in a different connection we should have introduced the term categorical.

r_1, r_2, \dots, r_i upon which no substitution is made in the process, then the same deductive process will be valid if we have given the original functions with an arbitrary substitution of the r 's as described in II provided this substitution is also made throughout the process.

LEMMA 2.—The most general process of obtaining an assertion from a given set of assertions in accordance with II and III can be reduced to first asserting a number of functions in accordance with II, and then applying II and III in such a way that no substitutions are made on the arguments of those functions.

THEOREM. *Every false function contains a finite set of untrue first order functions $\phi_1(p), \phi_2(p), \dots, \phi_v(p)$ such that whenever p is replaced by an untrue function at least one of these functions remains untrue.*

By the definition of false functions there must be some deductive process whereby from the given false function and true functions we assert p . By lemma 2 we can replace this process by another where from the given false function and true functions we obtain certain contained functions from which without substitution of the arguments we obtain p . Now first by lemma 1 we can equate to p all the arguments thus appearing and still have a valid deductive process for obtaining p . Denote the resulting untrue functions which are contained in the original false function by $\phi_1(p), \phi_2(p), \dots, \phi_v(p)$. Then secondly by lemma 1 we can replace p by any function ψ and still have a valid process which now consists in obtaining ψ from certain true functions and $\phi_1(\psi), \dots, \phi_v(\psi)$. If then each $\phi_i(\psi)$ were true, ψ , being obtained from true functions in accordance with II and III would be true. It follows that if ψ be untrue, some $\phi_i(\psi)$ must be untrue.

THEOREM. *Every false function contains an infinite number of untrue first order functions; and if the system has at least one false function of order greater than one, then each false function contains an infinite number of untrue functions of every order.*

By the above theorem the false function contains at least one untrue function $\phi_{i_1}(p)$. By the same theorem some $\phi_{i_2}, \phi_{i_3}(p)$ must be untrue, etc., through $\phi_{i_4}, \phi_{i_5}, \dots, \phi_{i_n}(p)$. These are all different being of different rank, and are all contained in the given function.

The last part of the theorem may then be proved by showing that by replacing equal by unequal variables in the infinity of functions thus gotten from the false function of order greater than one we get untrue functions of every order, and so by the above method an infinite number of every order in every false function.

We have immediately the

THEOREM. *A necessary and sufficient condition that a function of a closed system be true is that all contained first order functions be true.*

COROLLARY. *It is also necessary and sufficient that all those of rank greater than some finite integer ρ be true.*

In analogy with corresponding ideas in the system of 'Principia' define a completely untrue function as one in which all contained functions are untrue with a similar definition for completely false. We then have the interesting

THEOREM. *If a system has a completely untrue function, then every false function contains a completely untrue function.*

Every function contained in the completely untrue function makes at least one $\phi_i(p)$ of a false function untrue. If ψ is such a contained function which makes say $\phi_{i_1}(p)$ true, then ψ will be completely untrue, and all contained functions will make $\phi_{i_1}(p)$ true yet some remaining $\phi_i(p)$ untrue. By repeating this process we finally obtain a function ψ' such that all contained functions make each $\phi_i(p)$ of a set that remains untrue. Each such $\phi_i(\psi')$ will then be a completely untrue function in the given one.

COROLLARY. *If a closed system has a completely false function every false function contains a completely false function.*

If we call such a system completely closed we have the stronger

THEOREM. *In a completely closed system every false function $f(p_1, p_2, \dots, p_n)$ contains a completely false function $f(\psi_1(p), \psi_2(p), \dots, \psi_n(p))$ where each $\psi_i(p)$ is either true or completely false.*

By equating all variables to p in the function of the corollary we get such a completely false function where some ψ 's may be incompletely false. These are then eliminated by successively substituting for p functions which make them true.

COROLLARY. *A necessary and sufficient condition that a function of a completely closed system be true is that all contained first order functions found by substituting true or completely false functions for the arguments be true.*

This property begins to approximate to the truth-table method. It leads us easily to the following criterion for a completely closed postulational system being a truth-system which we state without proof.

THEOREM. *A necessary and sufficient condition that a completely closed postulational system be a truth-system is that a true first order function remains true whenever we replace a true or completely false constituent function by any other true or completely false first order function respectively.**

* In making a more complete study of the postulational generalization it would be desirable to classify all the systems that may result more or less in the way in which we have classified truth-systems through the associated systems of truth-tables. In this connection we might define the order of a set of postulates as the largest number of premises used in deriving a conclusion in III, and the order of a system as the lowest order a set of postulates deriving it can have. It is then of interest to note that whereas the set of postulates of the system of 'Principia' is of the second order, the system itself is of the first order.

m-VALUED TRUTH-SYSTEMS.*

11. The Generalized (\sim , \vee) System.—We have seen that the truth-table generalization, at least with regard to complete systems, is included in the postulational development. We now show that the latter is more general by presenting a new class of systems, distinct from the two-valued systems of symbolic logic, which can be generated by a completely closed set of postulates.

In these systems instead of the two truth-values +, - we have m distinct "truth-values" t_1, t_2, \dots, t_m where m is any positive integer. A function of order n will now have m^n configurations in its truth-table, so that there will be m^{mn} truth-tables of order n . Calling a system having all possible tables complete, we now show that the following two tables generate a complete system.

p	$\sim_m p$	p, q	$p \vee_m q$
t_1	t_2	$t_1 t_1$	t_1
t_2	t_3	\dots	\dots
\dots	\dots	$t_{i_1} t_{j_1}$	$t_{i_1} \leq j_1$
t_m	t_1	\dots	$t_2 \geq j_2$
		$t_{i_2} t_{j_2}$	
		\dots	
		$t_m t_m$	t_m

We see that $\sim_m p$, the generalization of $\sim p$, permutes the truth-values cyclically, while $p \vee_m q$, the generalization of $p \vee q$ has the higher of the two truth-values.†

To construct a function for any first order table, of which there are m^m note that

$$t_1(p) = . p \vee \sim_m p \vee_m \sim_m^2 p \vee_m \dots \sim_m^{m-1} p \quad Df,$$

where $\sim^2 p = . \sim \sim p \quad Df$, etc., has all its truth values t_1 . Then

$$\tau_{m_1}(p) = . \sim_m^{m-1} (\sim_m^{m-1} (\sim_m t_1(p) \vee_m \dots \vee_m \sim_m^{m-1} p) \vee_m \dots \vee_m \sim_m^{m_1} p) \quad Df$$

has all values t_m except the first which is t_{m_1} . Any first order table

p	$f(p)$
t_1	t_{m_1}
t_2	t_{m_2}
\dots	\dots
t_m	t_{m_m}

can then be constructed by the function

* See Lewis, *loc. cit.*, p. 222 for the term "Two-Valued Algebra."

† The higher truth-value has here the smaller subscript.

$$\tau_{m_1}(p) \cdot \vee_m \cdot \tau_{m_2}(\sim_m^{m-1} p) : \vee_m \cdot \tau_{m_3}(\sim_m^{m-2} p) : \dots \vee_m \cdots \tau_{m_m}(\sim_m p).$$

Construct now a function for the table

p	$\sim_m p$
t_1	t_m
t_2	t_{m-1}
\dots	\dots
t_m	t_1

and define $p \cdot_m q = \sim_m(\sim_m p \cdot \vee_m \cdot \sim_m q)$ *Df* which is the generalization of $p \cdot q$ and has the lower of the two truth values of its arguments. We can now construct a table all of whose values are t_m except for one configuration $t_{m_1}, t_{m_2}, \dots, t_{m_n}$ when it is $t_{m_{m_1 m_2 \dots m_n}} = t_\mu$ by the function

$$\tau_\mu(\sim_m^{m-m_1+1} p_1) \cdot_m \tau_\mu(\sim_m^{m-m_2+1} p_2) \cdot_m \cdots \tau_\mu(\sim_m^{m-m_n+1} p_n),$$

and so any table by constructing such a function for each configuration and then "summing up" by \vee_m .

12. Classification of Functions—the m dimensional Space Analogy.—The generalization of the classification of functions into positive, negative and mixed is afforded us by the following

THEOREM. *A function contains at least one function for every truth-table whose values are contained among the values of the given table.*

Let $t_{m_1} \dots t_{m_\mu}$ be the truth-values that appear in the table of a given function $f(p_1, p_2, \dots, p_n)$. Then we can pick out μ configurations having these values respectively. Construct functions $\phi_i(p)$ such that when p has the value t_{m_j} of one of these configurations, $\phi_i(p)$ have the value of p_i in that configuration. It is then easily seen that $f(\phi_1(p), \dots, \phi_n(p))$ has the value t_{m_j} whenever p has the value t_{m_j} . If then $\psi(q_1, q_2, \dots, q_l)$ have a table whose values are among the t_{m_j} 's, $f(\phi_1(\psi), \dots, \phi_n(\psi))$ will be a function contained in the given function with that table.

We are thus led to a classification of functions by means of their truth-tables such that the set of tables of ^{functions} contained in a given function is the same for all functions in a given class. We then have m classes of functions where but one truth-value appears, $[m(m-1)]/2!$ with two truth-values, \dots , $[m(m-1) \cdots (m-\mu+1)]/\mu!$ with μ truth-values, \dots , one class with all m truth-values. We thus have $2^m - 1$ classes of functions which when $m = 2$ reduces to the three classes of positive, negative and mixed functions.

These formulæ suggest an analogy which, if well founded, is of great interest. For this purpose replace the set of functions having all of a given set of μ truth-values by all functions whose values are among these μ values. If then we compare the functions of our complete system to the points of a

space of m dimensions,* the m classes of functions with but one truth-value would correspond to the m coordinate axes, the $[m(m - 1)]/2!$ classes of functions with no more than two truth-values to the $[m(m - 1)]/2!$ coordinate planes, etc., so that except for the absence of an origin all properties of determination and intersection within the coordinate configurations go over. If then we attach the name m -dimensional truth-space to our system, we observe the following difference, that whereas the highest dimensioned intuitionistic point space is three, the highest dimensioned intuitionistic proposition space is two. But just as we can interpret the higher dimensioned spaces of geometry intuitionistically by using some other element than point, so we shall later interpret the higher dimensioned spaces of our logic by taking some other element than proposition.

13. Truth-Table Characteristics of Asserted Functions.—The following analysis presupposes that in constructing a set of postulates for the system we at least wish to impose the

CONDITION.—*If a function is asserted, all functions with the same truth-table will be asserted.*

It follows from the theorem of the preceding section that under the given condition, *if a function is asserted, every function of the truth-space it determines is asserted.*

We can now prove that *if the system is to be completely closed its asserted functions must constitute a single truth-space contained in the given truth space.* For if there were at least two such spaces, then a function having all their truth-values would be false, and so would contain a completely false function. This in turn would contain functions with but one truth-value; and these being therefore in one of the two given spaces would be true which contradicts their being in a completely false function.

No loss of generality ensues if we take the truth values of this contained truth-space of asserted functions to be t_1, t_2, \dots, t_μ , where, to avoid degenerate cases $0 < \mu < m$. We now show that a completely closed set of postulates can be constructed for all such systems.

14. A Completely Closed Set of Postulates for the Systems.—I and II are determined directly as in the general case. To obtain III, construct a function $p \supset_m q$ whose table is given by the following: when the truth-value of p is that of q or lower, $p \supset_m q$ will have the value t_1 , while if the truth-value of p is above that of q , then if the value of p is t_μ or higher, $p \supset_m q$ will have the value of q , while if it is below t_μ , say t_ν and that of q is t_ν' , then the truth-value of $p \supset_m q$ will be $t_{\nu'-\nu+1}$. III will then be simply

* Or we might take the truth-table as element in which case the system is perhaps smoother than before.

"⊢ P"
 "⊢ P ⊨_m Q"
 produce
 "⊢ Q"

Now by generalizing each part *A*, *B*, *C*, *D* of the proof of the fundamental theorem of sec. 3 it can be shown that by the assertion of a finite number of functions with values from t_1 to t_μ all such can be obtained.* If then we assert these functions in IV we shall have every function in the μ -space asserted. Furthermore no others can be asserted for by the use of II and III we can only get functions with values from t_1 to t_μ by means of functions similarly restricted. This is obvious in II while in III if the value of *P* is from t_1 to t_μ while that of *Q* is below t_μ , then from the above definition of the table of $P \supset_m Q$ it would have the value of *Q* and so be below t_μ . But that contradicts the assumption that the premises had values from t_1 to t_μ .

This set of postulates will then give the proper set of true functions. Furthermore let us suppose that we assert a function with at least one value below t_μ . This will contain a function $\phi(p)$ with but one value, and that below t_μ . By II, $\phi(p)$ will be asserted. Furthermore since $\phi(p) \cdot \supset_m \phi(p)$ $\supset_m \sim_m \phi(p)$ has its value t_1 it will be asserted, and so we obtain by III $\sim_m \phi(p)$. Repetition of this process will finally give us a function $\psi(p)$ with but one value t_m . But $\psi(p) \cdot \supset_m p$ is asserted having but one value t_1 . We thus obtain the assertion of *p*. The system is therefore closed. And since all functions with values from $t_{\mu+1}$ to t_m are completely false, the system is completely closed.

15. Comparison of Systems.—As in the truth-table development we can generalize the systems by using arbitrary functions as primitives, and as was done there we can show how to generate a complete m -dimensioned system by one second order function, and how to give a completely closed set of postulates for all complete systems. The problem of determining all possible systems of m -dimensional truth-tables, however, is one we have not considered, though its solution would throw considerable light on the ordinary problem.

We turn now to the following

DEFINITIONS. *A closed system S with primitives f_1, f_2, \dots, f_n has a representation in a closed system S' with primitives f'_1, f'_2, \dots, f'_n if we can so replace the f's by functions in S' that a function in S will be true when and only when the correspondent in S' is true.*

Two systems are equivalent if each has a representation in the other.

Denote a complete m -dimensional truth-system with the asserted functions forming a truth-space of μ dimensions by ${}_m T_m$. We then have the

* Lack of space prevents us from giving the details.

THEOREM. Two complete truth-systems ${}_{\mu}T_m$ and ${}_{\mu'}T'_{m'}$ are equivalent when and only when $\mu = \mu'$ and $m = m'$.

The conditions are clearly sufficient since we can make truth-values correspond. To prove them necessary suppose $m > m'$. If we construct m^m functions of first order in T with different truth-tables then there will be two, $\phi_1(p), \phi_2(p)$ whose correspondents $\phi'_1(p), \phi'_2(p)$ have the same truth-tables since there are in T' only $m'^{m'}$ of first order. Let $\chi(p, q)$ have value t_1 when p and q have the same value and t_m otherwise. Then $\chi(\phi_1, \phi_1)$ is true; hence $\chi'(\phi'_1, \phi'_1)$ is. ϕ'_2 having the same table as ϕ'_1 , $\chi'(\phi'_1, \phi'_2)$ is true, and hence $\chi(\phi_1, \phi_2)$ the correspondent. But that would make ϕ_1 have the same table as ϕ_2 . Now suppose $\mu > \mu'$. If ϕ have all the values from t_1 to t_{μ} and no others there are μ^{μ} functions with values t_1 to t_{μ} of the form $\psi\phi(p)$. These will then be asserted and so the correspondents will be asserted and have values t'_1 to $t'_{\mu'}$. Since we can only have $\mu'^{\mu'}$ functions $\psi'\phi'(p)$ with different tables, we can find two of the μ^{μ} correspondents with the same table. The above contradiction then results as before.

For representation we have only found the

THEOREM. To represent ${}_{\mu}T_m$ in ${}_{\mu'}T'_{m'}$, it is necessary to have $\mu \leq \mu', m \leq m'$; it is sufficient to have $\mu \leq \mu', m - \mu \leq m' - \mu'$.

COROLLARY. A necessary and sufficient condition that ${}_{\mu}T_m$ have a representation in ${}_{\mu'}T'_{m'}$, is that $m \leq m'$.

It is of interest to note as a result that the only complete truth-systems equivalent to the system of 'Principia' are ${}_1T_2$'s; and though it can be represented in every complete truth-system, only ${}_1T_2$'s can be represented in it. We have thus verified our statement that we obtain essentially new logical systems.

16. Interpretation of m -valued Truth-systems in Terms of Ordinary Logic.—Let the elementary proposition of the (\sim_m, \vee_m) system be interpreted as an ordered set of $(m - 1)$ elementary propositions of ordinary logic $P = (p_1, p_2, \dots, p_{m-1})$ such that if one proposition is true all those that follow are true. P will be then be said to have the truth-value t_1 if all the p 's are true, t_2 if all but one are true, etc. Also P will be said to be true if at most $(\mu - 1)p$'s are false.

If $P = (p_1, p_2, \dots, p_{m-1}), Q = (q_1, q_2, \dots, q_{m-1})$ we define

$$P \vee_m Q . = . (p_1 \vee q_1, p_2 \vee q_2, \dots, p_m \vee q_m) \quad Df$$

$$\begin{aligned} \sim_m P . = . (\sim(p_1 \vee p_2 \vee \dots \vee p_{m-1}), \sim(p_1 \vee \dots \vee p_{m-1}) \cdot \vee . p_1 \cdot p_2, \dots, \\ \sim(p_1 \vee \dots \vee p_{m-1}) \cdot \vee . p_{m-2} \cdot p_{m-1}) \quad Df \end{aligned}$$

We easily justify these definitions by showing first that $P \vee_m Q$ and $\sim_m P$ are "elementary propositions" when P and Q are, and secondly that they

have the proper truth tables. Thus in $P \vee_m Q$ the first p, q , to be true is the first for which either p or q is true; also all later terms will have p or q true and so will be true. $P \vee_m Q$ is therefore elementary and has the required table.

But in spite of this representation $_1T_2$ still appears to be the fundamental system since its truth-values correspond entirely to the significance of true and completely false, whereas in $_\mu T_m$, $m > 2$ either $\mu > 1$ or $m - \mu > 1$, and this equivalence no longer holds. We must however take into account the fact that our development has been given in the language of $_1T_2$ and for that very reason every other kind of system appears distorted. This suggests that if we translate the entire development into the language of any one $_\mu T_m$ by means of its interpretation, then it would be the formal system most in harmony with regard to the two developments.

Reprinted from
Transactions of the
American Mathematical Society
Vol. 32 (1930), 723-781.

GENERALIZED DIFFERENTIATION*

BY
EMIL L. POST

INTRODUCTION

The history of the subject of generalized differentiation can be traced back to Leibnitz.[†] In the earlier literature the term fractional differentiation is used as an alternative, and the aim was to generalize the concept of the n th derivative of a function to non-integral values of n . Four different attacks on this problem may be noted. The earliest was that of Liouville, who expanded the functions operated upon in series of exponentials, and assumed, as a basis, $D^n e^{ax} = a^n e^{ax}$, where D symbolizes differentiation. Riemann considered power series with non-integral exponents as analogues of Taylor's series, and through their coefficients was led to the expression of generalized derivatives in terms of a definite integral plus an infinite series with arbitrary constant coefficients.[‡] Liouville's and Riemann's results proved to be in disagreement. Prior to the publication of Riemann's work, Grünwald was led by the restrictions of Liouville's method to generalize directly the definition of a derivative as the limit of a finite difference quotient, and, by rigorous methods, also arrived at definite integral formulas.[§] Grünwald's definition involved the idea of differentiation between limits which later resulted in the coordination of Liouville's and Riemann's results. This, for example, was effected by Krug,[¶] who introduced a new development, based on Cauchy's contour integral for ordinary derivatives, which also involved limits of differentiation, in terms of which he showed that Riemann's definite integral corresponded to finite lower limit, Liouville's development to lower limit $-\infty$. The Riemann-Grunwald definite integral form has become standard in the literature, and has been intensively studied.^{||}

* Presented to the Society, October 27, 1923; received by the editors June 29, 1929.

† For references, see S. Pincherle, *Équations et opérations fonctionnelles*, Encyclopédie des Sciences Mathématiques, Paris, 1912, tome 2, vol. 5, fasc. 1, pp. 1-81; also Eugene Stephens, *Symbolic calculus. Bibliography on general (or fractional) differentiation*, Washington University Studies, vol. 12 (1925), No. 2, pp. 137-152.

‡ B. Riemann, *Gesammelte Mathematische Werke*, Leipzig, 1876, pp. 331-344.

§ K. A. Grünwald, *Zeitschrift für Mathematik und Physik*, vol. 12 (1867), pp. 441-480.

¶ A. Krug, $d^n f(x)/dx^n$ regarded as a function of n , Akademie der Wissenschaften, Wien, Denkschriften, Mathematisch-Naturwissenschaftliche Klasse, vol. 57 (1890), pp. 151-228.

|| E.g. by A. Marchaud, *Sur les dérivées et sur les différences des fonctions de variables réelles*, Journal de Mathématiques, (9), vol. 6 (1927), pp. 337-425.

Another trend was introduced by the symbolic treatment of linear differential equations with constant coefficients as exemplified by the work of Boole. The polynomial operators thus occurring led naturally to the concept of an arbitrary operator $f(D)$, which was to be formally expanded in powers of D , and thus applied to the operand. This treatment has since been established on a rigorous basis for operators $f(D)$ corresponding to entire transcendental functions $f(z)$ of genus zero, operating upon functions which are analytic in a given region.* Boole's methods for linear differential equations with variable coefficients were extended by other writers to yield formal solutions in terms of operators algebraic in D ; but in many cases no further attempt was made to assign a meaning to such expressions.

We may think of the more recent work of Heaviside as the next step in this development.† Of course Heaviside's contribution assumes an importance far beyond this formal juggling of symbols, through its application to important physical problems, and its skillful methods for evaluating the operations that are used. This "operational calculus," however, was developed with physical intuition, rather than mathematical rigor, as guide. A more rigorous mathematical basis has since been supplied by Carson in terms of solutions of Laplace integral equations, and their use in a definite integral formula.‡ Carson's formulas include the Riemann-Grünwald definite integral for D^n as a special case when the real part of n is less than one; but his treatment is quite unrelated to the theory of entire operators of genus zero.§

In the present paper Grünwald's method of arriving at a definition for operators D^n is carried forward by means of an artifice of Arbogast to yield a definition of generalized differentiation for operators $f(D)$.¶ This definition is shown to include Carson's operators, and entire operators of genus zero, as special cases.|| The major part of the paper is devoted to operators $f(D)$,

* C. Bourlet, *Sur les opérations en général et les équations linéaires différentielles d'ordre infini*, Annales de l'Ecole Normale, (3), vol. 33 (1897), pp. 133-190.

† J. F. Ritt, *On a general class of linear homogeneous differential equations of infinite order with constant coefficients*, these Transactions, vol. 18 (1917), pp. 27-49.

‡ Oliver Heaviside, *Electromagnetic Theory*, London, 1922, vol. 2, chapters 7, 8.

§ J. R. Carson, *The Heaviside operational calculus*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 43-68; also numerous papers in the Bell System Technical Journal.

|| Another theory is developed by Norbert Wiener: *The operational calculus*, Mathematische Annalen, vol. 95 (1926), pp. 557-584.

¶ Our definition may therefore be called the extended Grünwald definition.

|| This is not strictly correct as far as Carson's development is concerned, since we derive his formulas only under certain hypotheses on the functions involved. It should be noted, however, that Carson does not explicitly state the domain of applicability of these formulas, and that the hypotheses in question are of considerable generality.

termed of type zero, which correspond to functions of a complex variable, $f(z)$, which are analytic in a certain sector of the z -plane of angle greater than π , and whose moduli satisfy within this sector the Poincaré inequality for entire functions of genus zero.* Existence theorems are established for such operators, and certain formal properties are investigated, such as the law of successive operations, and the generalized Leibnitz theorem for differentiation of a product.† The last three sections are chiefly devoted to operators given by a Laplace integral. For these we have only established the existence theorem, and investigated the application to Carson's development where not $f(z)$, but $f(z)/z$, is given by a Laplace integral.

Next to the definition of generalized differentiation itself, the writer wishes to call attention to the associated operator $A[f](t)$, defined by

$$A[f](t) = \lim_{\substack{\Delta x \rightarrow +0 \\ r \Delta x \rightarrow t}} \frac{(-1)^r f^{(r)}(1/\Delta x)}{r! \Delta x^{r+1}}.$$

The existence of this limit, for the cases considered, serves as the basis of our theory. It possesses the notable property of inverting the Laplace transformation, and enjoys numerous formal relations which flow directly from its definition. For $f(D)$ of type zero, it can be expressed by a contour integral, which is the Fourier integral with the vertical line contour replaced by two half-lines running into the negative half of the z -plane. The exponential factor of the integrand thus acquires a potency for convergence which should be of considerable use in most practical applications, where the loss of generality incurred by requiring analyticity in a sector of angle greater than π , as opposed to analyticity in a half-plane, is irrelevant, due to the presence of but a finite number of singularities.

In order to bring this paper to a conclusion, many topics have been excluded either because they involve an extension of our definition, or because they depend on the formulas flowing from our definition, rather than on the definition itself. Among these omitted topics are an extension of our definition to complex limits of differentiation, the application to Volterra's functions of composition of the closed cycle group, and the derivation of the Heaviside expansions.

* By thus breaking away from the classic assumption of analyticity in a half-plane, the restriction on the modulus of $f(z)$ is so greatly lightened (see footnote following proof of Theorem I) that we are able to include all operators $f(D)$ for which $f(z)$ is an entire function of genus zero, as well as all operators for which $f(z)$ is algebraic. The half-plane assumption would exclude these classes, as such, and would also hinder the theory by not permitting indiscriminate differentiation with respect to the upper limit.

† This treatment of the extended Grünwald definition follows very closely Grünwald's development of the original definition.

1. The definition. Our definition of generalized differentiation grows out of the expression of the n th derivative of a function as the limit of its n th difference quotient. For our purpose, this expression is best written in the form

$$D^n\phi(x) = \lim_{\Delta x \rightarrow 0}$$

$$\frac{\phi(x) - n\phi(x - \Delta x) + [n(n-1)/2!] \phi(x - 2\Delta x) - \cdots + (-1)^n \phi(x - n\Delta x)}{\Delta x^n},$$

where Δx is the negative of the increment of x as ordinarily defined. Since the right hand member of this equation continues to have meaning when n is not a positive integer, it can be used to define $D^n\phi(x)$ for arbitrary n .

When $n = -1$, the suggested definition becomes

$$D^{-1}\phi(x) = \lim_{\Delta x \rightarrow +0} [\phi(x)\Delta x + \phi(x - \Delta x)\Delta x + \phi(x - 2\Delta x)\Delta x + \cdots],$$

i.e., the limit of an infinite series. If, however, we arbitrarily terminate the series at the $(p+1)$ st term, where $x_0 < x - p\Delta x \leq x_0 + \Delta x$, the result, at least for $\phi(x)$ continuous, will be the definite integral of $\phi(x)$ with finite lower limit x_0 . Replacing the upper limit by X , we shall use the notation $\{D^{-1}\}_{x_0}^X \phi(x)$ for the limit of the finite sum, $\{D^{-1}\}_{-\infty}^X \phi(x)$ for the limit of the infinite series. Similarly for D^n , n arbitrary.*

To extend this definition still further, let us momentarily introduce the operator E^m , defined by the relation $E^m\phi(x) = \phi(x+m)$. The reader will have seen the analogy between the above expression for $D^n\phi(x)$ and the binomial series expansion. Through the operator E^m , this analogy becomes formal, and we can write, apart from refinements,

$$D^n\phi(x) = \lim_{\Delta x \rightarrow 0} \left(\frac{1 - E^{-\Delta x}}{\Delta x} \right)^n \phi(x) = \lim_{\Delta x \rightarrow 0} \left(\frac{\Delta}{\Delta x} \right)^n \phi(x).$$

This immediately suggests the desired definition of $f(D)\phi(x)$, i.e.,

$$f(D)\phi(x) = \lim_{\Delta x \rightarrow 0} f\left(\frac{\Delta}{\Delta x}\right) \phi(x) = \lim_{\Delta x \rightarrow 0} f\left(\frac{1 - E^{-\Delta x}}{\Delta x}\right) \phi(x),$$

where $f(1/\Delta x - E^{-\Delta x}/\Delta x)$ is to be expanded by Taylor's series, and formally applied to $\phi(x)$. As in the case of D^{-1} the limit of the infinite series will be written $\{f(D)\}_{-\infty}^X \phi(x)$. In the case of finite lower limit x_0 , the choice of p through the inequalities $x_0 < X - p\Delta x \leq x_0 + \Delta x$ allows Δx to approach zero independently of $X - x_0$, except for sign. p is then the largest integer less

* Up to this point the writer essentially retraces Grünwald's argument.

than $(X - x_0)/\Delta x$. We shall use the notation $p = \{(X - x_0)/\Delta x\}$. Though otherwise arbitrary, Δx must have the same sign as $X - x_0$. In this paper Δx will approach zero positively, that is, X will be greater than x_0 . Our completed definitions thus become

$$(1) \quad \{f(D)\}_{x_0}^X \phi(x) = \lim_{\Delta x \rightarrow +0} \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) = \lim_{\Delta x \rightarrow +0} \left\{ f\left(\frac{1 - E^{-\Delta x}}{\Delta x}\right) \right\}_{x_0}^X \phi(x)$$

$$= \lim_{\substack{\Delta x \rightarrow +0 \\ p = \{(X - x_0)/\Delta x\}}} \left[f\left(\frac{1}{\Delta x}\right) \phi(X) - \frac{f'\left(\frac{1}{\Delta x}\right)}{1! \Delta x} \phi(X - \Delta x) \right. \\ \left. + \dots + \frac{(-1)^p f^{(p)}\left(\frac{1}{\Delta x}\right)}{p! \Delta x^p} \phi(X - p\Delta x) \right].$$

$$(2) \quad \{f(D)\}_{-\infty}^X \phi(x) = \lim_{\Delta x \rightarrow +0} \left[f\left(\frac{1}{\Delta x}\right) \phi(X) - \frac{f'\left(\frac{1}{\Delta x}\right)}{1! \Delta x} \phi(X - \Delta x) + \dots \right].$$

The existence of the limits involved in these definitions is proved for various classes of operators in subsequent sections of the paper. In certain simple cases they can be evaluated directly. For example, consider $f(D) = \log D$, $\phi(x) \equiv 1$, and x_0 finite. Then by (1)

$$\left\{ \log\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X 1 = \log\left(\frac{1}{\Delta x}\right) - 1 - \frac{1}{2} - \dots - \frac{1}{p} \\ = - \left[1 + \frac{1}{2} + \dots + \frac{1}{p} - \log p \right] - \log(p\Delta x).$$

As $\Delta x \rightarrow +0$, p increases indefinitely, while $p\Delta x \rightarrow X - x_0$. Furthermore, as $p \rightarrow \infty$, the bracket has for limit γ , the Eulerian constant. We thus find

$$(3) \quad \{ \log(D) \}_{x_0}^X 1 = -\gamma - \log(X - x_0).$$

Besides proving existence theorems, with their associated formulas, we shall be concerned with the verification of the formal laws of generalized differentiation for our definitions. Of these the most important is the law of successive operations. This can be immediately verified for the finite difference operators on which our generalized derivatives are based. Assuming the necessary number of derivatives of f_1 and f_2 , we have

$$\begin{aligned}
& \left\{ f_1 \left(\frac{\Delta}{\Delta x} \right) \right\}_{x_0}^X \left\{ f_2 \left(\frac{\Delta}{\Delta x} \right) \right\}_{x_0}^x \phi(\xi) = f_1 \left(\frac{1}{\Delta x} \right) \left[f_2 \left(\frac{1}{\Delta x} \right) \phi(X) \right. \\
& - \frac{f'_2 \left(\frac{1}{\Delta x} \right)}{1! \Delta x} \phi(X - \Delta x) + \cdots + \frac{(-1)^p f_2^{(p)} \left(\frac{1}{\Delta x} \right)}{p! \Delta x^p} \phi(X - p \Delta x) \left. \right] \\
& - \frac{f'_1 \left(\frac{1}{\Delta x} \right)}{1! \Delta x} \left[f_2 \left(\frac{1}{\Delta x} \right) \phi(X - \Delta x) - \cdots + \frac{(-1)^{p-1} f_2^{(p-1)} \left(\frac{1}{\Delta x} \right)}{(p-1)! \Delta x^{p-1}} \phi(X - p \Delta x) \right] \\
& + \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\
& + \frac{(-1)^p f_1^{(p)} \left(\frac{1}{\Delta x} \right)}{p! \Delta x^p} \left[f_2 \left(\frac{1}{\Delta x} \right) \phi(X - p \Delta x) \right] \\
& = \left[f_1 \left(\frac{1}{\Delta x} \right) f_2 \left(\frac{1}{\Delta x} \right) \right] \phi(X) \\
& - \frac{\left[f_1 \left(\frac{1}{\Delta x} \right) f'_2 \left(\frac{1}{\Delta x} \right) + f'_1 \left(\frac{1}{\Delta x} \right) f_2 \left(\frac{1}{\Delta x} \right) \right]}{1! \Delta x} \phi(X - \Delta x) \\
& + \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\
& + (-1)^p \frac{\left[f_1 \left(\frac{1}{\Delta x} \right) f_2^{(p)} \left(\frac{1}{\Delta x} \right) + \frac{p}{1!} f'_1 \left(\frac{1}{\Delta x} \right) f_2^{(p-1)} \left(\frac{1}{\Delta x} \right) + \cdots + f_1^{(p)} \left(\frac{1}{\Delta x} \right) f_2 \left(\frac{1}{\Delta x} \right) \right]}{p! \Delta x^p} \\
& \cdot \phi(X - p \Delta x),
\end{aligned}$$

so that, by Leibnitz's theorem and our definition, we obtain

$$(4) \quad \left\{ f_1 \left(\frac{\Delta}{\Delta x} \right) \right\}_{x_0}^X \left\{ f_2 \left(\frac{\Delta}{\Delta x} \right) \right\}_{x_0}^x \phi(\xi) = \left\{ f_1 \left(\frac{\Delta}{\Delta x} \right) f_2 \left(\frac{\Delta}{\Delta x} \right) \right\}_{x_0}^X \phi(x).$$

The character of definition (1) is more apparent in the following form:

$$\begin{aligned}
(5) \quad & \{f(D)\}_{x_0}^X \phi(x) = \lim_{\substack{\Delta x \rightarrow +0 \\ p = \{(X-x_0)/\Delta x\}}} \sum_{r=0}^p A[f](r, \Delta x) \phi(X - r \Delta x) \Delta x; \\
& A[f](r, \Delta x) = \frac{(-1)^r f^{(r)} \left(\frac{1}{\Delta x} \right)}{r! \Delta x^{r+1}}.
\end{aligned}$$

(5) would lead directly to a definite integral if in place of $A[f](r, \Delta x)$ we had a function of $r\Delta x$. The way in which this difficulty can be partly overcome may be indicated by means of the operator D^n , n arbitrary. Using the notation $A[f(u)](r, \Delta x)$ when the form of f is specified, we have, in this case,

$$\begin{aligned} A[u^n](r, \Delta x) &= \frac{(-1)^r n(n-1) \cdots (n-r+1)}{r! \Delta x^{n+1}} \\ &= \frac{(-n)(-n+1) \cdots (-n+r-1)}{r! r^{-n-1}} (r\Delta x)^{-n-1}. \end{aligned}$$

The Gaussian form of the gamma function gives

$$\lim_{r \rightarrow \infty} \frac{(-n)(-n+1) \cdots (-n+r-1)}{r! r^{-n-1}} = \frac{1}{\Gamma(-n)}.$$

If at the same time $\Delta x \rightarrow +0$ in such a way that $r\Delta x \rightarrow t$, $t > 0$, it will follow that $A[u^n](r, \Delta x)$ will approach a function of t as limit. Symbolizing this function by $A[u^n](t)$, we thus find

$$(6) \quad A[u^n](t) = \lim_{\substack{\Delta x \rightarrow +0 \\ r\Delta x \rightarrow t}} A[u^n](r, \Delta x) = \frac{t^{-n-1}}{\Gamma(-n)}.$$

More generally we shall look for the existence of a limit function $A[f](t)$ corresponding to $A[f](r, \Delta x)$. When such a limit exists, then for Δx sufficiently small, and $r\Delta x$ greater than some positive ϵ , $A[f](r, \Delta x)$ will be approximated by $A[f](r\Delta x)$; and so, (5) will lead in part to a definite integral. Nevertheless, the restriction $r\Delta x > \epsilon$ leaves terms of (5) with small r unaccounted for. These will usually require separate treatment, where, however, the $A[f](t)$'s of operators connected with $f(D)$ will be used. The first step in our theory is therefore the establishing of the existence of $A[f](t)$ for a sufficiently wide class of operators $f(D)$. We pause a moment to consider a certain limit criterion which will unify the last stages of many of our proofs.

2. Limit criterion. If $F(\Delta x)$ can be expressed as a sum $G(\Delta x, \nu) + H(\Delta x, \nu)$, ν independent of Δx , such that

$$\lim_{\Delta x \rightarrow +0} G(\Delta x, \nu) = g(\nu), \quad \limsup_{\Delta x \rightarrow +0} |H(\Delta x, \nu)| \leq h(\nu), \quad \lim_{\nu \rightarrow \nu_0} h(\nu) = 0,$$

then $\lim_{\Delta x \rightarrow +0} F(\Delta x)$ exists, and is given by

$$(7) \quad \lim_{\Delta x \rightarrow +0} F(\Delta x) = \lim_{\nu \rightarrow \nu_0} g(\nu).$$

In fact the stated conditions show that

$$\limsup_{\Delta x \rightarrow +0} F(\Delta x) \leq g(\nu) + h(\nu), \quad \liminf_{\Delta x \rightarrow +0} F(\Delta x) \geq g(\nu) - h(\nu).$$

As these upper and lower limits are independent of ν , we find, by letting $\nu \rightarrow \nu_0$, and observing that $h(\nu) \rightarrow 0$,

$$\limsup_{\Delta x \rightarrow +0} F(\Delta x) = \liminf_{\Delta x \rightarrow +0} F(\Delta x) = \lim_{\nu \rightarrow \nu_0} g(\nu).$$

Since for any one ν the values of $g(\nu) + h(\nu)$ and $g(\nu) - h(\nu)$ are finite, the above inequalities yield the

COROLLARY. *Under the given hypothesis $\lim_{\Delta x \rightarrow +0} F(\Delta x)$, and hence also $\lim_{\nu \rightarrow \nu_0} g(\nu)$, is finite.**

3. Existence of $A[f](t)$ for $f(D)$ of type zero. An operator $f(D)$, corresponding to a function of a complex variable $f(z)$, will be said to be of *type zero* when $f(z)$ satisfies the following two conditions:

- (a) *$f(z)$ is analytic in a sector of the z -plane of angle greater than π bisected by the positive direction of the axis of reals,*
- (b) *for each real and positive κ , however small, there corresponds a real and positive K , such that*

$$|f(z)| \leq K e^{\kappa|z|},$$

for every z in the analytic sector.

We shall designate the positive acute angle made by the sides of the sector with the negative direction of the axis of reals by α .

Operators of finite order, defined later, which include all operators whose $f(z)$ is algebraic, are also of type zero. Other examples are e^{aD^m} , $m < 1$, and operators for which $f(z)$ is an entire function of genus zero.

We shall now prove the fundamental

THEOREM I. *If $f(D)$ is of type zero, $A[f](r, \Delta x)$ approaches a finite limit as $\Delta x \rightarrow +0$, $r\Delta x \rightarrow t$, for every real and positive t . The resulting function of t , $A[f](t)$, is given by the formula*

$$(8) \quad A[f](t) = \frac{1}{2\pi i} \int_C e^{tz} f(z) dz,$$

where C is formed by two rays within the analytic sector and parallel to its edges, with common end point on the axis of reals, and traversed so that, along it, the imaginary part of z increases.

The proof is based on Cauchy's second integral formula. For Δx sufficiently small, $1/\Delta x$ will be in the analytic sector posited by condition (a).

* Henceforth, existence will include finiteness.

If C''' is a simple contour about $1/\Delta x$, traversed counterclockwise, and also contained in this sector, then, by the formula in question, we shall have

$$f^{(r)}\left(\frac{1}{\Delta x}\right) = \frac{r!}{2\pi i} \int_{C'''} \frac{f(z)dz}{\left(z - \frac{1}{\Delta x}\right)^{r+1}};$$

and so, by the definition of $A[f](r, \Delta x)$ given in (5),

$$A[f](r, \Delta x) = -\frac{1}{2\pi i} \int_{C'''} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}}.$$

Assuming Δx sufficiently small, we can choose for C''' the contour formed by C'' , an arc of a circle, center the origin, radius $k/\Delta x$, $k > 1$, joined to C' , the finite portion of C cut off by this circle. If C'' is traversed in the same direction as C''' , but C' in the opposite direction, i.e., in the same direction as C , we can write schematically

$$A[f](r, \Delta x) = -\frac{1}{2\pi i} \int_{C''} + \frac{1}{2\pi i} \int_{C'}.$$

Along C'' we have $|z| = k/\Delta x$. Hence $|1 - z\Delta x| \geq k - 1$, so that, by inequality (b), we find

$$\left| \int_{C''} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| < \frac{K \cdot 2\pi k}{\Delta x(k-1)} \left[\frac{e^{k\kappa/(r\Delta x)}}{k-1} \right]^r.$$

Choose $k > 2$, and then κ so that $e^{k\kappa/(r\Delta x)} / (k-1) < 1$. The right hand member of the inequality will then approach zero as limit as $\Delta x \rightarrow +0, r\Delta x \rightarrow t$. The C'' contribution can therefore be neglected.

Now break up C' into $C_{l,m}$ and $C'_{l,m}$, where $C_{l,m}$ extends distances l and m respectively along the two segments of C' from their meeting point, while $C'_{l,m}$ consists of the rest of C' . For l and m sufficiently large, $R(z)$, the real part of z , will be negative along $C'_{l,m}$, and we shall have

$$\left| \int_{C'_{l,m}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| < K \int_{C'_{l,m}} \frac{e^{\kappa|z|} |dz|}{[1 - R(z)\Delta x]^{r+1}}.$$

If we set

$$1 - R(z)\Delta x = e^{-\lambda R(z)\Delta x},$$

we observe that λ stays positive and decreases monotonically as $-R(z)\Delta x$ increases. Now the largest value of $-R(z)\Delta x$ along any one $C'_{l,m}$ corresponds to $|z| = k/\Delta x$. As $\Delta x \rightarrow +0$ this largest value approaches $k \cos \alpha$ as limit. Hence for all Δx 's sufficiently small, and all corresponding $C'_{l,m}$'s, $-R(z)\Delta x$

remains less than some fixed positive quantity. The corresponding λ 's therefore have a positive lower bound λ_0 . Since $\lambda \geq \lambda_0 > 0$, we shall thereby have

$$1 - R(z)\Delta x > e^{-\lambda_0 R(z)\Delta x},$$

and so

$$\left| \int_{C'_{l,m}} \frac{f(z) dz}{(1 - z\Delta x)^{r+1}} \right| < K \int_{C'_{l,m}} e^{\kappa|z| + \lambda_0(r+1)\Delta x} |R(z)| |dz|.$$

Since $R(z)/|z| \rightarrow -\cos \alpha$, as $|z| \rightarrow \infty$, if we choose κ less than $\lambda_0 t \cos \alpha$ the last integral will be less than

$$\int_{C'_{l,m}} e^{-\mu|z|} |dz|,$$

with fixed positive μ , for l and m sufficiently large, and $(r+1)\Delta x$ sufficiently near t . The contour $C'_{l,m}$ depends on Δx . If we replace it by $\bar{C}'_{l,m}$, which consists of C with $C_{l,m}$ removed, and of which $C'_{l,m}$ is a part, we observe that the resulting integral converges, and so approaches zero as limit as l and m increase indefinitely. We therefore have

$$\limsup_{\substack{\Delta x \rightarrow +0 \\ r\Delta x \rightarrow t}} \left| \int_{C'_{l,m}} \frac{f(z) dz}{(1 - z\Delta x)^{r+1}} \right| \leq K \int_{\bar{C}'_{l,m}} e^{-\mu|z|} |dz|, \quad \lim_{\substack{l \rightarrow \infty \\ m \rightarrow \infty}} \int_{\bar{C}'_{l,m}} e^{-\mu|z|} |dz| = 0.$$

Finally $f(z)/(1 - z\Delta x)^{r+1}$ uniformly approaches $e^{tz}f(z)$ along $C_{l,m}$. Hence

$$\lim_{\substack{\Delta x \rightarrow +0 \\ r\Delta x \rightarrow t}} \int_{C_{l,m}} \frac{f(z) dz}{(1 - z\Delta x)^{r+1}} = \int_{C_{l,m}} e^{tz} f(z) dz.$$

We can therefore apply our limit criterion,* and obtain

$$\lim_{\substack{\Delta x \rightarrow +0 \\ r\Delta x \rightarrow t}} \int_{C'} \frac{f(z) dz}{(1 - z\Delta x)^{r+1}} = \lim_{\substack{l \rightarrow \infty \\ m \rightarrow \infty}} \int_{C_{l,m}} e^{tz} f(z) dz = \int_C e^{tz} f(z) dz.$$

Since we saw that the integral along C'' could be neglected, this establishes both the existence of $A[f](t)$, and the formula given for it.[†]

* This was stated for one independent variable Δx . It can obviously be extended to any number of independent variables, in this case two.

[†] If the analyticity condition imposed on $f(z)$ be weakened to analyticity in the half-plane to the right of some line $R(z) = a$, while within that half-plane the modulus of $f(z)$ satisfies the far stronger inequality

$$|f(z)| \leq K/|z|^{1+\epsilon}, \quad \epsilon > 0,$$

then essentially the same proof will yield the existence of $A[f](t)$, and its expression by means of a Fourier integral. Furthermore, to anticipate the later developments, the argument leading to (48) in §11 can be duplicated in this case, so that our methods would yield the classic solution of the Laplace integral equation.

By real-variable methods, it can be proved that if $A[f](t)$ exists for each t in a closed interval (t_1, t_2) , then, without any further hypothesis on the function f , or even on the form of $A[f](r, \Delta x)$, the following consequences hold:

(a) $A[f](t)$ is continuous in the closed interval (t_1, t_2) ;

(b) for each positive ϵ however small, there corresponds a positive η , such that, for $\Delta x < \eta$,

$$|A[f](r, \Delta x) - A[f](r\Delta x)| < \epsilon$$

for all r 's for which $r\Delta x$ is in the interval (t_1, t_2) .*

From our definition we have, for $x_0 < x_1 < X$,

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X \phi(x) = \left\{ f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^X \phi(x) + \sum_{r=q+1}^p A[f](r, \Delta x) \phi(X - r\Delta x) \Delta x,$$

where $p = \{(X - x_0)/\Delta x\}$, $q = \{(X - x_1)/\Delta x\}$. If we set $X - r\Delta x = x$, then, as r varies from $q+1$ to p , x stays in the closed interval (x_0, x_1) , while $r\Delta x$ stays in the interval $(X - x_1, X - x_0)$. Suppose then that $\phi(x)$ is continuous in (x_0, x_1) , while $A[f](t)$ exists in $(X - x_1, X - x_0)$. Then by (b), $A[f](r, \Delta x)$ can be replaced by $A[f](r\Delta x)$, i.e., by $A[f](X - x)$, in the above sum. But by (a), $A[f](X - x)$ will be continuous in x in the interval (x_0, x_1) . The new sum will therefore lead to a definite integral as $\Delta x \rightarrow +0$. Hence

THEOREM II. *If $A[f](t)$ exists in the interval $(X - x_1, X - x_0)$, where $x_0 < x_1 < X$ and if $\phi(x)$ is continuous in the interval (x_0, x_1) , then*

$$(9) \quad \left\{ f(D) \right\}_{x_0}^X \phi(x) = \left\{ f(D) \right\}_{x_1}^X \phi(x) + \int_{x_0}^{x_1} A[f](X - x) \phi(x) dx,$$

provided either of the indicated operations exist.

When $f(D)$ is of type zero $A[f](t)$ exists for all positive t 's. If then $\phi(x)$ is continuous in (x_0, X) , (9) will be valid with x_1 anywhere in this interval.

* The proof runs as follows: Choose any positive ϵ . For each point t' of the interval (t_1, t_2) , $A[f](t')$ exists, and is defined as $\lim A[f](r, \Delta x)$ as $\Delta x \rightarrow +0, r\Delta x \rightarrow t'$. Hence for each t' of (t_1, t_2) there is a positive η' such that $|A[f](r, \Delta x) - A[f](t')| < \epsilon/2$ provided $\Delta x < \eta'$ and $|r\Delta x - t'| < \eta'$. By letting $\Delta x \rightarrow +0$ and $r\Delta x \rightarrow t$ we obtain $|A[f](t) - A[f](t')| \leq \epsilon/2$ provided $|t - t'| < \eta'$ and t is in (t_1, t_2) . Hence $A[f](t)$ is continuous at every point t' of (t_1, t_2) .

Consider now the open intervals $(t' - \eta', t' + \eta')$ thus associated with the above pairs (t', η') . Every point of (t_1, t_2) is in fact the midpoint of such an interval. Hence by the Heine-Borel theorem a finite number of these intervals suffice to cover (t_1, t_2) . Each interval uniquely determines the corresponding t' and η' . Let η be the smallest of the η' 's of this finite set of intervals, and let $\Delta x < \eta$. Any $r\Delta x$ in (t_1, t_2) will be in one of these intervals. Since, for the (t', η') of this interval, $\Delta x < \eta \leq \eta'$ and $|r\Delta x - t'| < \eta'$, we will have $|A[f](r, \Delta x) - A[f](t')| < \epsilon/2$ and $|A[f](r\Delta x) - A[f](t')| \leq \epsilon/2$. Hence $|A[f](r, \Delta x) - A[f](r\Delta x)| < \epsilon$. That is, this inequality holds for every r and Δx for which $\Delta x < \eta$ and $r\Delta x$ is in (t_1, t_2) .

The suggestion made at the end of §1 is thus verified. Before we see how to deal with $\{f(D)\}_{x_0}^X \phi(x)$, we shall consider $\phi(x)$ a polynomial. In this case, $\{f(D)\}_{x_0}^X \phi(x)$ can be immediately evaluated.

4. Finite difference reduction formula; $\phi(x)$ a polynomial, unity. For the formula about to be derived it is essential to distinguish between $\Delta\phi(x)/\Delta x$, which in this paper is written $[\phi(x) - \phi(x - \Delta x)]/\Delta x$, and $\{\Delta/\Delta x\}_{x_0}^X \phi(\xi)$ which, while the same as $\Delta\phi(x)/\Delta x$ for $x - x_0 > \Delta x$, is but $\phi(x)/\Delta x$ for $x - x_0 \leq \Delta x$. While (4) would hold with $\{\Delta/\Delta x\}_{x_0}^X \phi(\xi)$ for $\{f_2(\Delta/\Delta x)\}_{x_0}^X \phi(\xi)$, it does not hold with $\Delta\phi(x)/\Delta x$ in place of $\{\Delta/\Delta x\}_{x_0}^X \phi(\xi)$. Instead, we obtain by the same method

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \frac{\Delta\phi(x)}{\Delta x} = \left\{ \frac{\Delta}{\Delta x} f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) - A[f(u)](p, \Delta x) \phi(x'_0);$$

$$p = \left\{ \frac{X - x_0}{\Delta x} \right\}, \quad x'_0 = X - (p + 1)\Delta x. *$$

It will be noticed that x'_0 does not vary with X as long as X changes by multiples of Δx . Also $x_0 - \Delta x < x'_0 \leq x_0$. If we replace $f(u)$ by $u^{-1}f(u)$ and rearrange its terms, this formula becomes

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) = \left\{ \left(\frac{\Delta}{\Delta x}\right)^{-1} f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \frac{\Delta\phi(x)}{\Delta x} + A[u^{-1}f(u)](p, \Delta x) \phi(x'_0).$$

By induction, we are thus led to the fundamental formula

$$(10) \quad \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) = \left\{ \left(\frac{\Delta}{\Delta x}\right)^{-m} f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \frac{\Delta^m \phi(x)}{\Delta x^m}$$

$$+ \sum_{\mu=0}^{m-1} A[u^{-\mu-1}f(u)](p, \Delta x) \frac{\Delta^\mu \phi(x'_0)}{\Delta x^\mu}.$$

It will be referred to as the finite difference reduction formula.

If $P(x)$ is a polynomial of degree n , we have

$$\frac{\Delta^{n+1} P(x)}{\Delta x^{n+1}} = 0.$$

Let then $\phi(x) = P(x)$, $m = n + 1$, in (10). It becomes

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X P(x) = \sum_{\mu=0}^n A[u^{-\mu-1}f(u)](p, \Delta x) \frac{\Delta^\mu P(x'_0)}{\Delta x^\mu}.$$

* The definition of $\{f(\Delta/\Delta x)\}_{x_0}^X \phi(x)$ only requires $\phi(x)$ to be defined in the interval (x_0, X) , whereas both members of this formula use values of x less than x_0 , when $(X - x_0)/\Delta x$ is not an integer. We must therefore arbitrarily define $\phi(x)$ for a suitable interval beyond x_0 to render the formula applicable. The validity of the formula, however, does not depend on the particular way in which this prolongation is effected.

If we now let $\Delta x \rightarrow +0$, it will follow that $p\Delta x \rightarrow X - x_0$, and $x'_0 \rightarrow x_0$. Hence if $A[u^{-\mu-1}f(u)](X-x_0)$ exists for $\mu=0, 1, \dots, n$, $\{f(D)\}_{x_0}^X P(x)$ will also exist, and will be given by

$$(11) \quad \{f(D)\}_{x_0}^X P(x) = \sum_{\mu=0}^n A[u^{-\mu-1}f(u)](X - x_0) P^{(\mu)}(x_0).$$

For $P(x) \equiv 1$, we can use $n=0$, and so obtain

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X 1 = A[u^{-1}f(u)](p, \Delta x).$$

Hence, if $A[u^{-1}f(u)](X-x_0)$ exists, $\{f(D)\}_{x_0}^X 1$ exists, and is given by

$$(12) \quad \{f(D)\}_{x_0}^X 1 = A[u^{-1}f(u)](X - x_0).$$

It is easily verified that if $f(D)$ is of type zero then $D^{-\mu-1}f(D)$ also is of type zero. Hence the existence theorem for $\phi(x)$ a polynomial is completely established for operators of type zero.

The relation (12) is of special interest and importance. Note that in spite of the finite difference relation preceding it, we cannot conclude that if $\{f(D)\}_{x_0}^X 1$ exists, $A[u^{-1}f(u)](X-x_0)$ also will exist; for in the former p depends on Δx , while for the latter $p\Delta x$ should vary independently of Δx .

5. Operators of finite order; existence theorems. In establishing existence theorems for $\{f(D)\}_{x_0}^X \phi(x)$, we shall find that the wider the class of operators $f(D)$ we consider, the greater the restrictions we have to impose on $\phi(x)$. Hence the following specialization of operators of type zero.

An operator $f(D)$ will be said to be of *finite order* ρ , ρ zero or a positive integer, when

- (a) $f(z)$ satisfies the analyticity condition for operators of type zero,
- (b) there exists a positive ϵ , and corresponding K , such that

$$|f(z)| \leq K |z|^{\rho-\epsilon}$$

for every z in the analytic sector for which $|z| > \delta$, $\delta > 0$.

According to this definition, if an operator is of order ρ , it is also of any order greater than ρ .

D^n , n arbitrary, is a typical example. All algebraic operators are of finite order. It is evident that operators of finite order are also of type zero. Hence §3 is immediately applicable.

We shall first derive an inequality for $A[f](r, \Delta x)$ which is essential for our existence proofs. With the notation of §3 we obtain by the new condition (b)

$$\left| \int_{C''} \frac{f(z)dz}{(1-z\Delta x)^{r+1}} \right| < 2\pi K \left(\frac{k}{\Delta x} \right)^{\rho+1-\epsilon} / (k-1)^{r+1}, \quad k/\Delta x > \delta.$$

For fixed Δx , and $r \geq \rho$, this expression approaches zero as limit as k increases indefinitely, i.e., as the radius of C'' is made to increase indefinitely. We can therefore replace C''' by C , the infinite contour of §3, and write, for $r \geq \rho$,

$$A[f](r, \Delta x) = \frac{1}{2\pi i} \int_C \frac{f(z)dz}{(1-z\Delta x)^{r+1}}.$$

Choose C so that its vertex is at $z=1/[2(r+1)\Delta x]$, and use condition (b) along it. This will be possible for $(r+1)\Delta x$ not too large. Through the change of variable $\xi = z(r+1)\Delta x$, we then obtain

$$|A[f](r, \Delta x)| \leq \frac{K}{2\pi} [(r+1)\Delta x]^{-\rho-1+\epsilon} \int_{C_{1/2}} \frac{|\xi|^{\rho-\epsilon} |d\xi|}{\left| 1 - \frac{\xi}{r+1} \right|^{r+1}},$$

where $C_{1/2}$ has its vertex at $\xi=1/2$. Since $R(\xi) \leq 1/2$ along $C_{1/2}$, we have

$$\frac{1}{\left| 1 - \frac{\xi}{r+1} \right|^{r+1}} \leq \frac{1}{\left(1 - \frac{R(\xi)}{r+1} \right)^{r+1}}.$$

Now we know that $[1+x/(r+1)]^{r+1}$ is an increasing function of x , $r \geq 0$, both for positive x , and for negative x with $|x| < 1$. We therefore have along $C_{1/2}$

$$\frac{1}{\left(1 - \frac{R(\xi)}{r+1} \right)^{r+1}} \leq \frac{1}{\left(1 - \frac{R(\xi)}{\rho+1} \right)^{\rho+1}}.$$

The integral along $C_{1/2}$, which depends only on r , is thus seen to be bounded for $r \geq \rho$, so that we obtain

$$|A[f](r, \Delta x)| < L' [(r+1)\Delta x]^{-\rho-1+\epsilon}.$$

A like inequality is obtained for $r < \rho$ by using for C''' a circle center $1/\Delta x$, radius $\theta/\Delta x$, $0 < \theta < 1$. We thus arrive at the following result:

For all Δx 's and r 's with $(r+1)\Delta x < a$, where a is some fixed positive quantity, we have the inequality

$$(13) \quad |A[f](r, \Delta x)| < L [(r+1)\Delta x]^{-\rho-1+\epsilon}.$$

Hence also, for $t < a$, we have

$$(14) \quad |A[f](t)| \leq Lt^{-\rho-1+\epsilon}.$$

The simplest formula for $\{f(D)\}_{x_0}^X \phi(x)$ results when $f(D)$ is of order zero. As at the end of §3, we shall write

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) = \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{X-h}^X \phi(x) + \sum_{r=q+1}^p A[f](r, \Delta x) \phi(X - r\Delta x) \Delta x;$$

$$q = \{h/\Delta x\}.$$

In the same manner also, we find for fixed positive h ,

$$\lim_{\Delta x \rightarrow +0} \sum_{r=q+1}^p A[f](r, \Delta x) \phi(X - r\Delta x) \Delta x = \int_{x_0}^{X-h} A[f](X - x) \phi(x) dx.$$

Now $\phi(x)$ is to be assumed continuous in (x_0, X) . Let M be the upper bound of $|\phi(x)|$ in this interval. For $h < a$ we can apply (13), with $\rho = 0$, and obtain

$$\left| \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{X-h}^X \phi(x) \right| < LM \sum_{r=0}^{\{h/\Delta x\}} [(r+1)\Delta x]^{-1+\epsilon} \Delta x.$$

By comparing the indicated sum with the integral of $\xi^{-1+\epsilon} d\xi$ between limits 0 and $h+\Delta x$ or Δx and $h+2\Delta x$, according as $0 < \epsilon < 1$, or $\epsilon \geq 1$, we see that

$$\limsup_{\Delta x \rightarrow +0} \left| \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{X-h}^X \phi(x) \right| \leq LM \frac{h^\epsilon}{\epsilon}.$$

Since $h^\epsilon/\epsilon \rightarrow 0$, as $h \rightarrow +0$, we can apply the limit criterion, and obtain

THEOREM III. If $f(D)$ is of order zero, and $\phi(x)$ is continuous in (x_0, X) , then $\{f(D)\}_{x_0}^X \phi(x)$ exists, and is given by

$$(15) \quad \{f(D)\}_{x_0}^X \phi(x) = \lim_{h \rightarrow +0} \int_{x_0}^{X-h} A[f](X - x) \phi(x) dx = \int_{x_0}^X A[f](X - x) \phi(x) dx.*$$

* If (8) is introduced in (15), we obtain

$$\{f(D)\}_{x_0}^X \phi(x) = \frac{1}{2\pi i} \int_K^X \left[\int_C e^{(X-x)z} f(z) dz \right] \phi(x) dx.$$

It may interest the reader to note that this formula might have suggested itself in the following formal manner. Symbolic use of Cauchy's second integral formula gives

$$f(D) = \frac{1}{2\pi i} \int_K \frac{f(z) dz}{z - D}.$$

On the other hand, the linear differential equation would suggest

$$\left\{ \frac{1}{Z - D} \right\}_{x_0}^X \phi(x) = - \int_{x_0}^X e^{(X-x)z} \phi(x) dx,$$

so that we would be led to

$$\{f(D)\}_{x_0}^X \phi(x) = - \frac{1}{2\pi i} \int_K \left[\int_{x_0}^X e^{(X-x)z} \phi(x) dx \right] f(z) dz.$$

Reversing the sense in which "contour" K is traversed, and changing the order of integration leads to the actual formula.

It will be noticed that the last integral may be improper for $x=X$; but of its convergence we are assured both by the limit criterion, and by the direct application of (14). To illustrate (15), we may take $f(D)=D^n$, $R(n)<0$. The operator is then of order zero. By (6) and (15), we thus get the standard Riemann-Grünwald form

$$(16) \quad \{D^n\}_{x_0}^X \phi(x) = \frac{1}{\Gamma(-n)} \int_{x_0}^X (X-x)^{-n-1} \phi(x) dx, \quad R(n) < 0.$$

We shall give two existence proofs for operators of arbitrary finite order. The first proof depends on the observation that if $f(D)$ is of order ρ , then $D^{-\rho}f(D)$ is of order zero, and so comes under Theorem III. Inasmuch as additional assumptions on $\phi(x)$ in a left neighborhood of $x=X$ will be required, we shall use Theorem II with x_1 in this neighborhood.

Formula (10), with x_1 in place of x_0 , and $m=\rho$, becomes

$$\begin{aligned} \left\{ f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^X \phi(x) &= \left\{ \left(\frac{\Delta}{\Delta x}\right)^{-\rho} f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^X \frac{\Delta^\rho \phi(x)}{\Delta x^\rho} \\ &+ \sum_{\mu=0}^{\rho-1} A[u^{-\mu-1}f(u)](q, \Delta x) \frac{\Delta^\mu \phi(x_1')}{\Delta x^\mu}, \quad q = \{(X-x_1)/\Delta x\}. \end{aligned}$$

Let $\phi(x)$ possess a continuous ρ th derivative in an interval (k, X) , $X>k$,* and assign to x_1 a value within this interval. By the law of the mean, and the continuity of $\phi^{(\rho)}(x)$ in (k, X) , it follows that, for sufficiently small Δx ,

$$\left| \frac{\Delta^\rho \phi(x)}{\Delta x^\rho} - \phi^{(\rho)}(x) \right| < \delta_{\Delta x}, \quad \lim_{\Delta x \rightarrow +0} \delta_{\Delta x} = 0,$$

for all x 's in (x_1, X) .† Hence if $\phi^{(\rho)}(x)$ be substituted for $\Delta^\rho \phi(x)/\Delta x^\rho$ in the above equation, the result will be changed in absolute value by no more than

$$\delta_{\Delta x} \sum_{r=0}^q |A[u^{-\rho}f(u)](r, \Delta x)| \Delta x.$$

Since $D^{-\rho}f(D)$ is of order zero, the proof of Theorem III shows this sum to be bounded, and hence the change in question to approach zero as limit as $\Delta x \rightarrow +0$. We then easily derive

* Here, as later, when a derivative is assumed to exist in a left neighborhood of X , the derivative at X is to be but a left derivative.

† See de la Vailée Poussin, *Cours d'Analyse*, vol. 1, 1914, §109, for the case $\rho=1$. By (1), §118, this is directly extended to arbitrary ρ .

THEOREM IV. If $f(D)$ is of order ρ , and if $\phi(x)$ is continuous in (x_0, X) , and possesses a continuous ρ th derivative in a left neighborhood of $x=X$, then $\{f(D)\}_{x_0}^X \phi(x)$ exists; and, if x_1 is within this neighborhood, is given by

$$(17) \quad \begin{aligned} \{f(D)\}_{x_0}^X \phi(x) &= \int_{x_1}^X A[u^{-\rho} f(u)](X-x)\phi^{(\rho)}(x)dx \\ &+ \sum_{\mu=0}^{\rho-1} A[u^{-\mu-1} f(u)](X-x_1)\phi^{(\mu)}(x_1) + \int_{x_0}^{x_1} A[f](X-x)\phi(x)dx. \end{aligned}$$

For illustration, again take D^n , with $\rho-1 \leq R(n) < \rho$. We obtain

$$(18) \quad \begin{aligned} \{D^n\}_{x_0}^X \phi(x) &= \frac{1}{\Gamma(\rho-n)} \int_{x_1}^X (X-x)^{\rho-n-1} \phi^{(\rho)}(x)dx \\ &+ \sum_{\mu=0}^{\rho-1} \frac{(X-x_1)^{\mu-n}}{\Gamma(\mu-n+1)} \phi^{(\mu)}(x_1) + \frac{1}{\Gamma(-n)} \int_{x_0}^{x_1} (X-x)^{-n-1} \phi(x)dx. * \end{aligned}$$

For the second proof note that, except for successive operations, the upper limit X plays the part of a constant. It can therefore appear as such in the operand $\phi(x)$. Now suppose that a continuous $\phi(x)$ satisfies the inequality

$$|\phi(x)| \leq N(X-x)^\rho$$

in a left neighborhood of $x=X$. This, with (13), leads to the identical inequalities that gave us Theorem III. Hence (15) holds for such a $\phi(x)$.

Let then a continuous $\phi(x)$ possess a finite ρ th left derivative at $x=X$, and hence also left derivatives of all lower orders. Then, as a result of a first theorem in Taylor's expansion, we have with finite N , in a left neighborhood of $x=X$,

$$|\phi(x)-P(x)| < N(X-x)^\rho, \quad P(x) = \sum_{\mu=0}^{\rho-1} \frac{\phi^{(\mu)}(X)}{\mu!} (x-X)^\mu.$$

If $\phi(x)$ is continuous in (x_0, X) , $\{f(D)\}_{x_0}^X [\phi(x)-P(x)]$ will exist, as seen above, and be given by (15). On the other hand, $P(x)$ is but a polynomial in x , so that $\{f(D)\}_{x_0}^X P(x)$ exists, and could be evaluated by (11). Hence

THEOREM V. If $f(D)$ is of order ρ , and if $\phi(x)$ is continuous in (x_0, X) , and possesses a finite ρ th left derivative at $x=X$, then $\{f(D)\}_{x_0}^X \phi(x)$ exists, and is given by

$$(19) \quad \begin{aligned} \{f(D)\}_{x_0}^X \phi(x) &= \int_{x_0}^X A[f](X-x) \left[\phi(x) - \sum_{\mu=0}^{\rho-1} \frac{\phi^{(\mu)}(X)}{\mu!} (x-X)^\mu \right] dx \\ &+ \{f(D)\}_{x_0}^X \sum_{\mu=0}^{\rho-1} \frac{\phi^{(\mu)}(X)}{\mu!} (x-X)^\mu. \end{aligned}$$

* This reduces to Grünwald's result if $x_1=x_0$.

Though this theorem requires less of $\phi(x)$ than Theorem IV, we shall nevertheless find Theorem IV more useful in our development. To illustrate (19), we shall take the operator $\log D$, whose order is one. By direct calculation we have $A[\log u](t) = -1/t$. Hence, by (19) and (3), we get

$$(20) \quad \{\log D\}_{x_0}^X \phi(x) = \int_{x_0}^X \frac{\phi(X) - \phi(x)}{X - x} dx - [\gamma + \log(X - x_0)]\phi(X).$$

It will be seen that this second proof uses the hypothesis that $f(D)$ is of finite order ρ only to enable us to assume the existence of $A[u^{-\mu}f(u)](t)$, for $t > 0$, $\mu = 0, 1, \dots, \rho$, and the validity of inequality (13). This suggests that we define $f(D)$ to be of *extended order* ρ if $A[u^{-\mu}f(u)](t)$ exists for $t > 0$, $\mu = 0, 1, \dots, \rho$, and (13) is satisfied. The first proof also uses the fact that $D^{-\mu}f(D)$ is of order zero. But it can be proved that if $f(D)$ is of extended order ρ , $D^{-\mu}f(D)$ is of extended order zero. Hence all of our existence theorems are valid for operators of extended finite order.

6. Existence theorems for operators of type zero. The crucial theorem for operators of type zero is the following:

THEOREM VI. *If $f(D)$ is of type zero, while $\phi(x)$ is analytic for $x_1 \leq x \leq X$, with the radius of convergence at x_1 greater than $X - x_1$, then $\{f(D)\}_{x_1}^X \phi(x)$ exists, and is given by the convergent series*

$$(21) \quad \{f(D)\}_{x_1}^X \phi(x) = A[u^{-1}f(u)](X - x_1)\phi(x_1) + A[u^{-2}f(u)](X - x_1)\phi'(x_1) + \dots$$

As a consequence of this theorem, if $\phi(x)$ is continuous in (x_0, X) , but analytic in some left neighborhood of X , then by choosing x_1 in this neighborhood, with $X - x_1$ less than half of the radius of convergence of $\phi(x)$ at X , both (9) and (21) become applicable, and so give

THEOREM VII. *If $f(D)$ is of type zero, while $\phi(x)$ is continuous in (x_0, X) , and analytic in a left neighborhood of $x = X$, then $\{f(D)\}_{x_0}^X \phi(x)$ exists, and, if x_1 is chosen as indicated above, is given by*

$$(22) \quad \{f(D)\}_{x_0}^X \phi(x) = \sum_{\mu=0}^{\infty} A[u^{-\mu-1}f(u)](X - x_1)\phi^{(\mu)}(x_1) + \int_{x_0}^{x_1} A[f](X - x)\phi(x)dx.$$

It may be noted that with 1 for $f(D)$, formula (21) reduces to the Taylor expansion of $\phi(x)$ for $x = x_1$. More generally, (21) is the result of operating on this Taylor expansion with $\{f(D)\}_{x_1}^X$, term by term.

Our proof of Theorem VI is an extension of the first existence proof for operators of finite order. As in that proof, we use the finite difference reduction formula, with limits x_1, X , viz.

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_1}^X \phi(x) = \left\{ \left(\frac{\Delta}{\Delta x}\right)^{-m} f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_1}^X \frac{\Delta^m \phi(x)}{\Delta x^m} + \sum_{\mu=0}^{m-1} A [u^{-\mu-1} f(u)](q, \Delta x) \frac{\Delta^\mu \phi(x_1')}{\Delta x^\mu}.$$

On the other hand, we shall let m vary with Δx , in fact equal $\{\epsilon/\Delta x\}$, with fixed and positive ϵ . Our proof will require the following condition on ϵ :

$$0 < \epsilon < \frac{l - (X - x_1)}{2},$$

where l is between $X - x_1$ and the radius of convergence of $\phi(x)$ at x_1 . Note that the left hand member of the reduction formula uses values of x only in the interval (x_1, X) , whereas in the right hand member the values spread over the interval $(x_1 - m\Delta x, X)$. Since $m\Delta x < \epsilon < l$, all of these x 's fall in the interval of convergence of the Taylor expansion of $\phi(x)$ at x_1 ; and so this Taylor expansion may be used to define $\phi(x)$ for $x < x_1$ for the purpose of our proof.

This proof consists essentially in establishing the following:

$$(a) \text{ For fixed } N, \lim_{\Delta x \rightarrow +0} \sum_{\mu=0}^{N-1} A [u^{-\mu-1} f(u)](q, \Delta x) \frac{\Delta^\mu \phi(x_1')}{\Delta x^\mu} = \sum_{\mu=0}^{N-1} A [u^{-\mu-1} f(u)](X - x_1) \phi^{(\mu)}(x_1).$$

$$(b) \text{ With } m = \{\epsilon/\Delta x\}, \lim_{\Delta x \rightarrow +0} \left\{ \left(\frac{\Delta}{\Delta x}\right)^{-m} f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_1}^X \frac{\Delta^m \phi(x)}{\Delta x^m} = 0.$$

$$(c) \text{ With } m = \{\epsilon/\Delta x\}, \limsup_{\Delta x \rightarrow +0} \sum_{\mu=N}^{m-1} A [u^{-\mu-1} f(u)](q, \Delta x) \frac{\Delta^\mu \phi(x_1')}{\Delta x^\mu} \leq h(N); \lim_{N \rightarrow \infty} h(N) = 0.$$

Since (b) allows us to neglect its term of the reduction formula, (a) and (c) together give us Theorem VI by means of the limit criterion.

(a) This is immediate.

(b) Formula (4) allows us to write

$$\left\{ \left(\frac{\Delta}{\Delta x}\right)^{-m} f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_1}^X \frac{\Delta^m \phi(x)}{\Delta x^m} = \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_1}^X \left\{ \left(\frac{\Delta}{\Delta \xi}\right)^{-m} \right\}_{x_1}^X \frac{\Delta^m \phi(\xi)}{\Delta \xi^m}, \Delta \xi = \Delta x.$$

By the law of the mean we have

$$\frac{\Delta^m \phi(\xi)}{\Delta \xi^m} = \phi^{(m)}(\xi - \theta m \Delta x); \quad 0 < \theta < 1.*$$

The Taylor expansion of $\phi(x)$ at x_1 enables us to define the function of a complex variable $\phi(z)$. Let M be the upper bound of the modulus of $\phi(z)$ on, and hence also within, the circle center x_1 , radius l . The smallest distance from $\xi - \theta m \Delta x$ to this circle is $l - |\xi - \theta m \Delta x - x_1|$, which certainly exceeds $l - \epsilon - (\xi - x_1)$ for all ξ 's in (x_1, X) . By applying the standard inequality of complex variable theory to $|\phi^{(m)}(\xi - \theta m \Delta x)|$, we thus obtain

$$\left| \frac{\Delta^m \phi(\xi)}{\Delta \xi^m} \right| < \frac{Mm!}{[l - \epsilon - (\xi - x_1)]^m},$$

so that we can write

$$\left| \left\{ \left(\frac{\Delta}{\Delta \xi} \right)^{-m} \right\}_{x_1}^x \frac{\Delta^m \phi(\xi)}{\Delta \xi^m} \right| < \sum_{s=0}^{\{(x-x_1)/\Delta x\}} A[u^{-m}](s, \Delta x) \frac{Mm!}{[l - \epsilon - (x - s\Delta x - x_1)]^m} \Delta x.$$

It is easily seen that

$$A[u^{-m}](s, \Delta x) = \frac{m(m+1) \cdots (m+s-1)}{s!} \Delta x^{m-1} < \frac{(\epsilon + s\Delta x)^{m-1}}{(m-1)!}.$$

Now the inequalities imposed on ϵ make $(\epsilon + s\Delta x)/[l - \epsilon - (x - s\Delta x - x_1)]$ an increasing function of $s\Delta x$. Since $s\Delta x < x - x_1$, we thereby obtain, on replacing $s\Delta x$ by $x - x_1$, the inequality

$$\frac{\epsilon + s\Delta x}{l - \epsilon - (x - s\Delta x - x_1)} < \frac{\epsilon + (x - x_1)}{l - \epsilon},$$

and so find in succession

$$\begin{aligned} \left| \left\{ \left(\frac{\Delta}{\Delta \xi} \right)^{-m} \right\}_{x_1}^x \frac{\Delta^m \phi(\xi)}{\Delta \xi^m} \right| &< \frac{M(x - x_1)}{\epsilon} \cdot m \left[\frac{\epsilon + (x - x_1)}{l - \epsilon} \right]^m, \\ \left| \left\{ f \left(\frac{\Delta}{\Delta x} \right) \right\}_{x_1}^x \left\{ \left(\frac{\Delta}{\Delta \xi} \right)^{-m} \right\}_{x_1}^x \frac{\Delta^m \phi(\xi)}{\Delta \xi^m} \right| & \\ &< \frac{M(X - x_1)}{\epsilon} \cdot m \left[\frac{\epsilon + (X - x_1)}{l - \epsilon} \right]^m \sum_{r=0}^g |A[f](r, \Delta x)| \Delta x. \end{aligned}$$

* This, as also the use of the law of the mean in §5, requires $\phi(x)$ to be a real function of the real variable x . If $\phi(x)$ is a complex function of a real variable, its real and imaginary parts separately will satisfy the assumption of existence and continuity of derivatives or of analyticity stated for $\phi(x)$ so that the demonstrations given are valid for these parts. Combining the two results thus obtained therefore yields the same result for the complex $\phi(x)$.

If δ is a real fixed quantity in the analytic sector of $f(z)$, we can choose for the C''' of §3 a circle, center $1/\Delta x$, radius $1/\Delta x - \delta$, and so, through the corresponding contour integral and condition (b) of §3, obtain

$$|A[f](r, \Delta x)| < \frac{Ke^{\kappa(2/\Delta x - \delta)}}{(1 - \delta\Delta x)^r}.$$

$\sum_{q=0}^{\infty} |A[f](r, \Delta x)| \Delta x$ is then less than $e^{2\kappa/\Delta x}$ times the sum of a geometric progression which is easily seen to have a finite limit as $\Delta x \rightarrow +0$. The essential factor is thus $e^{2\kappa/\Delta x}$. Now the inequalities imposed on ϵ can be re-written

$$0 < \frac{\epsilon + (X - x_1)}{l - \epsilon} < 1.$$

Hence, by choosing κ sufficiently small, we obtain

$$\lim_{\Delta x \rightarrow +0} m \left[\frac{\epsilon + (X - x_1)}{l - \epsilon} \right]^m e^{2\kappa/\Delta x} = 0, \quad m = \left\{ \frac{\epsilon}{\Delta x} \right\},$$

thereby establishing (b).

(c) As in part (b), we obtain

$$\left| \frac{\Delta^\mu \phi(x'_1)}{\Delta x^\mu} \right| < \frac{M\mu!}{(l - \epsilon)^\mu}.$$

On the other hand

$$A[u^{-\mu-1}f(u)](q, \Delta x) = - \frac{1}{2\pi i} \int_W \frac{z^{-\mu-1}f(z)}{(1 - z\Delta x)^{q+1}} dz,$$

where W encloses $1/\Delta x$, but excludes the origin. The contour W will consist of W_1 , that part of the C''' of §3 for which $|z| > R$, $R < 1/\Delta x$, joined to W_2 , the arc of the circle center the origin, radius R , cut off by C''' , and excluding, with W_1 , the origin. Such a contour will be valid when R exceeds the distance of the vertex of C from the origin. We have

$$\left| \frac{1}{2\pi} \int_{W_2} \frac{z^{-\mu-1}f(z)}{(1 - z\Delta x)^{q+1}} dz \right| < K \frac{R^{-\mu} e^{\kappa R}}{(1 - R\Delta x)^{q+1}}.$$

Now $R^{-\mu}/(1 - R\Delta x)^{q+1}$ has a minimum with respect to R for

$$R = \frac{\mu}{(q+1)\Delta x + \mu\Delta x}.$$

This value of R is less than $1/\Delta x$, and exceeds $\mu/[(X - x_1) + \epsilon]$, so that, for sufficiently large μ , it will give a valid contour. With it we have

$$\frac{R^{-\mu} e^{\kappa R}}{(1 - R\Delta x)^{q+1}} = \frac{[(q+1)\Delta x + \mu\Delta x]^\mu}{\mu^\mu} e^{\kappa\mu/(q+1)\Delta x + \mu\Delta x} \left(1 + \frac{\mu}{q+1}\right)^{q+1},$$

and, as $\mu < m$, we obtain

$$\left| \frac{1}{2\pi} \int_{W_1} \frac{z^{-\mu-1} f(z)}{(1 - z\Delta x)^{q+1}} dz \right| < K [(X - x_1) + \epsilon]^\mu e^{\kappa\mu/(X-x_1)} \mu^{-\mu} e^\mu.$$

For W_1 we have $|z| > R$. Since W_1 is part of C''' , we easily find that

$$\left| \frac{1}{2\pi} \int_{W_1} \frac{z^{-\mu-1} f(z)}{(1 - z\Delta x)^{q+1}} dz \right| < \frac{[(X - x_1) + \epsilon]^{\mu+1}}{\mu^{\mu+1}} \cdot \frac{1}{2\pi} \int_{C'''} \frac{|f(z)|}{|1 - z\Delta x|^{q+1}} |dz|.$$

Now our discussion of $A[f](r, \Delta x)$ in §3 proves that this integral has a finite limit as $\Delta x \rightarrow +0$. Symbolizing the product of $1/(2\pi)$ and this limit by $||A[f](X-x_1)||$, we are thus led to the following form for the $h(N)$ of the statement of this part of the proof, viz.,

$$h(N) = MK \sum_{\mu=N}^{\infty} \left\{ \left[\frac{(X - x_1) + \epsilon}{l - \epsilon} \right]^\mu e^{\kappa\mu/(X-x_1)} \frac{\mu!}{\mu^\mu e^{-\mu}} \right\} \\ + M(l - \epsilon) ||A[f](X-x_1)|| \sum_{\mu=N}^{\infty} \left\{ \left[\frac{(X - x_1) + \epsilon}{l - \epsilon} \right]^{\mu+1} \frac{\mu!}{\mu^{\mu+1}} \right\}.$$

Recalling that we have $0 < [(X - x_1) + \epsilon]/(l - \epsilon) < 1$, we see that the second series converges. By choosing κ sufficiently small to have

$$\frac{(X - x_1) + \epsilon}{l - \epsilon} e^{\kappa/(X-x_1)} < 1,$$

the convergence of the first series is assured. As a consequence of their convergence, these series approach zero as limit as N increases indefinitely. The same is therefore true of $h(N)$. (c) has thus been proved, and with it Theorem VI.

We turn now to the formal properties mentioned in the introduction.

7. Differentiation with respect to the upper limit. We can obtain the derivative of $A[f](t)$ under a general hypothesis. If $A[f(u)](t)$ and $A[uf(u)](t)$ exist for $0 < t_1 \leq t < t_2$,* then the terms of the relation

$$\{Df(D)\}_{-t}^0 1 = \{Df(D)\}_{-t_1}^0 1 + \int_{-t}^{-t_1} A[uf(u)](-\xi) d\xi$$

* The relation given below shows that if $A[uf(u)](t)$ exists in a neighborhood, and $A[f(u)](t)$ exists for one t in that neighborhood, it exists for every t therein.

obtained from Theorem II, exist, since by (12) this relation becomes

$$A[f(u)](t) = A[f(u)](t_1) + \int_{t_1}^t A[uf(u)](\eta)d\eta.$$

By keeping t_1 constant, and differentiating with respect to t , we get

THEOREM VIII. *If $A[f(u)](t)$ and $A[uf(u)](t)$ exist in a certain neighborhood of t , then $(d/dt) A[f(u)](t)$ exists, and is given by*

$$(23) \quad \frac{d}{dt} A[f(u)](t) = A[uf(u)](t).$$

By the use of (9), we immediately obtain the

COROLLARY. *If $A[u^{-1}f(u)](X-x_0)$ and $A[f(u)](X-x_0)$ exist in a certain neighborhood of X , then $(d/dX) \{f(D)\}_{x_0}^X$ exists, and is given by*

$$(24) \quad \frac{d}{dX} \{f(D)\}_{x_0}^X = \{Df(D)\}_{x_0}^X.$$

In extending this corollary to $\phi(x)$ as operand, it is desirable to replace the X left neighborhood of our existence theorems by a complete neighborhood, so that d/dX can stand for the derivative as ordinarily used. If, however, we retain the left neighborhood, then the following still remains valid, provided d/dX is understood to mean left derivative, an observation that is essential to most of our applications of these formulas.

For $f(D)$ either of finite order, or of type zero, we have, as a result of (23),

$$\frac{d}{dX} \int_{x_0}^{x_1} A[f(u)](X-x)\phi(x)dx = \int_{x_0}^{x_1} A[uf(u)](X-x)\phi(x)dx,$$

since the continuity of $A[uf(u)](X-x)$, for $x_0 \leq x \leq x_1$, and a neighborhood of X , permits differentiation under the integral sign. Hence, by (9), we shall have $(d/dX) \{f(D)\}_{x_0}^X \phi(x) = \{Df(D)\}_{x_0}^X \phi(x)$, provided we first prove this relation for lower limit x_1 . We shall choose x_1 , so that the special expansions hold.

First let $Df(D)$ be of order zero. $f(D)$ is then also of order zero, so that we can use (15), and write

$$\{f(D)\}_{x_0}^X \phi(x) = \lim_{h \rightarrow +0} \int_{x_1}^{X-h} A[f(u)](X-x)\phi(x)dx.$$

By Leibnitz's rule we have

$$\frac{d}{dX} \int_{x_1}^{X-h} A[f(u)](X-x)\phi(x)dx = A[f(u)](h)\phi(X-h) + \int_{x_1}^{X-h} A[uf(u)](X-x)\phi(x)dx.$$

Furthermore, by (14), we have

$$|A[uf(u)](X-x)| \leq L(X-x)^{-1+\epsilon}, \quad |A[f(u)](h)| \leq L'h^{*\epsilon}.$$

These inequalities show that the result of the last differentiation uniformly approaches $\int_{x_1}^X A[uf(u)](X-x)\phi(x)dx$ as limit in a neighborhood of X , as $h \rightarrow +0$. Hence by a well known criterion of differentiation,[†] we have, as desired,

$$\frac{d}{dX} \{f(D)\}_{x_1}^X \phi(x) = \int_{x_1}^X A[uf(u)](X-x)\phi(x)dx.$$

If now $Df(D)$ is of order ρ , we can use the first existence theorem for operators of finite order with $f(D)$ as of order ρ , and write

$$\{f(D)\}_{x_1}^X \phi(x) = \{D^{-\rho} f(D)\}_{x_1}^X \phi^{(\rho)}(x) + \sum_{\mu=0}^{\rho-1} A[u^{-\mu-1} f(u)](X-x_1) \phi^{(\mu)}(x_1).$$

Since $D \cdot D^{-\rho} f(D)$ is of order zero, we can use the special case just proved, along with (23), in differentiating both members of this equation, thereby obtaining on the right the expansion of $\{Df(D)\}_{x_1}^X \phi(x)$. Hence

THEOREM IX. *If $Df(D)$ is of order ρ , and if $\phi(x)$ is continuous in (x_0, X) , and has a continuous ρ th derivative in a neighborhood of X , then*

$$(25) \quad \frac{d}{dX} \{f(D)\}_{x_0}^X \phi(x) = \{Df(D)\}_{x_0}^X \phi(x).$$

COROLLARY. *If $D^n f(D)$ is of order ρ , and if $\phi(x)$ is continuous in (x_0, X) , and has a continuous ρ th derivative in a neighborhood of X , then*

$$(26) \quad \frac{d^n}{dX^n} \{f(D)\}_{x_0}^X \phi(x) = \{D^n f(D)\}_{x_0}^X \phi(x).$$

If $f(D)$ is replaced by $D^{-\rho} f(D)$, and n by ρ , the hypothesis of this corollary, restricted to a left neighborhood of X , reduces to the hypothesis of Theorem IV. Under this hypothesis, (26) therefore reduces to

* The first because $\rho=0$ for $uf(u)$; the second because the ρ of $f(u)$ may be taken to be -1 . Though ρ , as defined in §5, is either zero or a positive integer, it can be assigned a negative value, as is convenient here, with (14) remaining valid, provided a positive ϵ can still be chosen.

† Goursat, *Cours d'Analyse*, vol. 1, 1910, p. 74.

$$(27) \quad \{f(D)\}_{x_0}^X \phi(x) = \frac{d^\rho}{dX^\rho} \int_{x_0}^X A[u^{-\rho} f(u)](X-x)\phi(x) dx.$$

Formula (27) is a direct extension of Riemann's form for D^n for $R(n) > 0$.

The extension of Theorem IX to operators of type zero offers no difficulty. We first observe that the result of differentiating series (21) term by term is the expansion of $\{Df(D)\}_{x_0}^X \phi(x)$. Hence it is only necessary to prove the resulting series uniformly convergent in a neighborhood of X to obtain the desired result. Since $Df(D)$ is also of type zero, no loss of generality ensues if we prove the series for $\{f(D)\}_{x_0}^X \phi(x)$ uniformly convergent. Turning to the discussion of (c) §6, we observe that the single series in which $h(N)$ can be written is a majorant for the corresponding part of the series for $\{f(D)\}_{x_0}^X \phi(x)$. Now the discussion of §3 shows that we can write

$$\|A[f](X-x_0)\| = \frac{1}{2\pi} \int_C e^{R(z)(X-x_0)} |f(z)| |dz|.$$

This integral is continuous in X , and hence bounded for the neighborhood of X in question. On the other hand, through the choice of κ made in (c) §6, we shall have, for a sufficiently small neighborhood of X ,

$$0 < \frac{(X-x_0)+\epsilon}{l-\epsilon} e^{\kappa/(X-x_0)} < \lambda < 1.$$

As a result, the terms of $\{f(D)\}_{x_0}^X \phi(x)$ will be less in absolute value than those of a convergent series of positive constants, and so the series is uniformly convergent. We thus have

THEOREM X. *If $f(D)$ is of type zero, and $\phi(x)$ satisfies the hypothesis of Theorem VII, extended to a complete neighborhood of X , then $(d/dX)\{f(D)\}_{x_0}^X \phi(x)$ exists, and is given by (25).*

8. The law of successive operations. In the present section we shall consider certain conditions under which the relation

$$(28) \quad \{f_1(D)\}_{x_0}^X \{f_2(D)\}_{x_0}^z \phi(\xi) = \{f_1(D)f_2(D)\}_{x_0}^X \phi(x)$$

is valid. Inasmuch as $\{f_2(D)\}_{x_0}^z \phi(\xi)$, which serves as operand for $f_1(D)$, may be discontinuous at $x=x_0$, we shall introduce the following extension of our fundamental definition. If $\phi(x)$ is discontinuous for $x=x_0$, while $\{f(D)\}_{x_0+h}^X \phi(x)$ exists for sufficiently small h , then $\{f(D)\}_{x_0}^X \phi(x)$ is to be defined by

$$(29) \quad \{f(D)\}_{x_0}^X \phi(x) = \lim_{h \rightarrow +0} \{f(D)\}_{x_0+h}^X \phi(x)$$

provided the limit in question exists. Under the hypothesis of Theorem II, changed to allow for a discontinuity of $\phi(x)$ for $x=x_0$, the existence of this limit is equivalent to the convergence of the improper integral thus occurring in formula (9). This convergence is amply assured if $\phi(x)$ satisfies in a right neighborhood of x_0 the inequality

$$(30) \quad |\phi(x)| \leq M(x - x_0)^{-1+\eta}, \quad \eta > 0.$$

It will be convenient to say that $\phi(x)$ then has $M(x - x_0)^{-1+\eta}$ as majorant in the neighborhood. With (30) to replace continuity of $\phi(x)$ at $x=x_0$, Theorem II, with formula (9), continues to hold. Our existence theorems therefore go over, as do also the results of the preceding section.*

Consider, however, the special case $f_2(D) = D^n$, n other than zero, or a positive integer. We have, by (12) and (6),

$$\{D^n\}_{x_0}^x 1 = A[u^{n-1}](x - x_0) = (x - x_0)^{-n}/\Gamma(-n + 1).$$

This is discontinuous for $x=x_0$ when $R(n) > 0$, so that, without the above extension of our definition, (28) could only hold for $R(n) < 0$. With this extension, we can have $R(n) < 1$, since $\{D^n\}_{x_0}^x 1$ then satisfies (30). On the other hand, for $R(n) > 1$, let $f_1(D) = D^{-1}$. We have

$$\{D^{-1}\}_{x_0+h}^x \{D^n\}_{x_0}^x 1 = \int_{x_0+h}^x \frac{(x - x_0)^{-n}}{\Gamma(-n + 1)} dx,$$

which diverges as $h \rightarrow +0$. The left hand member of (28) therefore fails to exist in this case.

Volterra has encountered the same difficulty in the related theory of functions of composition; but his solution appears to the writer to be but a verbal evasion.[†] Among other possibilities, the difficulty might be removed by a study of the commonly neglected arbitrary series that Riemann adds to the definite integral in his formula for D^n . In the absence of a definitive solution, the writer leaves the breach open to view.[‡] This possible failure

* The differentiation of an improper integral under the integral sign that is required here is easily justified by the criterion of differentiation referred to in §7.

† V. Volterra, *Functions of composition*, The Rice Institute Pamphlet, vol. 7 (1920), p. 202.

‡ This difficulty does not appear in the theory for infinite lower limit; but the validity of our definition for finite lower limit is thereby rendered questionable. It is to be noted, however, that this failure occurs in the first place for the commonly accepted Riemann-Grünwald form for D^n . Furthermore, in all other respects the theory for finite lower limit is satisfactory; and if not for its own value, it would still be required as a foundation for the theory for infinite lower limit. Finally, in the various tentative modifications of the definition considered by the writer, the present theory remains the indispensable basis for the extension.

of (28), because of the non-existence of the second of the two successive operations, forces us to restrict $f_2(D)$ to finite orders zero or one, the latter only made possible by the extension of definition just given. Because of this restriction of $f_2(D)$, we shall only consider $f_1(D)$'s of finite order.

Our first consideration is to find conditions on $\phi(x)$ which will insure the existence of the left hand member of (28). For this purpose $f_1(D)$ may be taken of order zero, the extension to arbitrary finite order being easily made. We shall then want $\{f_2(D)\}_{x_0}^x \phi(\xi)$ to exist and be continuous for $x_0 < x \leq X$, and to satisfy in a right neighborhood of x_0 an inequality of type (30).

First let $f_2(D)$ be of order zero. Then, to insure the existence of $\{f_2(D)\}_{x_0}^x \phi(\xi)$, we shall want $\phi(\xi)$ to be continuous for $x_0 < \xi \leq X$, and to have $M(\xi - x_0)^{-1+\eta}$, $\eta > 0$, as majorant in a right neighborhood of x_0 . The same conditions turn out to be sufficient to yield the remaining requirements for $\{f_2(D)\}_{x_0}^x \phi(\xi)$. In fact, note that, due to its continuity, $\phi(\xi)$ will have $M'(\xi - x_0)^{-1+\eta}$ as majorant over the whole interval (x_0, X) . By applying (15), and using $Lt^{-1+\epsilon}$ as majorant for $A[f_2](t)$, in accordance with (14), we easily establish the continuity of $\{f_2(D)\}_{x_0}^x \phi(\xi)$. Furthermore, by the substitution of majorants, and reduction to the first Eulerian integral, (15) yields

$$|\{f_2(D)\}_{x_0}^x \phi(\xi)| \leq LM'B(\epsilon, \eta)(x - x_0)^{-1+\epsilon+\eta},$$

which is of type (30).

When $f_2(D)$ is of order one, the use of the first existence theorem for operators of finite order requires $\phi(\xi)$ to possess a continuous first derivative for $x_0 < \xi \leq X$, since x must vary from x_0 to X . This time, $\phi'(\xi)$ is assumed to have $M(\xi - x_0)^{-2+\eta}$, $\eta > 0$, as majorant. Since no loss of generality ensues if η is assumed less than one, $\phi(\xi)$ will then have $M'(\xi - x_0)^{-1+\eta}$ as majorant. By applying (17) with $x - x_1 = (x - x_0)/2$, i.e. by writing

$$\begin{aligned} \{f_2(D)\}_{x_0}^x \phi(\xi) &= \int_{x_1}^x A[u^{-1}f(u)](x - \xi)\phi'(\xi)d\xi + A[u^{-1}f_2(u)](x - x_1)\phi(x_1) \\ &\quad + \int_{x_0}^{x_1} A[f_2(u)](x - \xi)\phi(\xi)d\xi, \end{aligned}$$

and using the majorants $Lt^{-2+\epsilon}$ and $L't^{-1+\epsilon}$ for $A[f_2](t)$ and $A[u^{-1}f_2(u)](t)$ respectively, as given by (14), in conjunction with the above majorants for $\phi'(\xi)$ and $\phi(\xi)$, we establish the stated requirements for $\{f_2(D)\}_{x_0}^x \phi(\xi)$, under the added condition $\epsilon + \eta > 1$.

The desired conditions are therefore the following:

(a) for $f_2(D)$ of order zero: $\phi(x)$ is continuous for $x_0 < x \leq X$, with $M(x-x_0)^{-1+\eta}$, $\eta > 0$, as majorant;

(b) for $f_2(D)$ of order one: $\phi'(x)$ exists and is continuous for $x_0 < x \leq X$, with $M(x-x_0)^{-2+\eta}$ as majorant, where $\eta > 0$, and $\epsilon + \eta > 1$.

We proceed now to establish (28) for $f_1(D)$ of order zero under the above conditions on $\phi(x)$. The following preliminary result is fundamental.

As in the derivation of (4) §1, we find

$$(31) \quad A[f_1(u)f_2(u)](p, \Delta x) = \sum_{r=0}^p A[f_1(u)](r, \Delta x)A[f_2(u)](p-r, \Delta x)\Delta x.$$

If we let $p = \{(X-x_0)/\Delta x\}$, and have $\Delta x \rightarrow +0$, then, by a slight modification of the method used in establishing (15), we obtain

$$(32) \quad A[f_1(u)f_2(u)](X-x_0) = \int_{x_0}^X A[f_1(u)](X-x)A[f_2(u)](x-x_0)dx,*$$

provided $f_1(D)$ and $f_2(D)$ are both of order zero. By (15) this becomes

$$(33) \quad \{f_1(D)\}_{x_0}^X A[f_2(u)](x-x_0) = A[f_1(u)f_2(u)](X-x_0).$$

Let then $f_1(D)$ and $f_2(D)$ first both be of order zero, with $\phi(x)$ satisfying condition (a) given above. We can in this case apply (15), and write

$$\{f_1(D)\}_{x_0}^X \{f_2(D)\}_{x_0}^x \phi(\xi) = \int_{x_0}^X A[f_1(u)](X-x) \int_{x_0}^x A[f_2(u)](x-\xi) \phi(\xi) d\xi dx.$$

The improper double integral corresponding to this iterated integral exists, since the latter, and hence the former, is absolutely convergent.† Consequently the order of integration can be changed to give

$$\{f_1(D)\}_{x_0}^X \{f_2(D)\}_{x_0}^x \phi(\xi) = \int_{x_0}^X \left[\int_{\xi}^X A[f_1(u)](X-x) A[f_2(u)](x-\xi) dx \right] \phi(\xi) d\xi.$$

By (32) and (15) this reduces to the desired relation (28).

Now let $f_1(D)$ be of order zero, $f_2(D)$ of order one, with $\phi(x)$ satisfying condition (b). Let x_1 be chosen between x_0 and X . We have, by Theorem IV,

$$\begin{aligned} \{f_1(D)\}_{x_1}^X \{f_2(D)\}_{x_1}^x \phi(\xi) &= \{f_1(D)\}_{x_1}^X \{D^{-1}f_2(D)\}_{x_1}^x \phi'(\xi) \\ &\quad + \phi(x_1) \{f_1(D)\}_{x_1}^X A[u^{-1}f_2(u)](x-x_1). \end{aligned}$$

* This formula is of special interest in connection with Volterra's functions of composition of the closed cycle group. (See V. Volterra, loc. cit., pp. 181-251.) If $A[f_1(u)](X-x)$ is written $g_1(X-x)$, and $A[f_2(u)](x-x_0)$, $g_2(x-x_0)$, then (32) shows $A[f_1(u)f_2(u)](X-x_0)$ to be Volterra's $g_1 g_2^*$, so that this symbolic product of the g 's corresponds to the actual product of the related f 's.

† See de la Vallée Poussin, *Cours d'Analyse*, vol. 2, 1925, pp. 19-22.

Since $D^{-1}f_2(D)$ is of order zero, (28) can be applied to the first term, and (33) to the second term, of the right hand member to give

$$\begin{aligned}\{f_1(D)\}_{x_1}^X \{f_2(D)\}_{x_1}^x \phi(\xi) &= \{D^{-1}f_1(D)f_2(D)\}_{x_1}^X \phi'(x) + A[u^{-1}f_1(u)f_2(u)](X-x_1)\phi(x_1) \\ &= \{f_1(D)f_2(D)\}_{x_1}^X \phi(x).\end{aligned}$$

(29) then shows that to demonstrate (28) in this case, we need but establish the relation

$$\lim_{x_1 \rightarrow x_0} \{f_1(D)\}_{x_1}^X \{f_2(D)\}_{x_1}^x \phi(\xi) = \{f_1(D)\}_{x_0}^X \{f_2(D)\}_{x_0}^x \phi(\xi).$$

Now we have directly

$$\begin{aligned}\{f_1(D)\}_{x_0}^X \{f_2(D)\}_{x_0}^x \phi(\xi) &- \{f_1(D)\}_{x_1}^X \{f_2(D)\}_{x_1}^x \phi(\xi) \\ &= \int_{x_0}^{x_1} A[f_1(u)](X-x)[\{f_2(D)\}_{x_0}^x \phi(\xi)] dx \\ &\quad + \{f_1(D)\}_{x_1}^X \left[\int_{x_0}^{x_1} A[f_2(u)](x-\xi)\phi(\xi) d\xi \right].\end{aligned}$$

The first term of the second member of this equation is immediately seen to approach zero as limit as $x_1 \rightarrow x_0$. As for the second term, return to the discussion of the conditions imposed on $\phi(x)$. We can assume without loss of generality the inequalities $0 < \epsilon < 1$, $0 < \eta < 1$. Since $\epsilon + \eta > 1$ we can choose a positive λ such that $\lambda < \epsilon + \eta - 1$. We therefore also have $\lambda < \epsilon$. Then in connection with the integral appearing in the second term, we have for the corresponding majorants

$$\begin{aligned}\int_{x_0}^{x_1} (x-\xi)^{-2+\epsilon} (\xi-x_0)^{-1+\eta} d\xi &< (x-x_1)^{-1+\lambda} \int_{x_0}^{x_1} (x_1-\xi)^{-1+\epsilon-\lambda} (\xi-x_0)^{-1+\eta} d\xi \\ &= (x-x_1)^{-1+\lambda} (x_1-x_0)^{\epsilon+\eta-1-\lambda} B(\epsilon-\lambda, \eta).\end{aligned}$$

This second term will therefore be less in absolute value than

$$M''(X-x_1)^\mu (x_1-x_0)^{\epsilon+\eta-1-\lambda}.$$

Since $\epsilon + \eta - 1 - \lambda > 0$, this term also approaches zero as limit as $x_1 \rightarrow x_0$. The desired relation is thus established, and with it, (28).

The extension to $f_1(D)$ of arbitrary finite order ρ_1 , with $f_2(D)$ of order zero or one, is now easily made. $D^{-\rho_1}f_1(D)$ will be of order zero, so that, under the above conditions on $\phi(x)$,

$$\{D^{-\rho_1}f_1(D)\}_{x_0}^X \{f_2(D)\}_{x_0}^x \phi(\xi) = \{D^{-\rho_1}f_1(D)f_2(D)\}_{x_0}^X \phi(x).$$

We have observed that the results of §7 remain valid under (29) and (30). Note also that the derivative appearing in Theorem IX and Corollary I is continuous in a left neighborhood of $x=X$. Suppose then that $D^{\rho_1}f_2(D)$ is of order ρ , and that $\phi(x)$ possesses a continuous ρ th derivative in a left neighborhood of $x=X$. Then $\{f_2(D)\}_{x_0}^x \phi(\xi)$ will possess a continuous ρ_1 th derivative in a left neighborhood of $x=X$. Now $D^{\rho_1} \cdot D^{-\rho_1} f_1(D)$ is of order ρ_1 ; $D^{\rho_1} \cdot D^{-\rho_1} f_1(D) f_2(D)$ is of order not greater than ρ . Hence we can operate on both sides of the above equation by d^{ρ_1}/dX^{ρ_1} in accordance with the corollary of Theorem IX, and by (26) obtain (28).

These results can be stated as

THEOREM XI. *If $f_1(D)$ and $f_2(D)$ are of orders ρ_1 and ρ_2 respectively, with ρ_2 equal to zero or one, while $D^{\rho_1}f_2(D)$ is of order ρ ; and if $\phi(x)$ has a continuous ρ th derivative in a left neighborhood of $x=X$, a continuous ρ_2 th derivative for $x_0 < x \leq X$, while for this interval $|\phi^{(\rho_2)}(x)| \leq M(x-x_0)^{-\rho_2-1+\eta}$, where η is positive and such that $D^{-\eta}f_2(D)$ is of order zero,* then (28) is valid.*

The restriction of $f_2(D)$ to order zero or one was necessary in order that the result should apply to $\phi(x)=1$, the simplest case. However, by requiring $\phi(x)$ and a sufficient number of its derivatives to vanish for $x=x_0$, the order of $f_2(D)$ can be arbitrarily large. In fact under the specific conditions stated below, formula (17) gives

$$\{f_2(D)\}_{x_0}^x \phi(\xi) = \{D^{-\rho_2+1}f_2(D)\}_{x_0}^x \phi^{(\rho_2-1)}(\xi).$$

As $D^{-\rho_2+1}f_2(D)$ will then be of order one, the theorem applies. Hence the

COROLLARY. *The relation (28) holds for $\rho_2 > 1$, provided the corresponding conditions of the theorem are replaced by the following: $\phi^{(\rho_2-2)}(x)$ exists, and is continuous, for $x_0 \leq x \leq X$, with $\phi^{(\mu)}(x_0) = 0$, for $\mu = 0, 1, \dots, \rho_2 - 2$; $\phi^{(\rho_2)}(x)$ exists, and is continuous, for $x_0 < x \leq X$, with $|\phi^{(\rho_2)}(x)| \leq M(x-x_0)^{-2+\eta}$, where η is positive, and such that $D^{-\eta}f_2(D)$ is of order $\rho_2 - 1$.*

The reader can apply these results to more than two successive operations, and to the commutativity of the operators.

When either of the f 's is a polynomial, no restriction of the order of $f_2(D)$ is needed. In fact, if $P(d/dx)$ is a polynomial in d/dx we obtain, by (26),

$$(34) \quad P\left(\frac{d}{dX}\right) \{f(D)\}_{x_0}^x \phi(x) = \{P(D)f(D)\}_{x_0}^x \phi(x),$$

under the condition that, if $P(D)f(D)$ is of order ρ , $\phi(x)$ satisfies the hypothesis of Theorem IX. On the other hand, through (17), we get

* This is another way of stating the condition $\epsilon + \eta > 1$ used in the proof.

$$(35) \quad \left\{ f(D) \right\}_{x_0}^X P \left(\frac{d}{dx} \right) \phi(x) = \left\{ P(D)f(D) \right\}_{x_0}^X \phi(x) \\ - \sum_{\mu=0}^{n-1} A [u^{n-\mu-1} f(u)] (X - x_0) P_\mu(x_0),$$

where

$$P(D) = \sum_{\nu=0}^n a_\nu D^{n-\nu}, \quad P_\mu(x_0) = \sum_{\nu=0}^\mu a_\nu \phi^{(\mu-\nu)}(x_0).$$

For (35), $\phi(x)$ is to possess a continuous n th derivative in (x_0, X) , and a continuous ρ th derivative in a left neighborhood of X , where ρ is the order of $P(D)f(D)$. Formulas (34) and (35) can easily be extended to the case where $f(D)$ is of type zero. They probably also admit of extension to the case where $P(D)$ is not a polynomial, but of type zero over the plane. (See §14.) Though the disagreement between (35) and (28) was to be expected as a result of the finite difference reduction formula, it may suggest a solution of the difficulty discussed above.

9. Leibnitz's theorem generalized. The operators $f^n(D)$, $n = 1, 2, 3, \dots$, are intimately associated with the generalization of Leibnitz's Theorem for repeated differentiation of a product of two functions. A simple relation between $A[f^{(n)}(u)](t)$ and $A[f(u)](t)$ follows directly from our definition. We have

$$A[f'(u)](r, \Delta x) = \frac{(-1)^r f^{(r+1)}(1/\Delta x)}{r! \Delta x^{r+1}} = - (r+1) \Delta x A[f(u)](r+1, \Delta x).$$

Hence, for $t > 0$, $A[f'(u)](t)$ and $A[f(u)](t)$ coexist, and satisfy the relation

$$(36) \quad A[f'(u)](t) = - t A[f(u)](t). *$$

By induction, we obtain

$$(37) \quad A[f^{(n)}(u)](t) = (-t)^n A[f(u)](t).$$

It is easily proved by the use of Cauchy's integral formula with circular contour for $f^{(n)}(z)$, and the corresponding conditions on $f(z)$, that, if $f(z)$ is of type zero over a certain sector, $f^{(n)}(z)$ is also of type zero over any sector interior to the sector of $f(z)$, and of the same angle. Likewise it can be shown that if $f(z)$ is of finite order ρ , $f^{(n)}(z)$ is of order $\rho-n$, or zero (according as $\rho > n$, or $\leq n$), over any sector interior to that of $f(z)$, and of angle less than that of the sector of $f(z)$. However, the finite difference relation given above immediately shows that if $f(D)$ is of generalized order ρ , $f'(D)$ is of

* Comparison of this formula with (23) suggests a duality which is strikingly borne out, under special hypotheses, in a study of $A[f](t)$ for complex t , coupled with the extension to type one suggested in §17.

generalized order $\rho - 1$ (or zero if $\rho = 0$), with a corresponding extension to $f^{(n)}(D)$. A more complete statement is that $D^n f^{(n)}(D)$ is of the same order or generalized order as $f(D)$.

Turning now to the theorem desired, we shall first prove the special case

$$(38) \quad \{f(D)\}_{x_0}^X x\phi(x) = X \{f(D)\}_{x_0}^X \phi(x) + \{f'(D)\}_{x_0}^X \phi(x),$$

where $\phi(x)$ satisfies the conditions associated in our theorems with the existence of $\{f(D)\}_{x_0}^X \phi(x)$. We have, by definition,

$$\begin{aligned} \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X x\phi(x) &= \sum_{r=0}^{((X-x_0)/\Delta x)} \frac{(-1)^r f^{(r)}(1/\Delta x)}{r! \Delta x^r} (X - r\Delta x)\phi(X - r\Delta x) \\ &= X \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X \phi(x) + \left\{f'\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X-\Delta x} \phi(x) \\ &= X \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X \phi(x) + \left\{f'\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X \phi(x) \\ &\quad - \Delta x \left\{\frac{\Delta}{\Delta x} f'\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X \phi(x). \end{aligned}$$

Since $Df'(D)$ is of the same type, or finite order, as $f(D)$, $\{Df'(D)\}_{x_0}^X \phi(x)$ will exist, so that the last term vanishes with Δx . Hence (38).

We can rewrite (38) in the form

$$\{f(D)\}_{x_0}^X (x - X)\phi(x) = \{f'(D)\}_{x_0}^X \phi(x),$$

and obtain, by induction,

$$(39) \quad \{f(D)\}_{x_0}^X (x - X)^n \phi(x) = \{f^{(n)}(D)\}_{x_0}^X \phi(x).$$

If then $P(x)$ is a polynomial of degree m , we can expand it in powers of $(x - X)$, and obtain, by (39), the terminating Leibnitz expansion

$$\begin{aligned} \{f(D)\}_{x_0}^X P(x)\phi(x) &= P(X) \{f(D)\}_{x_0}^X \phi(x) + \frac{P'(X)}{1!} \{f'(D)\}_{x_0}^X \phi(x) \\ &\quad + \cdots + \frac{P^{(m)}(X)}{m!} \{f^{(m)}(D)\}_{x_0}^X \phi(x). \end{aligned}$$

This method will now be extended to $\{f(D)\}_{x_0}^X \psi(x)\phi(x)$.

We shall assume $\psi(x)$ to be analytic in the closed interval (x_0, X) with the radius of convergence of its Taylor expansion at $x = X$ greater than $(X - x_0)$. Then, for this interval, we have

$$\psi(x) = \psi(X) + \frac{\psi'(X)}{1!}(x - X) + \frac{\psi''(X)}{2!}(x - X)^2 + \cdots.$$

When $f(D)$ is of finite order ρ , $f^{(\rho)}(D)$ is of order zero. Writing the remainder after ρ terms of the $\psi(x)$ expansion $(x-X)^{\rho}\chi(x)$ we have, by (39),

$$\{f(D)\}_{x_0}^X(x-X)^{\rho}\chi(x)\phi(x) = \{f^{(\rho)}(D)\}_{x_0}^X\chi(x)\phi(x).$$

Since $f^{(\rho)}(D)$ is of order zero, the integral formula (15) is applicable. Furthermore, as the series for $\chi(x)$ is uniformly convergent over (x_0, X) , while the integral for $\{f^{(\rho)}(D)\}_{x_0}^X\phi(x)$ is absolutely convergent, we can integrate term by term. This is the same as operating with $f^{(\rho)}(D)$ term by term, or with $f(D)$ in the original series. We thus obtain, by (39), the generalized Leibnitz expansion

$$(40) \quad \{f(D)\}_{x_0}^X\psi(x)\phi(x) = \sum_{n=0}^{\infty} \frac{\psi^{(n)}(X)}{n!} \{f^{(n)}(D)\}_{x_0}^X\phi(x).$$

For $f(D)$ of type zero, we will designate the remainder after m terms of the $\psi(x)$ series by $R_m(x)$, and prove directly that

$$\lim_{m \rightarrow \infty} \{f(D)\}_{x_0}^X R_m(x)\phi(x) = 0.$$

Applying Theorem II to this expression, we observe that $R_m(x)$ uniformly approaches zero as limit along (x_0, x_1) as $m \rightarrow \infty$, so that the integral likewise has zero for limit. On the other hand, for x_1 sufficiently near X , the expansion of Theorem VI is applicable to $\{f(D)\}_{x_1}^X R_m(x)\phi(x)$. The first N terms can be directly treated. On the other hand, the $h(N)$ of part (c) of §6, with $R_m(x)\phi(x)$ in place of $\phi(x)$, is an upper bound for the absolute value of the rest of this expansion. We thus see that

$$|\{f(D)\}_{x_0}^X R_m(x)\phi(x)| \leq M_m L_m,$$

where M_m is the upper bound of the function of the complex variable $R_m(z)\phi(z)$ over the circle center x_1 , radius l , while L_m depends only on $f(u)$, $(X-x_1)$, and the radius of convergence of the expansion of $R_m(z)\phi(z)$ about $z=x_1$. Now the radius of convergence of $R_m(z)$ is the same as that of $\psi(z)$, so that L_m is independent of m . On the other hand, $R_m(z)$ uniformly approaches zero over the circle of radius l , so that $M_m \rightarrow 0$, as $m \rightarrow \infty$. The limit in question is thus proved to be zero, and we can state

THEOREM XII. *If $\{f(D)\}_{x_0}^X\phi(x)$ exists under the hypotheses of any of our existence theorems, and if $\psi(x)$ is analytic along the closed interval (x_0, X) , with the radius of convergence of its Taylor expansion at $x=X$ greater than $(X-x_0)$, then $\{f(D)\}_{x_0}^X\psi(x)\phi(x)$ is given by the generalized Leibnitz expansion (40).*

As a corollary, we have, by letting $\phi(x)$ identically equal 1,

$$(41) \quad \{f(D)\}_{x_0}^X \psi(x) = \psi(X) \{f(D)\}_{x_0}^X 1 + \frac{\psi'(X)}{1!} \{f'(D)\}_{x_0}^X 1 + \dots,$$

which is to be compared with series (21), which may be written

$$(42) \quad \{f(D)\}_{x_0}^X \psi(x) = \psi(x_0) \{f(D)\}_{x_0}^X 1 + \psi'(x_0) \{D^{-1}f(D)\}_{x_0}^X 1 + \dots.$$

In each case $\psi(x)$ is analytic along the closed interval (x_0, X) ; but in the first the radius of convergence at $x=X$ exceeds $X-x_0$, in the second that at $x=x_0$ exceeds $X-x_0$.

We reserve for §14 the application of (41) to entire operators.

10. $e^{ax}\phi(x)$ as operand; $f(D+a)$. From our definition we have

$$A[f(u+a)](r, \Delta x) = \frac{(-1)^r f^{(r)}\left(\frac{1}{\Delta x} + a\right)}{r! \Delta x^{r+1}}.$$

For real a , and sufficiently small Δx , the equation

$$\frac{1}{\Delta x} + a = \frac{1}{\Delta_1 x}$$

results in a positive $\Delta_1 x$. We can then write

$$A[f(u+a)](r, \Delta x) = (1 - a\Delta_1 x)^{r+1} A[f(u)](r, \Delta_1 x).$$

As $\Delta x \rightarrow +0$, and $r\Delta x \rightarrow t$, we have simultaneously $\Delta_1 x \rightarrow +0$, $r\Delta_1 x \rightarrow t$. Hence under the sole condition that a is real, we obtain

$$(43) \quad A[f(u+a)](t) = e^{-at} A[f(u)](t).$$

When $f(z)$ is of type zero, and hence also when it is of finite order, this result is obtained immediately, for a both real and complex, by the use of the contour integral formula (8) of Theorem I.

The theorem of this section will be restricted to $f(D)$ of finite order. When $f(D)$ is of order zero, $f(D+a)$ also is. By (15), we have

$$\begin{aligned} \{f(D)\}_{x_0}^X e^{ax}\phi(x) &= \int_{x_0}^X A[f(u)](X-x) e^{ax}\phi(x) dx \\ &= e^{aX} \int_{x_0}^X e^{-a(X-x)} A[f(u)](X-x) \phi(x) dx, \end{aligned}$$

and so, by (43), and (15) again, we obtain

$$(44) \quad \{f(D)\}_{x_0}^X e^{ax}\phi(x) = e^{aX} \{f(D+a)\}_{x_0}^X \phi(x).$$

When $f(D)$ is of order ρ , $D^{-\rho}f(D)$ is of order zero. We then have

$$\begin{aligned}\{f(D)\}_{z_0}^X e^{az} \phi(x) &= d^\rho/dX^\rho \{D^{-\rho}f(D)\}_{z_0}^X e^{az} \phi(x) \\ &= e^{aX} (d/dX + a)^\rho \{(D + a)^{-\rho} f(D + a)\}_{z_0}^X \phi(x) \\ &= e^{aX} \{f(D + a)\}_{z_0}^X \phi(x);\end{aligned}$$

the first by (26), the second by (44) for operators of order zero and polynomials in d/dX , the last by (34). We can therefore state

THEOREM XIII. *If $f(D)$ is of finite order, and $\phi(x)$ satisfies the hypothesis of the existence Theorem IV, then relation (44) is valid.*

By letting $\phi(x) = 1$, (44) and (8) yield a contour integral for $\{f(D)\}_{z_0}^X e^{az}$.

11. $x_0 = -\infty$; $\phi(x) = e^{ax}$. Let $f(z)$ be analytic to the right of the line $R(z) = c$. Then

$$\begin{aligned}\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{-\infty}^X e^{ax} &= f\left(\frac{1}{\Delta x}\right) e^{aX} - \frac{f'\left(\frac{1}{\Delta x}\right)}{1! \Delta x} e^{a(X-\Delta x)} + \dots \\ &= e^{aX} f\left(\frac{1}{\Delta x} - \frac{e^{-a\Delta x}}{\Delta x}\right),\end{aligned}$$

provided $f(1/\Delta x - e^{-a\Delta x}/\Delta x)$ can be expanded in powers of $e^{-a\Delta x}/\Delta x$. Now

$$|e^{-a\Delta x}/\Delta x| = e^{-R(a)\Delta x}/\Delta x = 1/\Delta x - R(a) + \epsilon_{\Delta x}, \quad \lim_{\Delta x \rightarrow +0} \epsilon_{\Delta x} = 0.$$

Hence if $R(a) > c$, this absolute value will be less than $1/\Delta x - c$, for Δx sufficiently small. As the radius of convergence of the Taylor expansion of $f(z)$ about $1/\Delta x$ is at least $1/\Delta x - c$, the above is valid, so that

$$\{f(D)\}_{-\infty}^X e^{ax} = \lim_{\Delta x \rightarrow +0} e^{aX} f\left(\frac{1 - e^{-a\Delta x}}{\Delta x}\right) = e^{aX} f(a).$$

Conversely, if $f(z)$ is an analytic function of z , and $\{f(D)\}_{-\infty}^X e^{az}$ exists for some a , $\{f(\Delta/\Delta x)\}_{-\infty}^X e^{az}$ must converge for all positive Δx 's less than some positive ϵ . Then $f(z)$ will be analytic within all circles center $1/\Delta x$, radius $|e^{-a\Delta x}/\Delta x|$ where $0 < \Delta x < \epsilon$. As these circular regions cover the half-plane $R(z) > R(a)$, $f(z)$ is analytic to the right of the line $R(z) = R(a)$. Hence

THEOREM XIV. *The necessary and sufficient condition that $\{f(D)\}_{-\infty}^X e^{az}$ exist for some a , when $f(z)$ is an analytic function of z , is that $f(z)$ is analytic to the right of some line $R(z) = c$. In that case $\{f(D)\}_{-\infty}^X e^{az}$ exists for $R(a) > c$, and is given by*

$$(45) \quad \{f(D)\}_{-\infty}^X e^{az} = e^{aX} f(a).$$

If $c < 0$, we can let $a = 0$, and so obtain

$$(46) \quad \{f(D)\}_{-\infty}^x 1 = f(0).$$

The above results were obtained without assuming $f(z)$ of type zero, though the analyticity condition of Theorem XIV is a weaker form of the analyticity condition for an operator of type zero. If $f(z)$ is of type zero, we can apply Theorem XV of the next section to obtain

$$(47) \quad \lim_{x_0 \rightarrow -\infty} \{f(D)\}_{x_0}^x e^{ax} = \{f(D)\}_{-\infty}^x e^{ax}.$$

This can also be obtained directly by expressing the remainder after p terms of the Taylor expansion $\{f(\Delta/\Delta x)\}_{-\infty}^x e^{ax}$ as a contour integral, finding for its limit as $\Delta x \rightarrow +0$, $p = \{(X-x_0)/\Delta x\}$, an integral like that of formula (8), and observing that this approaches zero as limit as $x_0 \rightarrow -\infty$. If we further restrict $f(z)$ to order zero, we can use (15) in (47), and by the change of variable $X-x=t$, obtain through (45)

$$(48) \quad \int_0^\infty e^{-at} A[f](t) dt = f(a). *$$

That is, $A[f](t)$ satisfies the Laplace integral equation when $f(z)$ is of order zero. The derivation of (48) suggests that we may consider relation (45) a generalization of the Laplace integral equation.

12. $x_0 = -\infty$: general case. Our definition, for $x_0 = -\infty$, is

$$\{f(D)\}_{-\infty}^x \phi(x) = \lim_{\Delta x \rightarrow +0} \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{-\infty}^x \phi(x).$$

Hence the relation

$$(49) \quad \{f(D)\}_{-\infty}^x \phi(x) = \lim_{x_0 \rightarrow -\infty} \{f(D)\}_{x_0}^x \phi(x)$$

is subject to proof. Since, in our definition, we consider the limit of an infinite series, some restriction on the behavior of $\phi(x)$ as $x \rightarrow -\infty$ will have to be introduced to insure the existence of the limits involved. The same condition will turn out to be adequate for (49).

We assume $f(D)$ to be of type zero. Then $f(z)$ will certainly be analytic

* When $A[f](t)$ can be directly found from its definition, formula (48) leads to the evaluation of a corresponding definite integral. Thus, (6), the formula for $A[u^n](t)$, yields the well known infinite integral for the gamma function for $R(n) < 0$, while (3), as transformed by (12), results in an infinite integral for the Eulerian constant γ . In a similar manner, formula (8) leads to the evaluation of corresponding contour integrals.

to the right of some line $R(z) = c$. Let d designate the lower limit of these c 's. Then, for Δx sufficiently small, and $r\Delta x$ greater than a positive h , there is a positive L for each b greater than d such that

$$(50) \quad |A[f](r, \Delta x)| < Le^{b(r\Delta x)}.$$

To prove this fundamental inequality, we reconsider the discussion of the contour integral for $A[f](r, \Delta x)$ given in §3. For Δx sufficiently small, and $r\Delta x > h$, r will be sufficiently large to have, along C'' ,

$$\left| \int_{C''} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| < L_1 \lambda^r, \quad 0 < \lambda < 1.$$

Since, for Δx sufficiently small, we shall have λ less than $e^{b\Delta x}$ even with b negative, a relation like (50) holds for this contribution to $A[f](r, \Delta x)$. Along $C'_{l,m}$, we can have, with l and m sufficiently large, $R(z) < b'$, where $b' < b$, and $b' < 0$. We can then write

$$\left| \int_{C'_{l,m}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| < \frac{1}{(1 - b'\Delta x)^{r+1-(h/\Delta x)}} \int_{C'_{l,m}} \frac{|f(z)||dz|}{|1 - z\Delta x|^{(h/\Delta x)}}.$$

The latter integral has a finite upper bound, for Δx sufficiently small, in accordance with §3. On the other hand, since $b' < b$, we have, again for sufficiently small Δx ,

$$\frac{1}{1 - b'\Delta x} < e^{b\Delta x}.$$

Hence an inequality (50) exists for the $C'_{l,m}$ contribution. Finally, we may have to change $C_{l,m}$ so that it will consist of the segment of $R(z) = b''$, $d < b'' < b$, cut off by the former $C_{l,m}$, joined to the part of the former $C_{l,m}$ to the left of $R(z) = b''$. Since along this new $C_{l,m}$ we have

$$|1 - z\Delta x| \geq 1 - b''\Delta x,$$

we find that

$$\left| \int_{C_{l,m}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| \leq \frac{\bar{l}K'}{(1 - b''\Delta x)^{r+1}},$$

where \bar{l} is the length of $C_{l,m}$, and K' the upper bound of $|f(z)|$ along it. Hence, as in the case of $C'_{l,m}$, a relation (50) holds. By combining these three results we get (50) itself.

If in (50) we let $\Delta x \rightarrow +0$, $r\Delta x \rightarrow t$, we obtain, for $t > h$,

$$(51) \quad |A[f](t)| \leq Le^{bt}.$$

We can now prove

THEOREM XV. *If $f(D)$ is of type zero, with specified d , and if, for some positive M , and real c greater than d , $\phi(x)$ satisfies the inequality*

$$|\phi(x)| \leq M e^{cx}$$

for every x less than some real x_1 , relation (49) will be valid.

Choose b between c and d . By (50), and the condition on $\phi(x)$, we have

$$\left| \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{-\infty}^X \phi(x) - \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) \right| \\ < LM e^{cx} \sum_{r=p+1}^{\infty} e^{-(c-b)r\Delta x} \Delta x, \quad p = \left\{ \frac{X - x_0}{\Delta x} \right\}$$

provided $x_0 < x_1$, and also $x_0 < X - h$. Since this geometric progression converges, $\{f(\Delta/\Delta x)\}_{-\infty}^X \phi(x)$ also converges. Furthermore, on summing this progression, we find its limit, as $\Delta x \rightarrow +0$, to be $[LM/(c-b)]e^{cx}e^{-(c-b)(X-x_0)}$. This expression approaches zero as limit as $x_0 \rightarrow -\infty$. We can therefore apply the limit criterion, with x_0 in place of v , to the identity

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{-\infty}^X \phi(x) = \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) \\ + \left[\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{-\infty}^X \phi(x) - \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^X \phi(x) \right],$$

and obtain the theorem.

The use of the limit criterion assumes the existence of $\{f(D)\}_{x_0}^X \phi(x)$ for all x_0 's less than X . Granting this, its corollary gives the existence of both members of (49). Hence

COROLLARY I. *If $\phi(x)$ is continuous for $x < X$, and satisfies the X -neighborhood conditions of any of our existence theorems for finite x_0 , then, under the above hypothesis, both members of (49) exist, and are finite.*

It is also not difficult to establish

COROLLARY II. *If the X -neighborhood condition of Corollary I is not known to be fulfilled, but instead $\{f(D)\}_{-\infty}^X \phi(x)$ is known to exist, then the existence of $\{f(D)\}_{x_0}^X \phi(x)$ for finite x_0 follows.*

The rest of this section will be devoted to the law of successive operations for $x_0 = -\infty$. For finite x_0 , the order of $f_2(D)$ had to be zero or one, unless $\phi(x)$, and a sufficient number of its derivatives, vanished at $x = x_0$. In the present case this restriction on the order of $f_2(D)$ disappears. We shall, however, still restrict ourselves to operators of finite order.

First, in Theorem XV, let $f(D)$ be of order zero. Then, by (49) and (15), we obtain

$$(52) \quad \{f(D)\}_{-\infty}^X \phi(x) = \int_{-\infty}^X A[f](X-x)\phi(x)dx.$$

Break up this integral over limits $(-\infty, X-a)$, and $(X-a, X)$. By applying to the resulting integrals inequality (51), with b between c and d , and the hypothesis of Theorem XV, we obtain, for $X < x_1$, and $a > h$,

$$\left| \int_{-\infty}^X A[f](X-x)\phi(x)dx \right| < M e^{cX+|c|a} \int_0^a |A[f](t)| dt + [LM/(c-b)]e^{-a(c-b)} \cdot e^{cX} = M' e^{cX}.$$

That is, $|\{f(D)\}_{-\infty}^X \phi(x)|$ satisfies the same inequality that $|\phi(x)|$ satisfies except for the factor M . It is also easily seen that $\{f(D)\}_{-\infty}^X \phi(x)$ is continuous in X .

Let then $f_1(D)$ and $f_2(D)$ both be of order zero, with d 's equal to d_1 and d_2 respectively. If, in the hypothesis of Theorem XV, we have $c > d_1$ as well as $c > d_2$, our last result shows that the existent $\{f_2(D)\}_{-\infty}^X \phi(x)$ can be used as operand for $\{f_1(D)\}_{-\infty}^X$, so that we can write

$$\{f_1(D)\}_{-\infty}^X \{f_2(D)\}_{-\infty}^x \phi(\xi) = \int_{-\infty}^X A[f_1](X-x) \left[\int_{-\infty}^x A[f_2](x-\xi)\phi(\xi)d\xi \right] dx.$$

Since this iterated integral is absolutely convergent, the corresponding improper double integral exists.* It can therefore be rewritten

$$\int_{-\infty}^X \left[\int_x^X A[f_1](X-\xi)A[f_2](\xi-x)d\xi \right] \phi(x)dx.$$

Now the d of $f_1(D)f_2(D)$ does not exceed both d_1 and d_2 . We can therefore use (52), with $f_1(D)f_2(D)$ as operator, $\phi(x)$ as operand, and, by applying (32), obtain

$$(53) \quad \{f_1(D)\}_{-\infty}^X \{f_2(D)\}_{-\infty}^x \phi(\xi) = \{f_1(D)f_2(D)\}_{-\infty}^X \phi(x).$$

Before (53) can be extended to operators of arbitrary finite order, several preliminary results must be obtained. First, we have

$$(54) \quad \frac{d}{dX} \{f(D)\}_{-\infty}^X \phi(x) = \{Df(D)\}_{-\infty}^X \phi(x)$$

under the same X -neighborhood conditions as for finite x_0 , along with the

* The reasoning is the same as in §8 for which a reference has been given.

usual inequality on $|\phi(x)|$.* In fact, our inequalities easily show that the integral resulting from differentiating $\int_{-\infty}^{x_2} A[f](X-x)\phi(x)dx$ under the integral sign is uniformly convergent in X . Secondly, we have to discard the reduction formula (17), since the d of $D^{-\rho}f(D)$ will be greater than that of $f(D)$ when the latter is negative and ρ is not zero. By §10 and formula (17) with x_1 and x_0 identical, we obtain in its place

$$(55) \quad \begin{aligned} \{f(D)\}_{x_0}^X \phi(x) &= \{(D-l)^{-\rho}f(D)\}_{x_0}^X \left(\frac{d}{dx} - l \right)^\rho \phi(x) \\ &+ \sum_{\mu=0}^{\rho-1} A[(u-l)^{-\mu-1}f(u)](X-x_0) \cdot \left[\left(\frac{d}{dx} - l \right)^\mu \phi(x) \right]_{x=x_0}, \end{aligned}$$

where, for validity, $\phi(x)$ and its first ρ derivatives are continuous in (x_0, X) .

To allow x_0 to approach $-\infty$ as limit in (55), we shall want $\phi(x)$, and its first ρ derivatives, to be continuous for $x \leq X$. Also, for some c greater than d , and some positive M , we must have, for $x < x_1$,

$$|\phi^{(\mu)}(x)| \leq M e^{cx}; \quad \mu = 0, 1, 2, \dots, \rho. \dagger$$

As a result, we also have, for $x < x_1$,

$$\left| \left(\frac{d}{dx} - l \right)^\mu \phi(x) \right| \leq M' e^{cx}; \quad \mu = 0, 1, 2, \dots, \rho.$$

Now choose l less than c . The d of $(D-l)^{-\mu-1}f(D)$ will then also be less than c . Choose b between this d and c . Then, by (51),

$$|A[(u-l)^{-\mu-1}f(u)](X-x_0)| \leq L_\mu e^{bx}.$$

As $b < c$, these two inequalities give us

$$\lim_{x_0 \rightarrow -\infty} A[(u-l)^{-\mu-1}f(u)](X-x_0) \left[\left(\frac{d}{dx} - l \right)^\mu \phi(x) \right]_{x=x_0} = 0.$$

We thus obtain

$$(56) \quad \{f(D)\}_{-\infty}^X \phi(x) = \{(D-l)^{-\rho}f(D)\}_{-\infty}^X \left(\frac{d}{dx} - l \right)^\rho \phi(x).$$

Though (56) reduces all operators of finite order to operators of order zero, for which we have already established the law of successive operations, its conditions are stronger than those yielded by the following more extended

* See the first footnote of §8.

† If $\phi(x)$ behaves "regularly" as $x \rightarrow -\infty$, the inequalities for $\mu > 0$ will follow from the one for $\mu = 0$.

treatment. The operators $f_1(D)$ and $f_2(D)$ will now be of arbitrary finite orders ρ_1 and ρ_2 respectively with corresponding d_1 and d_2 . We shall require the inequalities

$$|\phi^{(\mu)}(x)| \leq M_1 e^{c_1 x}, \quad \mu = 0, 1, 2, \dots, \rho_2, \quad \text{for } x < x_1, \quad \text{with } c_1 > d_1.$$

When $d_2 \geq c_1$, we shall also require, to insure the existence of $\{f_2(D)\}_{-\infty}^x \phi(\xi)$,

$$|\phi(x)| \leq M_2 e^{c_2 x}, \quad \text{for } x < x_1, \quad \text{with } c_2 > d_2.$$

In the latter case $c_2 > c_1$, so that, for some x_2 , we shall have for $x < x_2$, $M_2 e^{c_2 x} < M_1 e^{c_1 x}$. In either case, we can prove that, if l is less than c_1 ,

$$\left| \frac{d^\mu}{dx^\mu} \{ (D - l)^{-\rho_2} f_2(D) \}_{-\infty}^x \phi(\xi) \right| \leq M_3 e^{c_1 x}, \quad \text{for } x < x_3, \quad \mu = 0, 1, 2, \dots, \rho_2.$$

In fact, through (54) extended, we can write

$$\begin{aligned} \frac{d^\mu}{dx^\mu} \{ (D - l)^{-\rho_2} f_2(D) \}_{-\infty}^x \phi(\xi) &= \{ D^\mu (D - l)^{-\rho_2} f_2(D) \}_{-\infty}^x \phi(\xi) \\ &+ \int_{-\infty}^{x-a} A [u^\mu (u - l)^{-\rho_2} f_2(u)] (x - \xi) \phi(\xi) d\xi. \end{aligned}$$

By reducing the first term of the right hand member of this equation by formula (17) with $\rho = \mu$, $X = x$, $x_1 = x_0 = x - a$, we can easily apply our inequalities, by methods already made familiar, to obtain the desired result. We can therefore use (56), with ρ_2 in place of ρ , $(D - l)^{-\rho_2} f_2(D)$ in place of $f_2(D)$, and $\{ (D - l)^{-\rho_2} f_2(D) \}_{-\infty}^x \phi(\xi)$ in place of $\phi(x)$ to obtain

$$\begin{aligned} \{ (D - l)^{-\rho_2} f_1(D) \}_{-\infty}^X \{ (D - l)^{-\rho_2} f_2(D) \}_{-\infty}^x \phi(\xi) \\ = \{ (D - l)^{-(\rho_1 + \rho_2)} f_1(D) \}_{-\infty}^X \left(\frac{d}{dx} - l \right)^{\rho_2} \{ (D - l)^{-\rho_2} f_2(D) \}_{-\infty}^x \phi(\xi). \end{aligned}$$

Now reduce the first member of this equation by our proved result for operators of order zero, the second member by operating with $(d/dx - l)^{\rho_2}$ formally, as is justified by (54). We thus obtain

$$\{ (D - l)^{-(\rho_1 + \rho_2)} f_1(D) f_2(D) \}_{-\infty}^X \phi(x) = \{ (D - l)^{-(\rho_1 + \rho_2)} f_1(D) \}_{-\infty}^X \{ f_2(D) \}_{-\infty}^x \phi(\xi).$$

Operating on both sides of this equation by $(d/dx - l)^{\rho_1 + \rho_2}$ yields our law of successive operations for operators of arbitrary finite orders.

In this discussion we have merely concerned ourselves with the convergence difficulties introduced by letting $x_0 = -\infty$. When we also supply the discussion of the existence of derivatives tacitly assumed above, a discussion that does not introduce anything essentially new, we obtain

THEOREM XVI. Let $f_1(D)$ and $f_2(D)$ be of finite orders ρ_1 and ρ_2 respectively, with corresponding d_1 and d_2 , and let $D^{\rho_1}f_2(D)$ be of order ρ . Let $\phi(x)$ and its first ρ_2 derivatives be continuous for $x \leq X$, and also let the first ρ derivatives of $\phi(x)$ be continuous in a left neighborhood of $x = X$. Finally, let us have the inequalities

$$|\phi^{(\mu)}(x)| \leq M_1 e^{c_1 x}, \text{ for } x < X, \mu = 0, 1, 2, \dots, \rho_2, \text{ with } c_1 > d_1,$$

and, if c_1 does not exceed d_2 , let us also have

$$|\phi(x)| \leq M_2 e^{c_2 x}, \text{ for } x < X, \text{ with } c_2 > d_2.$$

Then formula (53) is valid, and both of its members exist.

13. $x_0 = -\infty$: special case. When $d = 0$, the existence theorem of the preceding section still requires an exponential inequality for $|\phi(x)|$ as $x \rightarrow -\infty$. This can be replaced by an algebraic inequality if certain assumptions are made about the behavior of $f(z)$ in the neighborhood of those singularities that lie on the axis of imaginaries. The operator D^m is typical.

Let then $f(z)$ be of type zero, and analytic to the right of the axis of imaginaries. We shall first assume the origin to be the only singularity with real part zero. The assumptions required are the following:

(a) $f(z)$ is analytic within a circular sector of radius λ , and angle greater than π , which has its vertex at the origin, and is bisected by the positive half of the real axis,

(b) in this sector, for some real m and positive K , $|f(z)|$ satisfies the inequality

$$|f(z)| \leq K |z|^m.$$

Let S designate the interior of the infinite analytic sector of the type zero condition, S' the rest of the z -plane; s the interior of the circular sector of radius λ , s' the rest of the interior of the circle. It will be convenient to have λ small enough for this circle to lie wholly within S' . Since the origin is the only singularity with real value zero, there will be a line $R(z) = -\kappa_1$, $\kappa_1 > 0$, to the right of which $f(z)$ is analytic, except for z 's in s' . Now in the argument leading to inequality (50) of the preceding section, replace d by $-\kappa_1$, and choose the b'' of that argument between 0 and $-\kappa_1$, and also sufficiently near zero to have the line $R(z) = b''$ cut the sides of s' . If then we remove from C''' , as modified in that argument, the portion of this line that is in s' , we shall have along the resulting open contour C^{iv}

$$\left| \int_{C^{iv}} \frac{f(z) dz}{(1 - z\Delta x)^{r+1}} \right| < L_1 e^{-\kappa(r\Delta x)}$$

where $-\kappa$ is between 0 and b'' , and hence is negative.* For the purposes of this section we further modify C''' as follows. Let C_s^v consist of two segments in s , starting from a point $z = \delta$ on the real axis, running parallel to the sides of s , and terminated by $R(z) = b''$. Such a contour is possible for δ sufficiently small. Join this to C^v , where we remove from C^v the portions of $R(z) = b''$ that lie between the ends of C_s^v . Then C_s^v with the shortened C^v forms an admissible C''' . Now if C^v is thus shortened in the above inequality, the result is simply strengthened. As for C_s^v , replace δ by $1/[2(r+1)\Delta x]$, and transform z by $\zeta = (r+1)\Delta x z$. Using condition (b), we obtain

$$\left| \int_{C_s^v} \frac{f(z) dz}{(1 - z\Delta x)^{r+1}} \right| \leq \frac{K}{[(r+1)\Delta x]^{m+1}} \int_{C_{1/2}^v} \frac{|\zeta|^m |\, d\zeta |}{\left| 1 - \frac{\zeta}{r+1} \right|^{r+1}}.$$

If the sides of $C_{1/2}^v$ be extended to infinity, we obtain the same integral that occurred in §11, except that m replaces $\rho - \epsilon$. This integral was there shown to converge, and be bounded with respect to r for $r \geq \rho$. The same is thus true of the $C_{1/2}^v$ integral for $r > m$. Combining these results, we are enabled to state that, for Δx sufficiently small, and $r\Delta x$ greater than a fixed positive h , we have with a positive L ,

$$(57) \quad |A[f](r, \Delta x)| < \frac{L}{(r\Delta x)^{m+1}}.$$

Hence, also, for t greater than h , we have

$$(58) \quad |A[f](t)| \leq \frac{L}{t^{m+1}} \cdot \dagger$$

This result can be immediately extended to the case where there are a finite number of singularities of the above type on the axis of imaginaries. If these singularities are at points $z = z_i$, then on each we impose conditions like (a) and (b), where in (a) the vertex of the sector is now at z_i , while the inequality in (b) is replaced by

$$|f(z)| \leq K_i |z - z_i|^{m_i}.$$

The line $R(z) = -\kappa_1$ can be chosen once for all, and C''' modified near each of the singularities. As a result, (57) and (58) will follow, provided m is

* Inequality (b) of this section, in conjunction with the choice of b'' made here, allows us to conclude that $|f(z)|$ has a finite upper bound along that part of $R(z) = b''$ used in C^v , as is required in this proof.

† This inequality can be used as the basis of a simple derivation of certain of the Heaviside asymptotic expansions, and extensions thereof.

the least of the m_i 's. For ease of reference we shall speak of a function and corresponding operator of this kind as being of *degree m*. We can now state

THEOREM XVII. *If an operator $f(D)$ of type zero, with $d=0$, is of degree m , and if for $x < x_1, x_1 < 0$, we have*

$$|\phi(x)| \leq M(-x)^{m-\eta}$$

with positive η and M , the equivalence relation (49) will hold.

COROLLARY. *With this hypothesis replacing that of Theorem XV, Corollary I and Corollary II of the latter theorem continue to hold here.*

In the proof of this theorem the geometric progression of the preceding section is replaced by a series that can be written

$$LM \sum_{r=p+1}^{\infty} \left(1 - \frac{X}{r\Delta x}\right)^{m+1} \frac{1}{(r\Delta x - X)^{1+\eta}} \Delta x,$$

provided $x_0 < x_1$, and, also, $x_0 < X - h$. Since $[1 - X/(r\Delta x)]^{m+1}$ is bounded as $\Delta x \rightarrow +0$, and $x_0 \rightarrow -\infty$, while we have

$$\sum_{r=p+1}^{\infty} \frac{1}{(r\Delta x - X)^{1+\eta}} \Delta x < \int_{X-x_0-\Delta x}^{\infty} (t - X)^{-1-\eta} dt = \frac{(-x_0 - \Delta x)^{-\eta}}{\eta},$$

recourse can again be had to the limit criterion to give the theorem.

The case $\phi(x) \equiv 1$ is of special interest. When $d < 0$, the discussion of §11 applies, as well as the resulting formula (46). When $d > 0$, the same discussion shows $\{f(D)\}_{-\infty}^X 1$ to be non-existent. When $d=0$, (46) still holds, provided $f(z)$ is analytic at the origin; but the equivalence relation (49) seems to require further assumptions. The following theorem, which we state without proof, contains several simpler cases, and can be considered the extension of (46) for $d=0$.

THEOREM XVIII. *If $f(z)$ is of type zero, with $d=0$, and if, except for the origin, it is of degree of greater than -1 , while for the analytic sector vertexed at the origin it approaches a finite limit K as $z \rightarrow 0$ in the sector, then*

$$(59) \quad \{f(D)\}_{-\infty}^X 1 = K = \lim_{x_0 \rightarrow -\infty} \{f(D)\}_{x_0}^X 1.$$

We turn now to the law of successive operations. As before, $f_1(z)$ and $f_2(z)$ will be of finite order. The case of most interest is the one in which both d_1 and d_2 are zero, the treatment of the cases where but one d is zero then being obvious. On the whole, the development of the preceding section can be followed.

First let $f(D)$ be of order zero, while otherwise it, and $\phi(x)$, satisfy the hypothesis of Theorem XVII. We can again use (46), and break up the infinite integral over limits $(-\infty, X-a)$, $(X-a, X)$. Choose a greater than h , and let X be less than x_1 . We then easily obtain

$$\begin{aligned} \left| \int_{X-a}^X A[f](X-x)\phi(x)dx \right| &\leq M(-X+\theta a)^{m-\eta} \int_0^a |A[f](t)| dt < M''(-X)^{m-\eta}, \\ \left| \int_{-\infty}^{X-a} A[f](X-x)\phi(x)dx \right| &< LM \int_{-\infty}^{X-a} \frac{(-x)^{m-\eta}}{(X-x)^{m+1}} dx \\ &= LM(-X)^{-\eta} \int_0^{-X/a} (1+t)^{m-\eta} t^{\eta-1} dt, \end{aligned}$$

where, in the latter, we have set $t = -X/(X-x)$. It is convenient to assume $X < -a$, so that $-X/a > 1$. We can then break up this integral over limits $(0, 1)$, and $(1, -X/a)$. The first part converges, and is independent of X . For the second, we can write, with some fixed k ,

$$\int_1^{-X/a} (1+t)^{m-\eta} t^{\eta-1} dt < k \int_1^{-X/a} t^{m-1} dt.$$

The case $m=0$ can be avoided. For $m \neq 0$ we thus obtain the inequalities

$$(60) \quad m > 0: \quad | \{f(D)\}_{-\infty}^X \phi(x) | < M'(-X)^{m-\eta},$$

$$(61) \quad m < 0: \quad | \{f(D)\}_{-\infty}^X \phi(x) | < M'(-X)^{-\eta}.$$

Now let $f_1(D)$ and $f_2(D)$ both be of order zero, with d_1 and d_2 zero, and associated m_1 and m_2 . No loss of generality ensues if we assume the d of $f_1(D)f_2(D)$ to be zero.* The corresponding m_3 is evidently not less than m , the least of the three quantities m_1 , m_2 , m_1+m_2 . We can then show that if $\phi(x)$ satisfies the hypothesis of Theorem XVII for m so defined, the law of successive operations will be satisfied. It will be sufficient to show that $\{f_2(D)\}_{-\infty}^x \phi(\xi)$, $\{f_1(D)\}_{-\infty}^X \{f_2(D)\}_{-\infty}^x \phi(\xi)$ and $\{f_1(D)f_2(D)\}_{-\infty}^X \phi(x)$ exist in accordance with Theorem XVII, since the argument of the preceding section will then be valid here. The inequality of Theorem XVII is the primary consideration. Since the three operators involved, $f_2(D)$, $f_1(D)$, $f_1(D)f_2(D)$, are of degrees m_2 , m_1 , and m_3 respectively, we have to show that the corresponding operands $\phi(\xi)$, $\{f_2(D)\}_{-\infty}^x \phi(\xi)$ and $\phi(x)$ satisfy inequalities with exponents $m_2-\eta_2$, $m_1-\eta_1$ and $m_3-\eta_3$ respectively, where the η 's are all positive. Note that our hypothesis makes $\phi(x)$ satisfy an inequality with

* This d may be less than zero; but in that case the actual inequalities are even stronger than on the assumption that it is zero.

exponent $m - \eta$ where η is positive. This suffices for the first and last cases; for by identifying this inequality with the desired inequalities, we are setting $m_2 - \eta_2 = m - \eta$, $m_3 - \eta_3 = m - \eta$. Since m does not exceed m_2 or m_3 , we thus have $\eta_2 \geq \eta$, $\eta_3 \geq \eta$, that is, η_2 and η_3 are positive, as was desired. For η_1 , we must use the inequalities (60) and (61). The case $m_2 = 0$ can be avoided since m_2 can always be decreased in the inequalities on $|f_2(z)|$, at least for $|z - z_i| \leq 1$, and by making this decrease less than η there will at worst result a corresponding decrease in m and η in the original inequality on $\phi(x)$, with the new η still positive. For $m_2 > 0$, we see from (60) that $\{f_2(D)\}_{-\infty}^x \phi(\xi)$ satisfies an inequality with exponent equal to that of the inequality for $\phi(\xi)$. That is, we may set $m_1 - \eta_1 = m - \eta$, and as before, obtain $\eta_1 \geq \eta$. For $m_2 < 0$, (61) shows that the exponent for $\{f_2(D)\}_{-\infty}^x \phi(\xi)$ is but $-\eta_2$, if $m_2 - \eta_2$ is the exponent for $\phi(\xi)$. That is, we may set $m_1 - \eta_1 = -\eta_2$, where we have identically $m_2 - \eta_2 = m - \eta$. By combining the two equations, we obtain $\eta_1 = m_1 + m_2 - m + \eta \geq \eta$. A positive η_1 therefore results. The sufficiency of the $m - \eta$ exponent for $\phi(x)$ has thus been demonstrated.

It is unnecessary to go into the details of the extension of these results to operators of arbitrary finite orders, as the steps of the procedure of the previous section are easily verified here. We can thus state

THEOREM XIX. *Let $f_1(D)$ and $f_2(D)$ be of finite orders ρ_1 and ρ_2 respectively with d_1 and d_2 zero, and associated m_1 and m_2 , and let $D^{\rho_1} f_2(D)$ be of order ρ . Let $\phi(x)$ and its first ρ_2 derivatives be continuous for $x \leq X$, and also let the first ρ derivatives of $\phi(x)$ be continuous in a left neighborhood of $x = X$. Finally, for all x 's such that $x < x_1 < 0$, let*

$$|\phi(x)| \leq M(-X)^{m-\eta}, \quad |\phi^{(\mu)}(x)| \leq M_1(-X)^{m-\eta}, \quad \mu = 1, 2, \dots, \rho_2,$$

where η and η_1 are positive, and m is the least of the three quantities m_1 , m_2 , $m_1 + m_2$. Then formula (53) holds, and both of its members exist.

14. Entire operators; $A[f](t) \equiv 0$. If $f(z)$ is an entire function of the complex variable z , and satisfies condition (b) of §3 over the whole z -plane, then $f(D)$ will be said to be of *type zero over the plane*. Since entire transcendental functions of genus zero are known to satisfy this condition, their corresponding operators are of type zero over the plane. Polynomials in D are also included.

We first prove that, for such operators, $A[f](t)$ vanishes identically. In formula (8) of §3 choose C so that its vertex is at the origin. Let \bar{C}_N be the portion of the line $R(z) = -N$, $N > 0$, between the half-lines of C . Since $f(z)$ is entire, the integral in (8), limited to that part of C which is to the right of this line, will equal the same integral over \bar{C}_N . Hence we have

$$A[f](t) = \frac{1}{2\pi i} \lim_{N \rightarrow \infty} \int_{C_N} e^{tz} f(z) dz.$$

By choosing κ of condition (b) less than $t \cos \alpha$, this limit is seen to be zero.

An immediate consequence of this result is that, for continuous $\phi(x)$, $\{f(D)\}_{x_0}^X \phi(x)$ is independent of the finite lower limit x_0 , since by (9),

$$\{f(D)\}_{x_0}^X \phi(x) = \{f(D)\}_{x_1}^X \phi(x).$$

The same is evidently true for $x_0 = -\infty$, if $\phi(x)$ satisfies the hypothesis of Theorem XV. We can now easily show that when $f(D)$ is of type zero over the plane, $\{f(D)\}_{x_0}^X \phi(x)$ reduces to the formal result obtained by expanding $f(D)$ in powers of D . The function $\phi(x)$ is of course assumed to be analytic in a left neighborhood of $x = X$. We can then choose x_1 sufficiently near X to apply the Leibnitz expansion (41) to $\{f(D)\}_{x_1}^X \phi(x)$. We thus have

$$\{f(D)\}_{x_0}^X \phi(x) = \phi(X) \{f(D)\}_{x_1}^X 1 + \frac{\phi'(X)}{1!} \{f'(D)\}_{x_1}^X 1 + \dots$$

Now $f^{(n)}(D)$ is of type zero over the plane for every n . Hence

$$\{f^{(n)}(D)\}_{x_1}^X 1 = \{f^{(n)}(D)\}_{-\infty}^X 1 = f^{(n)}(0),$$

the last by (46). We therefore have the standard expansion

$$(62) \quad \{f(D)\}_{x_0}^X \phi(x) = f(0)\phi(X) + \frac{f'(0)}{1!}\phi'(X) + \frac{f''(0)}{2!}\phi''(X) + \dots$$

Suppose now, conversely, that $\{f(D)\}_{x_0}^X \phi(x)$ is independent of x_0 in a certain x_0 interval. Differentiation of (9) with respect to x_0 gives

$$A[f](X - x)\phi(x_0) = 0.$$

If then $\phi(x_0)$ does not vanish in this interval, $A[f](t)$ must vanish over a corresponding interval. Now formula (8) shows $A[f](t)$ to be an analytic function of t for every real and positive t . Hence if $A[f](t)$ vanishes over an interval, it vanishes identically. The question thus raised is answered by

THEOREM XX. *If $f(D)$ is of finite order, $A[f](t)$ vanishes identically when, and only when, $f(D)$ is a polynomial; if $f(D)$ is of type zero, $A[f](t)$ vanishes identically when, and only when, $f(D)$ is of type zero over the plane.*

The direct part of this theorem has already been demonstrated. As for the converses, let $f(D)$ first be of finite order ρ . Then $f^{(\rho)}(D)$ will be of order zero, so that, for $R(a) > d$, we shall have, by (48),

$$f^{(\rho)}(a) = \int_0^\infty e^{-at} A[f^{(\rho)}](t) dt.$$

Since under our hypothesis $A[f](t) \equiv 0$, (37) gives us $A[f^{(\rho)}](t) \equiv 0$, so that $f^{(\rho)}(a) = 0$ for all a 's in at least a half-plane. $f(a)$ is therefore a polynomial in that half-plane, and, being analytic, is a polynomial throughout.

Now let $f(D)$ be of type zero, with $A[f](t) \equiv 0$. Then, for $R(a) > d$,

$$\{f(D)\}_{-\infty}^X e^{ax} = \{f(D)\}_{x_1}^X e^{ax}.$$

We can apply (45) to the first member of this equation; and, as e^{ax} is an entire function of x , (41) will yield a convergent series for the second. We thus get

$$f(a) = \{f(D)\}_{x_1}^X 1 + \frac{\{f'(D)\}_{x_1}^X 1}{1!} a + \frac{\{f''(D)\}_{x_1}^X 1}{2!} a^2 + \dots$$

Since this power series in a converges in a half-plane of a , its radius of convergence must be infinite, and so $f(a)$ is entire. To prove that it is of type zero over the plane, note that, from the series, we have

$$\{f^{(n)}(D)\}_{x_1}^X 1 = f^{(n)}(0).$$

Now let X be any positive number, however small, and choose x_1 between 0 and X . Formula (41) will then yield the convergent series

$$\{f(D)\}_{x_1}^X x^{-1} = X^{-1}f(0) - X^{-2}f'(0) + X^{-3}f''(0) - \dots$$

Since this converges for every positive X we must have

$$\lim_{n \rightarrow \infty} [f^{(n)}(0)]^{1/n} = 0.$$

This condition on the coefficients of the expansion of an entire $f(z)$ in powers of z is equivalent to condition (b) of §3 holding over the z -plane. Hence $f(z)$ is of type zero over the plane.

The scope of Theorem XX is clarified by the following two observations. First, $f(z)$ may be entire, and of type zero in a sector of angle greater than π , and yet not of type zero over the plane; secondly, $f(z)$ may be a transcendental function of type zero over the plane without being of genus zero. For the first case consider the function

$$f(z) = \int_0^\infty e^{-zt} e^{-t^2} dt.$$

Since the integral converges uniformly over every bounded region of the

z -plane, $f(z)$ is an entire analytic function of z . Now it can easily be seen that, if t is considered as a complex variable, the positive half of the real t axis, used in the above integral for contour, can be replaced by any half-line from the origin which makes an angle θ between $-\pi/4$ and $\pi/4$ with that positive t axis. For any one θ , we find

$$|f(z)| < \frac{A_\theta}{R(e^{i\theta}z) + B_\theta},$$

for $R(e^{i\theta}z) > -B_\theta$. By combining the inequalities for $\theta = \theta_1$ and $\theta = -\theta_1$ we thus easily see that $f(z)$ is in fact of order zero over any sector of angle less than $3\pi/2$ bisected by the positive half of the real z axis. It is therefore also of type zero. That it is not of type zero over the plane is seen by considering negative real values of z , for which we have, with arbitrarily large N ,

$$f(z) > \int_0^N e^{-zt} e^{-t^2} dt > \frac{e^{-N^2}}{(-z)} [e^{N(-z)} - 1].$$

For the second observation consider

$$f(z) = \prod_{n=2}^{\infty} \left[1 - \frac{z^2}{(n \log n)^2} \right],$$

whose zeros are $\pm n \log n$, $n = 2, 3, \dots$. As $\sum 1/(n \log n)$ does not converge, $f(z)$ is not of genus zero. On the other hand

$$|f(z)| < \prod_{n=2}^{\infty} \left[1 + \frac{|z|^2}{(n \log n)^2} \right] < P_N(|z|) \frac{\sinh(\pi |z| / \log N)}{\pi |z| / \log N},$$

where $P_N(|z|)$ is the polynomial in $|z|$ formed from the first $N-2$ factors. By expressing the hyperbolic sine in terms of exponentials, we see that condition (b) of §3 is satisfied over the plane, that is, $f(z)$ is of type zero over the plane.*

15. The Laplace integral equation. Except for a few results that followed directly from our definition of generalized differentiation, the preceding sections studied that definition for operators which we called of type zero. In the present and following sections, we shall make an independent study of operators given as Laplace integrals. We have seen in §11 that $A[f](t)$ satisfies the Laplace integral equation (48) when $f(D)$ is of order zero. We turn now to an extension of the converse of this result.

It has been shown in the literature that if $\int_a^\infty \psi(t) e^{-xt} dt$, where $\psi(t)$ is

* For the standard derivation of the same inequalities in the case of functions of genus zero see Borel, *Fonctions Entières*, chapter 3.

continuous for $t \geq a$, converges for $\xi = \xi_0$, then it converges for all ξ 's with $R(\xi) > R(\xi_0)$, and represents an analytic function of ξ in that half-plane.* This proof is readily extended to the case where $\psi(t)$ is continuous only for $t > a$, by breaking up the interval of integration. With this in mind we state

THEOREM XXI. *Let $\psi(t)$ be a continuous function of t for $t > 0$, and let $\int_0^\infty \psi(t) e^{-\xi t} dt$, considered improper at both its limits, converge for some value of ξ . Then, if*

$$\int_0^\infty \psi(t) e^{-\xi t} dt = f(\xi)$$

in the resulting half-plane of convergence, we will have, for $t > 0$,

$$(63) \quad \psi(t) = A[f](t) = \lim_{\substack{\Delta x \rightarrow +0 \\ r \Delta x \rightarrow t}} \frac{(-1)^r f^{(r)}(1/\Delta x)}{r! \Delta x^{r+1}}.$$

Before we turn to the proof, note that when $f(D)$ is of order zero, $A[f](t)$ may be discontinuous for $t = 0$. Hence the assumption of continuity for $\psi(t)$ only for positive t . Such an assumption of mere continuity is in accord with the literature. Since we have observed that for $f(D)$ not only of order zero, but of type zero, $A[f](t)$ is analytic for positive t , we see that the present development must be a quite independent study of our fundamental definition.

The proof of the convergence theorem for Laplace integrals can easily be extended to show that successive derivatives of the integral can be found by differentiating under the integral sign. We therefore find, for the $f(\xi)$ of our theorem,

$$A[f](r, \Delta x) = \frac{(-1)^r f^{(r)}(1/\Delta x)}{r! \Delta x^{r+1}} = \int_0^\infty \frac{e^{-t/\Delta x} t^r}{r! \Delta x^{r+1}} \psi(t) dt,$$

provided $1/\Delta x$ is in the half-plane of convergence. Let $t = r\Delta x \cdot \tau$. With the help of Stirling's formula for $r!$ we obtain

$$A[f](r, \Delta x) = \frac{[r/(2\pi)]^{1/2}}{1+\epsilon} \int_0^\infty (e^{1-\tau})^r \psi(r\Delta x \cdot \tau) d\tau,$$

where $\epsilon \rightarrow 0$ as $r \rightarrow \infty$. The function $e^{1-\tau}$ attains a maximum value of one for $\tau = 1$. Hence $(e^{1-\tau})^r$ stays equal to one for $\tau = 1$ as $r \rightarrow \infty$, but otherwise becomes inappreciable, as r increases, except for an ever narrowing neighborhood of $\tau = 1$. This suggests that we break up the integral over limits $(0, 1-\lambda), (1-\lambda, 1+\lambda), (1+\lambda, \infty)$, where λ is between zero and one, and

* For references, see Pincherle, loc. cit., p. 40.

show that the first and third parts can be neglected in finding the limit of $A[f](r, \Delta x)$.

With this in mind, let b be a real number in the half-plane of convergence of the Laplace integral. That means, with our change of variable, that

$$\int_0^\infty e^{-b+r\Delta x \cdot r} \psi(r\Delta x \cdot r) dr$$

converges. Now rewrite the integrand $(e^{1-\tau})^r \psi(r\Delta x \cdot r)$ so that it reads $(e^{1-r+b\Delta x \cdot r})^r e^{-b+r\Delta x \cdot r} \psi(r\Delta x \cdot r)$. For fixed λ , and Δx sufficiently small, $(e^{1-r+b\Delta x \cdot r})^r$ will monotonically increase from $r=0$ to $r=1-\lambda$, and monotonically decrease from $r=1+\lambda$ to $r=\infty$. We can therefore apply the second law of the mean for integrals,* and obtain, with $0 \leq \theta \leq 1-\lambda$,

$$\begin{aligned} & \int_0^{1-\lambda} (e^{1-r+b\Delta x \cdot r})^r e^{-b+r\Delta x \cdot r} \psi(r\Delta x \cdot r) dr \\ &= [(e^{1-r+b\Delta x \cdot r})^r]_{r=+0} \int_0^\theta e^{-b+r\Delta x \cdot r} \psi(r\Delta x \cdot r) dr \\ & \quad + [(e^{1-r+b\Delta x \cdot r})^r]_{r=1-\lambda} \int_\theta^{1-\lambda} e^{-b+r\Delta x \cdot r} \psi(r\Delta x \cdot r) dr. \end{aligned}$$

The first term of the right hand member is always zero. As for the second term, the integral, expressed in the original variable t , is

$$\frac{1}{r\Delta x} \int_{r\Delta x \cdot \theta}^{r\Delta x(1-\lambda)} e^{-bt} \psi(t) dt,$$

and, due to the convergence of $\int_0^\infty e^{-bt} \psi(t) dt$, stays bounded irrespective of its limits of integration, as $r\Delta x$ approaches a finite limit. On the other hand, as $\Delta x \rightarrow +0$, $e^{1-r+b\Delta x \cdot r}$, for $r=1-\lambda$, becomes and remains less than a fixed positive quantity which is itself less than one. Since at the same time r must increase indefinitely, we thus obtain

$$\lim_{\substack{\Delta x \rightarrow +0 \\ r\Delta x \rightarrow t}} \frac{[r/(2\pi)]^{1/2}}{1 + \epsilon_r} [(e^{1-r+b\Delta x \cdot r})^r]_{r=1-\lambda} = 0.$$

Hence the $(0, 1-\lambda)$ contribution to $A[f](r, \Delta x)$ can be neglected in studying its limit. An entirely similar proof shows the same to be true of the $(1+\lambda, \infty)$ contribution. We thus have, provided either limit exists,

* In the present instance we need the second law of the mean for an improper integral that may be only conditionally convergent. If, however, we first apply this law to the corresponding integral with positive lower limit ϵ (see de la Vallée Poussin, *Cours d'Analyse*, vol. 2, 1925, pp. 1-3) and if we then let θ be a limit value of the corresponding θ_ϵ 's as $\epsilon \rightarrow +0$, we easily obtain the desired result.

$$\lim_{\substack{\Delta x \rightarrow +0 \\ r \Delta x \rightarrow t}} A[f](r, \Delta x) = \lim_{\substack{\Delta x \rightarrow +0 \\ r \Delta x \rightarrow t}} \frac{[r/(2\pi)]^{1/2}}{1 + \epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau} \tau)^r \psi(r \Delta x \cdot \tau) d\tau.$$

Since $(e^{1-\tau} \tau)^r$ is positive, and $\psi(r \Delta x \cdot \tau)$ is continuous in τ , we have

$$\begin{aligned} \frac{[r/(2\pi)]^{1/2}}{1 + \epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau} \tau)^r \psi(r \Delta x \cdot \tau) d\tau \\ = \psi[r \Delta x(1 + \theta\lambda)] \cdot \frac{[r/(2\pi)]^{1/2}}{1 + \epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau} \tau)^r d\tau, \end{aligned}$$

with $-1 < \theta < 1$. Now consider the special case $\psi(t) \equiv 1$. From the Laplace integral we find

$$f(\xi) = \int_0^\infty e^{-\xi t} dt = 1/\xi.$$

For this f direct calculation gives

$$A[f](r, \Delta x) = \frac{(-1)^r f^{(r)}(1/\Delta x)}{r! \Delta x^{r+1}} \equiv 1,$$

so that, from the above expression for $\lim A[f](r, \Delta x)$, we obtain

$$\lim_{\substack{\Delta x \rightarrow +0 \\ r \Delta x \rightarrow t}} \frac{[r/(2\pi)]^{1/2}}{1 + \epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau} \tau)^r d\tau \equiv 1.$$

Combining these results, we thus see that

$$\limsup_{\substack{\Delta x \rightarrow +0 \\ r \Delta x \rightarrow t}} A[f](r, \Delta x) \leq M_\lambda, \quad \liminf_{\substack{\Delta x \rightarrow +0 \\ r \Delta x \rightarrow t}} A[f](r, \Delta x) \geq m_\lambda,$$

where M_λ and m_λ are the upper and lower bounds of $\psi[t(1 + \theta\lambda)]$ as θ varies from -1 to 1 . Now let $\lambda \rightarrow +0$. The continuity of ψ makes M_λ and m_λ both approach $\psi(t)$ as limit. Hence the upper and lower limits of $A[f](r, \Delta x)$ both equal $\psi(t)$, i.e., $A[f](t)$ exists, and equals $\psi(t)$.

16. Operator expressed as a Laplace integral. With the help of the preceding section we shall now prove

THEOREM XXII. *If $f(\xi)$ is given by the convergent Laplace integral*

$$f(\xi) = \int_0^\infty \psi(t) e^{-\xi t} dt,$$

where $\psi(t)$ is continuous for positive t , and the integral is considered improper at both its limits, and if $\phi(x)$ is continuous in the closed interval (x_0, X) , then

$$(64) \quad \{f(D)\}_{x_0}^X \phi(x) = \int_{x_0}^X \psi(X-x) \phi(x) dx,$$

under either of the following two auxiliary conditions:

- (a) the improper integral $\int_0^h \psi(t) dt$ is absolutely convergent for positive h ;
- (b) $\phi(x)$ has a finite total variation in a left neighborhood of $x=X$.

It may be worth noting that (a) is the necessary and sufficient condition on a $\psi(t)$ continuous for positive t that $\int_{x_0}^X \psi(X-x) \phi(x) dx$ converge for every continuous $\phi(x)$; while (b) is the necessary and sufficient condition on a continuous $\phi(x)$ that this integral converge for every $\psi(t)$ which is continuous for positive t , and for which $\int_0^h \psi(t) dt$ converges.

Since the preceding section proves that $A[f](t)$ exists, and equals $\psi(t)$, it is sufficient for the proof of our theorem, as in the treatment of operators of order zero given in §5, to show that

$$\lim_{h \rightarrow 0} \limsup_{\Delta x \rightarrow 0} \left| \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x-h}^X \phi(x) \right| = 0.$$

We have by definition, and from the preceding section,

$$\begin{aligned} \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x-h}^X \phi(x) &= \sum_{r=0}^q A[f](r, \Delta x) \phi(X - r\Delta x) \Delta x, \quad q = \{h/\Delta x\}, \\ A[f](r, \Delta x) &= \int_0^\infty \frac{e^{-t/\Delta x} t^r}{r! \Delta x^{r+1}} \psi(t) dt. \end{aligned}$$

We shall now break up this integral over limits $(0, k)$, and (k, ∞) , where k is greater than h but independent of r and Δx , and consider the two sums into which the above sum is thus broken up.

The treatment of the sum arising from the integrals with limits (k, ∞) is independent of the special conditions (a), (b). Let

$$\alpha_r(t) = \frac{e^{-t/\Delta x} t^r}{r! \Delta x^{r+1}}.$$

We observe that $\alpha_r(t)$ is a monotonically decreasing function of t in the interval (k, ∞) , due to the inequalities $k > h > r\Delta x$. The same will then be true of $\alpha_r(t)e^{\theta t}$ for sufficiently small Δx . As in the last section, the second law of the mean for integrals is applicable, and gives

$$\left| \int_k^\infty \alpha_r(t) \psi(t) dt \right| = \alpha_r(k) e^{\theta k} \left| \int_k^\infty e^{-\theta t} \psi(t) dt \right| \leq N \alpha_r(k) e^{\theta k}.$$

The sum in question is thus not greater in absolute value than

$$MN \sum_{r=0}^q \alpha_r(k) e^{bk} \Delta x,$$

where M is the upper bound of $|\phi(x)|$ in (x_0, X) . The above inequalities $k > h > r\Delta x$ also show that $\alpha_r(k)$ increases with r , as r varies from 0 to q . Since $(q+1)\Delta x < h + \Delta x$, we find this sum to be less in absolute value than

$$(h + \Delta x) MN \alpha_q(k) e^{bk}.$$

As $\Delta x \rightarrow +0$, q increases indefinitely, and so $\alpha_q(k) \rightarrow 0$. This sum can therefore be neglected in our discussion.

For condition (a) it will be sufficient to observe that the sum for the $(0, k)$ integrals will not exceed in absolute value

$$M \int_0^k \left(\sum_{r=0}^q \frac{e^{-t/\Delta x} t^r}{r! \Delta x^r} \right) |\psi(t)| dt.$$

But we have directly

$$\sum_{r=0}^q \frac{e^{-t/\Delta x} t^r}{r! \Delta x^r} < e^{-t/\Delta x} \sum_{r=0}^{\infty} \frac{t^r}{r! \Delta x^r} = 1.$$

Hence this sum is less in absolute value than $M \int_0^k |\psi(t)| dt$, and hence also

$$\limsup_{\Delta x \rightarrow +0} \left| \left\{ f \left(\frac{\Delta}{\Delta x} \right) \right\}_{X-h}^X \phi(x) \right| \leq M \int_0^k |\psi(t)| dt.$$

Now let $h \rightarrow +0$, and at the same time let $k \rightarrow +0$, while keeping $k > h$. Due to the convergence of the last integral under condition (a), it will approach zero as limit with k , thus proving our result.

For condition (b), we shall write this sum in the form

$$\sum_{r=0}^q \phi(X - r\Delta x) \beta_r \Delta x, \quad \beta_r = \int_0^k \alpha_r(t) \psi(t) dt,$$

with $\alpha_r(t)$ as already defined, and use the identity

$$\begin{aligned} \sum_{r=0}^q \phi(X - r\Delta x) \beta_r \Delta x &= \phi(X) \sum_{r=0}^q \beta_r \Delta x + [\phi(X - \Delta x) - \phi(X)] \sum_{r=1}^q \beta_r \Delta x \\ &\quad + \cdots + [\phi(X - q\Delta x) - \phi(X - [q-1]\Delta x)] \beta_q \Delta x. \end{aligned}$$

We can write

$$\sum_{r=s}^q \beta_r \Delta x = \int_0^k \gamma_s(t) \psi(t) dt, \quad \gamma_s(t) = \sum_{r=s}^q \alpha_r(t) \Delta x.$$

Integration by parts gives the relation

$$\int_0^k \gamma_s(t)\psi(t)dt = \gamma_s(k) \int_0^k \psi(t)dt - \int_0^k \left[\int_0^t \psi(\tau)d\tau \right] \gamma'_s(t)dt.$$

Let N_k be the upper bound of $|\int_0^t \psi(\tau)d\tau|$ in the interval $0 \leq t \leq k$. Note that N_k is finite, since the convergence of the Laplace integral entails the convergence of this integral. We then have

$$\left| \int_0^k \gamma_s(t)\psi(t)dt \right| \leq N_k \left[\gamma_s(k) + \int_0^k |\gamma'_s(t)| dt \right].$$

Now we easily verify that

$$\gamma_s(k) < 1, \quad \gamma'_s(t) = \alpha_{s-1}(t) - \alpha_q(t).$$

Furthermore, we have, for $\alpha_{s-1}(t)$, and similarly for $\alpha_q(t)$,

$$\int_0^k \alpha_{s-1}(t)dt < \int_0^\infty \alpha_{s-1}(t)dt \equiv 1,$$

so that we find

$$\left| \sum_{r=s}^q \beta_r \Delta x \right| = \left| \int_0^k \gamma_s(t)\psi(t)dt \right| < 3N_k.$$

Returning to our identity we thus obtain

$$\left| \sum_{r=0}^q \beta_r \phi(X - r\Delta x) \Delta x \right| < 3N_k [|\phi(X)| + V_{X-h}^X \phi(x)]$$

where $V_{X-h}^X \phi(x)$ is the total variation of $\phi(x)$ in $(X-h, X)$. Under hypothesis (b), this is finite for sufficiently small h , and so does not then increase as $h \rightarrow 0$. On the other hand, N_k approaches zero as limit, as k , along with h , approaches zero as limit. Our theorem is thus proved under (b).

To illustrate (64), consider the operator $B(D, n)$. Through the change of variable $x = e^{-t}$, the usual integral for the beta function becomes

$$B(\zeta, n) = \int_0^\infty (1 - e^{-t})^{n-1} e^{-\zeta t} dt,$$

where $R(n) > 0$ (and $R(\zeta) > 0$). Since condition (a) is satisfied, we thus have, for $\phi(x)$ continuous in (x_0, X) ,

$$(65) \quad \{B(D, n)\}_{x_0}^X \phi(x) = \int_{x_0}^X [1 - e^{-(X-x)}]^{n-1} \phi(x) dx. *$$

* (65) in conjunction with (44) leads to the solution of certain linear differential equations with exponential polynomial coefficients by means of definite integrals. Note, of course, that $B(D, n)$ is of order zero for $R(n) > 0$.

17. Carson's form; $e^{-\alpha D}$ as operator. Carson has given a form for generalized differentiation which, with our notation, reads, if

$$\frac{f(\xi)}{\xi} = \int_0^\infty \chi(t) e^{-\xi t} dt,$$

then

$$(66) \quad \{f(D)\}_{x_0}^X \phi(x) = \frac{d}{dX} \int_{x_0}^X \chi(X-x) \phi(x) dx. *$$

We shall show that this form results from our definition in either of the following two cases:

(a) $\chi(t)$ is continuous for $t \geq 0$, with $\lim_{t \rightarrow \infty} e^{-\xi t} \chi(t) = 0$ for sufficiently large ξ ; $\chi'(t)$ exists for $t > 0$, and, with $\phi(x)$, satisfies the hypothesis of Theorem XXII.

(b) $\chi(t)$ is continuous for $t > 0$, with $\int_0^k |\chi(t)| dt$ convergent, and $\phi'(x)$ exists, and is continuous, for $x_0 \leq x \leq X$.

It may be noted that in case (a) Carson's form reduces to that of §16, while in case (b) it is related to that of §16 much as our operators of order one are related to those of order zero.

In case (a), let first $\chi(0) = 0$. We have directly

$$\frac{d}{dX} \int_{x_0}^X \chi(X-x) \phi(x) dx = \int_{x_0}^X \chi'(X-x) \phi(x) dx.$$

On the other hand, integration by parts of the integral for $f(\xi)/\xi$ gives

$$f(\xi) = \int_0^\infty \chi'(t) e^{-\xi t} dt,$$

so that, by Theorem XXII, we have also

$$\{f(D)\}_{x_0}^X \phi(x) = \int_{x_0}^X \chi'(X-x) \phi(x) dx.$$

(66) thus follows. If $\chi(0) = c$, we have

$$\int_0^\infty [\chi(t) - c] e^{-\xi t} dt = \frac{f(\xi) - c}{\xi}.$$

The case just proved can therefore be applied to $f(D) - c$, so that

* J. R. Carson, *The Heaviside operational calculus*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 43-68. Numerous papers on this subject by Carson are to be found in the Bell System Technical Journal.

$$\{f(D) - c\}_{x_0}^X \phi(x) = \frac{d}{dX} \int_{x_0}^X [\chi(X - x) - c] \phi(x) dx,$$

which easily reduces to (66)

In case (b), replace $X - x$ by a new variable. We thus get

$$\frac{d}{dX} \int_{x_0}^X \chi(X - x) \phi(x) dx = \int_{x_0}^X \chi(X - x) \phi'(x) dx + \chi(X - x_0) \phi(x_0)$$

where we returned to the old variable x in the end. Now by §§15 and 16

$$A[u^{-1}f(u)](t) = \chi(t), \quad \{D^{-1}f(D)\}_{x_0}^X \phi'(x) = \int_{x_0}^X \chi(X - x) \phi'(x) dx.$$

If we turn then to the first existence proof for operators of finite order, with $\rho = 1$, we see that it will hold here provided

$$\sum_{r=0}^p |A[u^{-1}f(u)](r, \Delta x)| \Delta x$$

is bounded as $\Delta x \rightarrow +0$. But the discussion of §16, with $u^{-1}f(u)$ in place of $f(u)$, shows that for $k > X - x_0$

$$\limsup_{\Delta x \rightarrow +0} \sum_{r=0}^p |A[u^{-1}f(u)](r, \Delta x)| \Delta x \leq \int_0^k |\chi(t)| dt,$$

which is here assumed finite. Hence formula (17), with $\rho = 1$, can be used to give

$$\{f(D)\}_{x_0}^X \phi(x) = \int_{x_0}^X \chi(X - x) \phi'(x) dx + \chi(X - x_0) \phi(x_0),$$

which is just what we found for the other member of (66).

More generally, if $\chi(t)$ and $\phi(x)$ satisfy the hypothesis of Theorem XXII, (66) is equivalent to the relation

$$\{f(D)\}_{x_0}^X \phi(x) = \frac{d}{dX} \{D^{-1}f(D)\}_{x_0}^X \phi(x),$$

which offers no difficulty when $f(D)$ is of finite order, or of type zero (see §7). However, the complete discussion of this relation when $f(\zeta)/\zeta$ is given by a Laplace integral offers considerable difficulty, and can only be made the subject of a separate investigation. We shall merely append an examination of the special operator e^{-ax} , $a > 0$, which, in addition to its relation to the Laplace integral treatment, throws considerable light on our previous work with operators of type zero.

This operator comes under Carson's treatment by assigning to $\chi(t)$ the value 0, for $0 \leq t \leq a$, 1, for $t > a$. Since $\chi(t)$ is not even continuous for $t > 0$, agreement of the results with our definition cannot be sought for in the above two cases, but must be obtained directly. From our definition,

$$\{e^{-a(\Delta/\Delta x)}\}_{x_0}^X \phi(x) = \sum_{r=0}^p \frac{e^{-a/\Delta x} a^r}{r! \Delta x^r} \phi(X - r\Delta x), \quad p = \left\{ \frac{X - x_0}{\Delta x} \right\}.$$

The coefficient of $\phi(X - r\Delta x)$ attains its largest value for $r\Delta x < a \leq (r+1)\Delta x$. As in the treatment of the Laplace integral equation, it can be shown that only values of $r\Delta x$ in a neighborhood of this value affect the limit as $\Delta x \rightarrow +0$. We thus similarly obtain, for continuous $\phi(x)$,

$$(67) \quad X - x_0 < a: \{e^{-aD}\}_{x_0}^X \phi(x) = 0; \quad X - x_0 > a: \{e^{-aD}\}_{x_0}^X \phi(x) = \phi(X - a), *$$

which agree exactly with Carson's results.[†] We may note in passing, that (31) gives the relation

$$A[e^{-auf(u)}](r, \Delta x) = \sum_{s=0}^r \frac{e^{-a/\Delta x} a^s}{s! \Delta x^s} A[f(u)](r-s, \Delta x),$$

from which, in a similar manner, we find

$$(68) \quad t < a: A[e^{-auf(u)}](t) = 0; \quad t > a: A[e^{-auf(u)}](t) = A[f(u)](t-a),$$

provided $f(D)$ is of type zero. If we replace $f(u)$ by $u^{-1}f(u)$, we also obtain, by (12),

$$(69) \quad X - x_0 < a: \{e^{-aD}f(D)\}_{x_0}^X 1 = 0; \quad X - x_0 > a: \{e^{-aD}f(D)\}_{x_0}^X 1 = \{f(D)\}_{x_0}^{X-a} 1.$$

Consider now some non-formal aspects of this and related operators. We found in §3 that, when $f(D)$ is of type zero, $A[f](t)$ exists for every positive t . By contrast, the formula

* For $X - x_0 = a$, terms on but one side of the maximum are included. Due to approximate symmetry of the coefficients with respect to this maximum, the result for $X - x_0 = a$ is seen to be $\phi(X - a)/2$.

[†] The reader may be interested in the following formal "derivation" of the formula of §16 from this formula. Writing symbolically

$$f(D) = \int_0^\infty \psi(t) e^{-tD} dt,$$

and noting that $\{e^{-tD}\}_{x_0}^X \phi(x)$ vanishes for $t > X - x_0$, we would be led to

$$\{f(D)\}_{x_0}^X \phi(x) = \int_0^\infty \psi(t) [\{e^{-tD}\}_{x_0}^X \phi(x)] dt = \int_0^{X-x_0} \psi(t) \phi(X-t) dt.$$

By replacing $X-t$ by x , we obtain the formula in question. A rigorous proof along these lines may be possible.

$$e^{-A}[e^{au}](r, \Delta x) = \frac{e^{-a/\Delta x} a^r}{r! \Delta x^r}$$

shows that $A[e^{-au}](t)$, while zero for every other positive t , fails to exist for $t=a$ through becoming infinite. Stronger still is the contrast furnished by e^{aD} , $a>0$, for which $A[e^{au}](t)$ fails to exist for every positive t not exceeding a certain positive α , and is zero for $t>\alpha$. Of course e^{-aD} and e^{aD} are not of type zero. They are however closely related to operators of type zero, since the corresponding analytic functions satisfy the analyticity condition (a), given in §3, for operators of type zero, and also the inequality of (b), not, however, for each positive κ , but only for $\kappa>\kappa_0$, $\kappa_0>0$. Calling operators satisfying (a), and this qualified (b), operators of type one, we easily find from §3 that, for them, $A[f](t)$ exists for t greater than some positive α . Also, by a modification of the treatment of part (b) of §6, we obtain the existence and expansion in series (21) of $\{f(D)\}_{x_0}^X \phi(x)$, provided $X-x_0$ is greater than this α and the radius of convergence of $\phi(x)$ at x_0 is greater than $X-x_0$ by more than a certain fixed positive β . These results were not included in the paper since the method used gave values to α and β larger than those demanded by the operators themselves. They serve, however, to clarify the non-existence of $A[e^{-au}](t)$ and $A[e^{au}](t)$ for certain t 's,* and the difference in form of $\{e^{-aD}\}_{x_0}^X \phi(x)$ for $X-x_0 < a$, and $X-x_0 > a$.

We may, in fact, think of our definition of generalized differentiation as not beginning to work, in the latter case, until $X-x_0 > a$. As a increases, this period of adjustment, as we may call it, increases. It is then interesting to observe that in the case of the operator e^{-D^2} this period of adjustment is never completed, since, for finite x_0 ,

$$(70) \quad \{e^{-D^2}\}_{x_0}^X \phi(x) \equiv 0.$$

Here, then, we must take $x_0 = -\infty$, for which, at least, (45) holds. e^{-D^2} can be considered of type higher than one. This failure of our definition of generalized differentiation, with its partial failure in the case of operators of type one, throws into greater relief its peculiar applicability to those operators we have called of type zero.

* That $A[e^{-au}](t)$ and $A[e^{au}](t)$ are identically zero for $t>\alpha$ corresponds to e^{-aD} and e^{aD} being operators of type one over the plane.

NEW YORK CITY, N. Y.

FINITE COMBINATORY PROCESSES—FORMULATION 1

EMIL L. POST

The present formulation should prove significant in the development of symbolic logic along the lines of Gödel's theorem on the incompleteness of symbolic logics¹ and Church's results concerning absolutely unsolvable problems.²

We have in mind a *general problem* consisting of a class of *specific problems*. A solution of the general problem will then be one which furnishes an answer to each specific problem.

In the following formulation of such a solution two concepts are involved: that of a *symbol space* in which the work leading from problem to answer is to be carried out,³ and a fixed unalterable *set of directions* which will both direct operations in the symbol space and determine the order in which those directions are to be applied.

In the present formulation the symbol space is to consist of a two way infinite sequence of spaces or boxes, i.e., ordinally similar to the series of integers . . . , -3, -2, -1, 0, 1, 2, 3, The problem solver or worker is to move and work in this symbol space, being capable of being in, and operating in but one box at a time. And apart from the presence of the worker, a box is to admit of but two possible conditions, i.e., being empty or unmarked, and having a single mark in it, say a vertical stroke.

One box is to be singled out and called the starting point. We now further assume that a specific problem is to be given in symbolic form by a finite number of boxes being marked with a stroke. Likewise the answer is to be given in symbolic form by such a configuration of marked boxes. To be specific, the answer is to be the configuration of marked boxes left at the conclusion of the solving process.

The worker is assumed to be capable of performing the following primitive acts.⁴

- (a) *Marking the box he is in (assumed empty),*
- (b) *Erasing the mark in the box he is in (assumed marked),*
- (c) *Moving to the box on his right,*
- (d) *Moving to the box on his left,*
- (e) *Determining whether the box he is in, is or is not marked.*

The set of directions which, be it noted, is the same for all specific problems and thus corresponds to the general problem, is to be of the following form. It is to be headed:

Start at the starting point and follow direction 1.

Received October 7, 1936. The reader should compare an article by A. M. Turing, *On computable numbers*, shortly forthcoming in the *Proceedings of the London Mathematical Society*. The present article, however, although bearing a later date, was written entirely independently of Turing's. *Editor.*

¹ Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, *Monatshefte für Mathematik und Physik*, vol. 38 (1931), pp. 173-198.

² Alonzo Church, *An unsolvable problem of elementary number theory*, *American Journal of Mathematics*, vol. 58 (1936), pp. 345-363.

³ Symbol space, and time.

⁴ As well as otherwise following the directions described below.

It is then to consist of a finite number of directions to be numbered 1, 2, 3, . . . n. The i th direction is then to have one of the following forms:

- (A) Perform operation O_i , [$O_i = (a), (b), (c), \text{ or } (d)$] and then follow direction j_i ,
- (B) Perform operation (e) and according as the answer is yes or no correspondingly follow direction j_i' or j_i'' ,
- (C) Stop.

Clearly but one direction need be of type C. Note also that the state of the symbol space directly affects the process only through directions of type B.

A set of directions will be said to be *applicable* to a given general problem if in its application to each specific problem it never orders operation (a) when the box the worker is in is marked, or (b) when it is unmarked.⁵ A set of directions applicable to a general problem sets up a deterministic process when applied to each specific problem. This process will terminate when and only when it comes to the direction of type (C). The set of directions will then be said to set up a *finite 1-process* in connection with the general problem if it is applicable to the problem and if the process it determines terminates for each specific problem. A finite 1-process associated with a general problem will be said to be a *1-solution* of the problem if the answer it thus yields for each specific problem is always correct.

We do not concern ourselves here with how the configuration of marked boxes corresponding to a specific problem, and that corresponding to its answer, symbolize the meaningful problem and answer. In fact the above assumes the specific problem to be given in symbolized form by an outside agency and, presumably, the symbolic answer likewise to be received. A more self-contained development ensues as follows. The general problem clearly consists of at most an enumerable infinity of specific problems. We need not consider the finite case. Imagine then a one-to-one correspondence set up between the class of positive integers and the class of specific problems. We can, rather arbitrarily, represent the positive integer n by marking the first n boxes to the right of the starting point. The general problem will then be said to be *1-given* if a finite 1-process is set up which, when applied to the class of positive integers as thus symbolized, yields in one-to-one fashion the class of specific problems constituting the general problem. It is convenient further to assume that when the general problem is thus 1-given each specific process at its termination leaves the worker at the starting point. If then a general problem is 1-given and 1-solved, with some obvious changes we can combine the two sets of directions to yield a finite 1-process which gives the answer to each specific problem when the latter is merely given by its number in symbolic form.

With some modification the above formulation is also applicable to symbolic logics. We do not now have a class of specific problems but a single initial finite marking of the symbol space to symbolize the primitive formal assertions of the logic. On the other hand, there will now be no direction of type (C). Consequently, assuming applicability, a deterministic process will be set up which is *unending*. We further assume that in the course of this process certain recognizable symbol groups, i.e., finite sequences of marked and unmarked boxes, will appear which are not further altered in the course of the process. These will be the derived assertions of the logic. Of course the set of directions corresponds to the deductive processes of the logic. The logic may then be said to be *1-generated*.

An alternative procedure, less in keeping, however, with the spirit of symbolic

⁵ While our formulation of the set of directions could easily have been so framed that applicability would immediately be assured it seems undesirable to do so for a variety of reasons.

logic, would be to set up a finite 1-process which would yield the n th theorem or formal assertion of the logic given n , again symbolized as above.

Our initial concept of a given specific problem involves a difficulty which should be mentioned. To wit, if an outside agency gives the initial finite marking of the symbol space there is no way for us to determine, for example, which is the first and which the last marked box. This difficulty is completely avoided when the general problem is 1-given. It has also been successfully avoided whenever a finite 1-process has been set up. In practice the meaningful specific problems would be so symbolized that the bounds of such a symbolization would be recognizable by characteristic groups of marked and unmarked boxes.

The root of our difficulty however, probably lies in our assumption of an infinite symbol space. In the present formulation the boxes are, conceptually at least, physical entities, e.g., contiguous squares. Our outside agency could no more give us an infinite number of these boxes than he could mark an infinity of them assumed given. If then he presents us with the specific problem in a finite strip of such a symbol space the difficulty vanishes. Of course this would require an extension of the primitive operations to allow for the necessary extension of the given finite symbol space as the process proceeds. A final version of a formulation of the present type would therefore also set up directions for generating the symbol space.⁶

The writer expects the present formulation to turn out to be logically equivalent to recursiveness in the sense of the Gödel-Church development.⁷ Its purpose, however, is not only to present a system of a certain logical potency but also, in its restricted field, of psychological fidelity. In the latter sense wider and wider formulations are contemplated. On the other hand, our aim will be to show that all such are logically reducible to formulation 1. We offer this conclusion at the present moment as a *working hypothesis*. And to our mind such is Church's identification of effective calculability with recursiveness.⁸ Out of this hypothesis, and because of its apparent contradiction to all mathematical development starting with Cantor's proof of the non-enumerability of the points of a line, independently flows a Gödel-Church development. The success of the above program would, for us, change this hypothesis not so much to a definition or to an axiom but to a *natural law*. Only so, it seems to the writer, can Gödel's theorem concerning the incompleteness of symbolic logics of a certain general type and Church's results on the recursive unsolvability of certain problems be transformed into conclusions concerning all symbolic logics and all methods of solvability.

COLLEGE OF THE CITY OF NEW YORK

⁶ The development of formulation 1 tends in its initial stages to be rather tricky. As this is not in keeping with the spirit of such a formulation the definitive form of this formulation may relinquish some of its present simplicity to achieve greater flexibility. Having more than one way of marking a box is one possibility. The desired naturalness of development may perhaps better be achieved by allowing a finite number, perhaps two, of physical objects to serve as pointers, which the worker can identify and move from box to box.

⁷ The comparison can perhaps most easily be made by defining a 1-function and proving the definition equivalent to that of recursive function. (See Church, loc. cit., p. 350.) A 1-function $f(n)$ in the field of positive integers would be one for which a finite 1-process can be set up which for each positive integer n as problem would yield $f(n)$ as answer, n and $f(n)$ symbolized as above.

⁸ Cf. Church, loc. cit., pp. 346, 356-358. Actually the work already done by Church and others carries this identification considerably beyond the working hypothesis stage. But to mask this identification under a definition hides the fact that a fundamental discovery in the limitations of the mathematicizing power of Homo Sapiens has been made and blinds us to the need of its continual verification.

Reprinted from the
 TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY
 Vol. 48, No. 2, pp. 208-350
 September, 1940

POLYADIC GROUPS

BY
 EMIL L. POST

TABLE OF CONTENTS

SECTION		PAGE
Introduction		209
I. GENERAL THEORY OF POLYADIC GROUPS		
1. Definition of a polyadic group		213
2. Identity, inverse, equivalence		214
3. The coset theorem		218
4. Subgroups and transforms; expansion in cosets		221
5. Reducibility		228
6. Arbitrary containing ordinary groups		238
7. Determination of all types of semi-abelianisms		242
8. On the construction of polyadic groups		245
II. FINITE POLYADIC GROUPS		
A. m -ADIC SUBSTITUTIONS AND SUBSTITUTION GROUPS		
9. The symmetric m -adic substitution group of degree n		248
10. 2^{m-1} -fold classification of m -adic substitutions; the m -adic alternating groups		250
11. Associated and containing ordinary groups; commutative m -adic substitutions		253
12. Further study of the complete m -adic δ -group and m -adic alternating groups		255
13. Transitive m -adic substitution groups		261
14. Intransitive m -adic substitution groups		262
15. Substitutions which are commutative with each of the substitutions of a transitive m -adic substitution group		263
16. Holomorphs of a regular m -adic substitution group		267
17. m -adic groups of μ -adic substitutions		272
18. Primitive and imprimitive (m, μ) substitution groups		273
19. Multiple transitivity; cyclically transitive m -adic substitution groups		276
20. Class of an m -adic substitution group		278
B. FINITE ABSTRACT POLYADIC GROUPS		
21. Cyclic polyadic groups; ordinary theory		282
22. Cyclic polyadic groups; polyadic theory		286
23. Abstract polyadic groups of the first three orders		293
24. Properties of transforms		295
25. Generation of polyadic groups by two groups, one invariant under the elements of the other		298
26. m -adic groups of order g prime to $m-1$		304
27. Sylow subgroups of order p^α with g/p^α prime to $m-1$		307
28. Representation of an arbitrary m -adic group as a regular m -adic substitution group		312
29. Invariant subgroups and quotient groups; the m -adic central quotient group		313

Presented to the Society, October 26, 1935; received by the editors January 4, 1940.

30. Commutator, semi-commutator, and quasi-commutator subgroups	316
31. The ϕ -subgroup of an m -adic group	322
32. Simply isomorphic m -adic groups; group of inner isomorphisms	324
33. Extension of Frobenius's theorem to m -adic groups	327
34. Representation of an abstract m -adic group as a transitive (m, μ) substitution group .	328

C. FINITE m -ADIC LINEAR GROUPS

35. m -adic linear transformations	330
36. m -adic collineations and collineation-groups	334
37. m -adic Hermitian invariants	337
38. Reduction to canonical form	340
39. m -adic invariants	344
40. Generalization of m -adic substitution and transformation groups	347

INTRODUCTION

The group concept is peculiar in the breadth of its application and the narrowness of its formulation. By modifying one or more of its restrictions there have resulted such concepts as that of semi-group, groupoid, mischgruppe, quasi-group, hypergroup, multigroup. In all of these generalizations of the group concept the group operation remains dyadic, that is, it is a function of two independent variables. Our present interest is in that generalization of the group concept which results when, while retaining all other of its special features, the group operation becomes polyadic, that is, a function of any finite number of independent variables.

As far back as 1904, E. Kasner thus considered generalizing the ordinary "group property," and called a set of elements closed under a k -adic operation a k -adic system⁽¹⁾. But the complete formulation of this generalization seems to have been first effected by Dörnte⁽²⁾ in 1928 in a paper containing an extensive theory of what he there terms n -groups, n being the number of independent variables in the operation. In 1932 Lehmer⁽³⁾ independently formulated and investigated the special concept he termed triplex, which, in Dörnte's terminology, is an abelian 3-group. Dörnte's m -group, to change

(¹) While the paper in question, *An extension of the group concept*, has not appeared in print, an abstract thereof will be found in the Bulletin of the American Mathematical Society, vol. 10 (1904), pp. 290-291. Though at one point of the abstract Kasner observes that "the law of combination of the general system is best exhibited by means of its k dimensional multiplication table," his original definition adds the requirement that the combination of no fewer than k elements shall be contained in the system—a requirement that is meaningless unless the k -adic operation itself is merely an extended product based on a prior dyadic operation. And the absence of any mention of an associative law, coupled with a reference to the inverse of an element, further suggests that, as in Miller's perfect cosets referred to below, this dyadic operation is understood to be that of some actual group in the ordinary sense containing the given system.

(²) W. Dörnte, *Untersuchungen über einen verallgemeinerten Gruppenbegriff*, Mathematische Zeitschrift, vol. 29 (1928), pp. 1-19.

(³) D. H. Lehmer, *A ternary analogue of abelian groups*, American Journal of Mathematics, vol. 54 (1932), pp. 329-338.

the symbol, is also our m -adic group, or, for unspecified m , our polyadic group⁽⁴⁾.

As examples of triadic systems, and these also are examples of triadic groups, Kasner mentions "the odd permutations in any number of letters, the ∞^2 central symmetries of the plane or the ∞^3 of space, the totality of dual or reciprocal transformations, the correlations contained in any projective group, the totality of conformal transformations of the plane which reverse angles." In the introduction of his paper Dörnte mentions, among other examples, residue classes modulo k as $(k+1)$ -groups, and in the body of his paper introduces many such arithmetical illustrations as exemplifiers of his abstract development. Apart from examples which are the subject of a major part of our theory, we may add the linear transformations of determinant an $(m-1)$ -st root of unity as an m -group, and, more significantly, the m -group consisting of all the substitutions of a group which, instead of carrying a fixed letter into itself, transform say $a_1 \rightarrow a_2, a_2 \rightarrow a_3, \dots, a_{m-1} \rightarrow a_1$. In all of these examples the polyadic operation is merely an extended product expressed in terms of a prior dyadic operation. On the other hand, lengths under the operation fourth proportional, now to be written $b:a=c:x$, constitute a 3-group in which, geometrically, the triadic operation is primary⁽⁵⁾. Even more so for an abstract m -group whose operation is given ab initio by an m -dimensional table.

While the abstract formulation of polyadic group must be credited to Dörnte, in its coset theorem the present paper may be said to solve the problem of determining the essential nature of a polyadic group. This basic result is to the effect that any m -adic group can have its class of elements so widened, and in that widened class a dyadic operation so introduced, that the enlarged class, under that operation as product, constitutes an ordinary group in which the class of elements of the m -adic group is a coset of an invariant subgroup of the ordinary group, and the operation of the m -adic group the product of m elements as elements of the ordinary group⁽⁶⁾. At first glance this theorem seems to be identical, for finite groups, with a result of Miller's

⁽⁴⁾ The present paper arose as a reaction to the importance ascribed to the group concept by C. J. Keyser in his *Mathematical Philosophy*, New York, 1922, Lecture XII. But see the next to the last paragraph of this introduction. We may note that an early attempt on our part to thus generalize the group concept on the basis of its fourfold characterization had failed. But on now turning to the twofold basis as given by Miller (*Finite Groups*, below, p. 52) we found generalization to be immediate.

⁽⁵⁾ Analytically, the operation becomes $x = (ac)/b$, and so a variant of $a - b + c$, easily seen to lead to a 3-group. This last is already present in Dörnte's paper, and generalized in his Theorem 7, §1. Note that geometrically even, the binary operation multiplication can nevertheless be defined, even if secondary. The about-to-be-mentioned coset theorem shows the same situation to obtain in general.

⁽⁶⁾ Cf. A. Suschkewitsch, *Über die Erweiterung der Semigruppe bis zur ganzen Gruppe*, Communications de la Société Mathématique de Kharkoff, (4), vol. 12 (1935), pp. 81-87.

of 1935⁽⁷⁾). But, apart from other differences in hypothesis, Miller obtains the coset conclusion by essentially *assuming* the given set of elements to be in an ordinary group⁽⁸⁾. However, as a result of the two theorems, finite polyadic group does become identical with Miller's "perfect co-set," some of whose properties he develops, provided the latter is understood to mean set of elements and polyadic operation thereon⁽⁹⁾.

In addition to differences in abstract development, the present paper goes beyond Dörnte's in generalizing the concepts of substitution and linear transformation in such a way that the resulting m -adic substitutions and m -adic linear transformations naturally lead to m -adic groups thereof (see §9 and §35 for their definition). These m -adic groups we study as generalizations of ordinary substitution and linear transformation groups. As incentive for this development, we have the theorem that any abstract m -adic group (finite) can be represented as a "regular" m -adic substitution group, a theorem which, indeed, first gave us our coset theorem. In the final section of the paper these concepts receive a wide extension which remains significant for ordinary groups. But they are then seen to be at least closely related to a type of ordinary group formulated by Specht⁽¹⁰⁾.

Intermediate between these generalizations of substitution group is one which includes m -adic groups of ordinary substitutions. Two of our examples given above are of this type. In this connection we may mention a work of Corral⁽¹¹⁾ referred to by Miller. With substitutions on a given finite set of letters in question, Corral calls a set of substitutions a perfect brigade if closed under the operation ABC , an imperfect brigade if closed under the operation $AB^{-1}C$. The former is then identical with a 3-group of ordinary substitutions, the latter with a schar of substitutions, schar in Baer's⁽¹²⁾ wider form of a concept due to Prüfer⁽¹³⁾. Prüfer's development had a great influence on

⁽⁷⁾ G. A. Miller, *Sets of group elements involving only products of more than n*, Proceedings of the National Academy of Sciences, vol. 21 (1935), pp. 45-47. All references to Miller other than to *Finite Groups* (below) concern this paper.

⁽⁸⁾ The closing statement in Kasner's abstract, which suggests an anticipation of our coset theorem for triadic groups, is more probably merely related thereto in similar fashion.

⁽⁹⁾ His condition that his set S contain no like subset is in error. Recognizing S as an $(n+1)$ -group of order h , we see from our §21 that his partial condition " h is a power of n " should be "every distinct prime factor of h is a factor of n ."

⁽¹⁰⁾ W. Specht, *Eine Verallgemeinerung der Permutationsgruppen*, Mathematische Zeitschrift, vol. 37 (1933), pp. 321-341.

⁽¹¹⁾ J. I. Corral, *Brigadas de Substitutiones*, Part I, Havana, 1934; Part II, Toledo, 1935.

⁽¹²⁾ R. Baer, *Zur Einführung des Scharbegriffs*, Journal für die reine und angewandte Mathematik, vol. 160 (1929), pp. 199-207. His abstract formulation occurs in the important footnote on page 202. (Condition III therein can be proved in its entirety, and so is unnecessary.) The same footnote proves, in our terminology (see §5), that every schar is reducible to an ordinary group. Had the same situation obtained for polyadic groups, there would have been no need of our coset theorem.

⁽¹³⁾ H. Prüfer, *Theorie der Abelschen Gruppen*, I, *Grundeigenschaften*, Mathematische Zeitschrift, vol. 20 (1924), pp. 165-187.

Dörnte who showed that by rewriting the operation $AB^{-1}C$ formally as ABC , Prüfer's schar becomes a special kind of 3-group. This reinterpretation is however no longer possible if the Prüfer hypothesis $AB^{-1}C = CB^{-1}A$ is deleted to give Baer's schar.

While Dörnte's development in large measure consists in extending Prüfer's schar results to n -groups, our own work correspondingly attempts to generalize ordinary group theory. Thus, at the very beginning of our developments, where Dörnte's recognizes no identity for an m -group with $m > 2$, we find that role played by certain sequences of $m - 1$ elements of the m -group, and are thus led to a development culminating in the coset theorem of §3. The remainder of Part I, which is really a theory of abstract polyadic groups finite or infinite, consists of largely unrelated topics, but each fundamental in the theory. Our program crystallizes in Part II which, in A , B , C , systematically generalizes most of the general topics of three chapters in the Miller, Blichfeldt, Dickson, *Finite Groups*⁽¹⁴⁾, that is, Miller's Chapters II and III on substitution groups and abstract groups respectively, and Chapter IX, Blichfeldt's introductory chapter on linear groups. The reader will find here certain developments which merely paraphrase the ordinary theory, others which are far richer in their polyadic form, and still others which have no counterpart in ordinary theory. On the whole, the amount that does go over is surprisingly large. The principal failure is the but partial extension of Sylow's theorem. To the student of ordinary groups we may point out, among other connections, that the generalizations quasi-abelianism and quasi-commutator subgroup of §30 remain significant for ordinary groups, that §5 also gives a polyadic superstructure to any ordinary group, and that the coset theorem could be used to translate polyadic group results independently arrived at into ordinary group properties. While much of Dörnte's paper becomes clarified by means of our coset theorem, and several of his developments are carried considerably further in our own work, the present paper by no means can be said to supplant Dörnte's. We are furthermore directly indebted to him for his concepts of semi-invariant subgroup and semi-abelian group.

Useful as the coset theorem is in establishing certain properties of polyadic groups, its very existence greatly minimizes the significance of that generalization. Nevertheless, we cannot agree with Miller who says "the generalization secured by using perfect cosets instead of groups is, however, only apparent." In its autonomous formulation, polyadic group is fundamentally a generalization of ordinary group and, indeed, it is as generalization that

(14) New York, 1916. Henceforth referred to as *Finite Groups*. Where in Part II the writer refers to the standard proof of an ordinary group result it is the proof in this text that is meant. We may note here that when an ordinary group term is applied without explicit definition to polyadic groups, its polyadic definition is entirely similar.

it lends itself to a corresponding development⁽¹⁵⁾. However, the final verdict will undoubtedly hang on the question of application⁽¹⁶⁾. For this end our concept of m -adic invariant is no doubt far too special (see §39). Genuine application of polyadic groups will probably therefore have to wait upon the formulation of an adequate concept of polyadic invariant.

We wish here to express our obligation to B. P. Gill to whose efforts we owe the completion of a major phase of our development (see §12). Had we completed the determination of the triadic linear groups in two variables mentioned in our preliminary report, this obligation would have been still greater. We are also indebted to R. Baer who, on two separate occasions, set us on the right path in the maze of ordinary group literature.

I. GENERAL THEORY OF POLYADIC GROUPS

1. Definition of a polyadic group. Given a class of elements C , and an operation $c(s_1s_2 \dots s_m)$, we shall say that the elements of C constitute an m -adic group G under c if the following two conditions are satisfied:

1. If any m of the $m+1$ symbols in an equation of the form

$$c(s_1s_2 \dots s_m) = s_{m+1}$$

represent elements in C , the remaining symbol also represents an element in C , and is uniquely determined by this equation.

2. The elements of C satisfy the associative law under c , that is, they satisfy

$$\begin{aligned} c(c(s_1s_2 \dots s_m)s_{m+1}s_{m+2} \dots s_{2m-1}) &= c(s_1c(s_2 \dots s_ms_{m+1})s_{m+2} \dots s_{2m-1}) \\ &= \dots = c(s_1s_2 \dots c(s_ms_{m+1}s_{m+2} \dots s_{2m-1})). \end{aligned} \quad (17)$$

(15) It is fundamental to remember, in this connection, that we are dealing not with a mere class of elements, but with a class of elements and an operation thereon; still better, with the properties of a class of elements under a given operation. Thus the genuineness of non-Euclidean geometry is not affected because it can be represented by certain constructions in Euclidean geometry. Had Miller's point of view been adopted, such a development as that of §5, for example, would hardly have been possible.

(16) E.g., such as the Galois theory in the case of ordinary groups, not applications, such as the examples given above, which are mere illustrations of polyadic groups or of the theory thereof. Much of Corral's development concerns a brigade Galois theory. But this seems to the writer to be merely a restatement of standard Galois theory in terms of brigades rather than a genuine application.

(17) This formulation, patterned by the author after Miller, is identical with Dörnte's except that Dörnte splits up our 1 into two parts, P_1 and P_s , according as S_{m+1} , or S_i , $i \neq m+1$, is to be determined. It is then readily proved by the methods of our next section that in P_s only the existence of the solution S_i need be postulated, its uniqueness being then provable. It can further be shown that this existence of a solution for S_i need only be universally postulated either for a single i with $1 < i < m$, or for both $i=1$ and $i=m$, the existence of a solution for S_i for all other i 's from 1 to m then being provable. If the second form be used in place of P_s , and the first can only be used for $m > 2$, the resulting set of postulates would be the exact generalization of the basis for ordinary groups used by Albert in his *Modern Higher Algebra*.

We shall also use Dörnte's phrase " m -group" for G . Though these conditions are vacuously satisfied when C is a null class, the ordinary group concept tacitly assumes the existence of at least one element, and so we make the same assumption here. An ordinary group is then immediately an m -adic group with $m=2$, that is, a dyadic group, or 2-group. Unlike Dörnte, we exclude the case $m=1$.

It is readily proved by induction that the number of elements entering into any combination of elements built up by the operation c is of the form $k(m-1)+1$, where, in fact, k is the number of c 's in the assumed symbolic expression of this "extended operation." As the basic operation $c(s_1s_2 \dots s_m)$ is on an ordered m -ad of elements, an extended operation built up by c 's orders the $k(m-1)+1$ elements appearing therein in a linear array $s_1, s_2, \dots, s_{k(m-1)+1}$. It is then readily proved that as a consequence of the associative law 2 the element given by such an extended operation depends only on the sequence $s_1, s_2, \dots, s_{k(m-1)+1}$, and is independent of the particular way in which parentheses are introduced in conjunction with the k c 's that must enter into such an expression. We are justified, then, in briefly writing any such extended operation $c(s_1s_2 \dots s_{k(m-1)+1})$.

2. Identity, inverse, equivalence. Let $a_1, a_2, \dots, a_{m-1}, a_m$ be elements of an m -adic group G satisfying the equation

$$c(a_1a_2 \dots a_{m-1}a_m) = a_m.$$

Assuming as we do that $m \geq 2$, we can, in fact, let a_m and $m-2$ of the $m-1$ elements a_1, a_2, \dots, a_{m-1} be arbitrary elements of G , and then determine the remaining element in accordance with 1 of §1 so that this equation will be satisfied. If now s be any element of G , we can likewise find s_2, s_3, \dots, s_m in G so that $c(a_ms_2s_3 \dots s_m) = s$. By our assumed equation we will have

$$c(c(a_1a_2 \dots a_{m-1}a_m)s_2s_3 \dots s_m) = c(a_ms_2s_3 \dots s_m).$$

Hence, by the associative law,

$$c(a_1a_2 \dots a_{m-1}c(a_ms_2s_3 \dots s_m)) = c(a_ms_2s_3 \dots s_m),$$

and so

$$c(a_1a_2 \dots a_{m-1}s) = s.$$

That is, if the equation $c(a_1a_2 \dots a_{m-1}s) = s$ holds for one s in G , it holds for every s in G . The sequence, or $(m-1)$ -ad, $\{a_1, a_2, \dots, a_{m-1}\}$ may then be called a left identity of G . In the same way we can show that if $c(sb_1b_2 \dots b_{m-1}) = s$ holds for one s in G , it holds for every s in G , and $\{b_1, b_2, \dots, b_{m-1}\}$ may be called a right identity of G .

We now prove that every left identity of G is a right identity, and conversely, thus arriving at the unique concept of an $(m-1)$ -ad as an identity of an m -adic group. Let $\{a_1, a_2, \dots, a_{m-1}\}$ be a left identity. Then $c(a_1a_2 \dots a_{m-1}a_1) = a_1$. By the associative law,

$$c(a_0a_1a_2 \cdots a_{m-2}c(a_{m-1}a_1a_2 \cdots a_{m-1})) = c(a_0c(a_1a_2 \cdots a_{m-2}a_{m-1}a_1)a_2 \cdots a_{m-1}).$$

Hence

$$c(a_0a_1a_2 \cdots a_{m-2}c(a_{m-1}a_1a_2 \cdots a_{m-1})) = c(a_0a_1a_2 \cdots a_{m-1}).$$

Since the first $m-1$ arguments of the two members of this equation are identical, the last must also be equal by 1, §1. Hence

$$c(a_{m-1}a_1a_2 \cdots a_{m-1}) = a_{m-1},$$

and $\{a_1, a_2, \dots, a_{m-1}\}$ is also a right identity. Similarly for the converse.

Our equation $c(a_1a_2 \cdots a_{m-1}a_1) = a_1$ shows that if $\{a_1, a_2, \dots, a_{m-1}\}$ is an identity, so is $\{a_2, \dots, a_{m-1}, a_1\}$. Hence also $\{a_3, \dots, a_{m-1}, a_1, a_2\}$, and so on. Of course we have used the preceding result on left identities being the same as right identities. In general, then, if $\{a_1, \dots, a_i, a_{i+1}, \dots, a_{m-1}\}$ is an identity, so is $\{a_{i+1}, \dots, a_{m-1}, a_1, \dots, a_i\}$. Otherwise stated, cyclic permutation of the elements of an identity leaves it an identity.

Our initial observation proved the existence of an identity for $m \geq 2$. Clearly, if $\{a_1, a_2, \dots, a_{m-1}\}$ is an identity, it is immaterial which $m-2$ of these elements were assumed arbitrarily. Hence all identities of an m -adic group can be obtained by arbitrarily assigning values to, say, a_1, a_2, \dots, a_{m-2} , and correspondingly determining a_{m-1} . If G be of finite order g , there are g^{m-1} $(m-1)$ -ads formed from elements of G . Hence G has g^{m-2} identities. There will be no ambiguity if we use similar terminology when g is infinite.

While the term identity will thus mean an $(m-1)$ -ad of the above kind, a corresponding development in connection with an extended operation on $k(m-1)+1$ arguments leads to what may be termed an extended identity in the form of a $k(m-1)$ -ad. Except for their number, extended identities enjoy the same properties as identities. Rather unsymmetrically we may say that $\{a_1, a_2, \dots, a_{k(m-1)}\}$ is an extended identity if $\{a_1, a_2, \dots, a_{m-2}, c(a_{m-1} \cdots a_{k(m-1)})\}$ is an identity.

The concept of identity immediately leads to that of inverse. For $m=2$, the inverse of an element s is an element which multiplied into s yields the identity. For $m > 2$, to obtain an identity from an element s we must annex $m-2$ other elements. We are thus led to an $(m-2)$ -ad as an inverse of s . Hence, for $m > 2$, an inverse of an element is an element when and only when $m=3$. $\{s_1, s_2, \dots, s_{m-2}\}$ is then an inverse of s if $\{s, s_1, s_2, \dots, s_{m-2}\}$ is an identity. As $\{s_1, s_2, \dots, s_{m-2}, s\}$ is then also an identity, we may therefore say that s is an inverse of the $(m-2)$ -ad $\{s_1, s_2, \dots, s_{m-2}\}$. We are thus led to define inverse for i -ads with arbitrary i .

First let $i < m-1$. We then define an inverse of an i -ad $\{s_1, s_2, \dots, s_i\}$ to be an $(m-i-1)$ -ad $\{s'_1, s'_2, \dots, s'_{m-i-1}\}$ such that $\{s_1, s_2, \dots, s_i, s'_1, s'_2, \dots, s'_{m-i-1}\}$ is an identity. As $\{s'_1, s'_2, \dots, s'_{m-i-1}, s_1, s_2, \dots, s_i\}$ is then also an identity, $\{s_1, s_2, \dots, s_i\}$ is an inverse of $\{s'_1, s'_2, \dots, s'_{m-i-1}\}$, so that we can talk of a pair of inverse polyads. When $i=m-1$ we must

have recourse to an extended identity, and are thus led to an $(m-1)$ -ad as inverse. $\{s'_1, s'_2, \dots, s'_{m-1}\}$ is then an inverse of $\{s_1, s_2, \dots, s_{m-1}\}$ if $\{s_1, s_2, \dots, s_{m-1}, s'_1, s'_2, \dots, s'_{m-1}\}$ is an extended identity. As before, $\{s_1, s_2, \dots, s_{m-1}\}$ is also an inverse of $\{s'_1, s'_2, \dots, s'_{m-1}\}$.

By means of inverses we easily solve an equation of the form

$$c(a_1 a_2 \cdots a_i s b_1 b_2 \cdots b_{m-i-1}) = s_0$$

for s ⁽¹⁸⁾. Let $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$, $\{b'_1, b'_2, \dots, b'_i\}$ be inverses of $\{a_1, a_2, \dots, a_i\}$, $\{b_1, b_2, \dots, b_{m-i-1}\}$ respectively. Operating on both sides of the above equation by $c(a'_1 a'_2 \cdots a'_{m-i-1} | b'_1 b'_2 \cdots b'_i)$, the bar indicating the missing argument, applying the associative law, and reducing the left-hand side by the property of identities we obtain

$$s = c(a'_1 a'_2 \cdots a'_{m-i-1} s_0 b'_1 b'_2 \cdots b'_i).$$

When a 's or b 's are missing, our inverse of an $(m-1)$ -ad serves the same purpose. Clearly an equation of the same type arising from an extended operation can always be reduced to the above type by means of the associative law. Our need of inverses of i -ads with $i > m-1$ is thus not pressing. However, they can be similarly introduced by means of extended identities. While such an inverse can always be a j -ad with $1 \leq j \leq m-1$, to preserve the symmetry of the inverse relationship we must allow $j > m-1$ as well, and thus have to introduce extended inverses. Thus if $i = k(m-1) + l$, $0 \leq l < m-1$, an inverse will be an $(m-l-1)$ -ad, while all extended inverses will have j in the form $\kappa(m-1) + (m-l-1)$.

The multiplicity of inverses when the latter are not single elements leads to the concept of equivalent i -ads. We can introduce that concept directly, however, as follows. Let $\{a_1, a_2, \dots, a_i\}$ and $\{b_1, b_2, \dots, b_i\}$ be such that for some specific $d_1, \dots, d_j, e_1, \dots, e_{m-i-j}$

$$c(d_1 \cdots d_j a_1 a_2 \cdots a_i e_1 \cdots e_{m-i-j}) = c(d_1 \cdots d_j b_1 b_2 \cdots b_i e_1 \cdots e_{m-i-j});$$

that is, replacing the sequence a_1, a_2, \dots, a_i by b_1, b_2, \dots, b_i in the specific operation given by the left-hand member of this equation leaves the result unaltered. Let $\{d'_1, \dots, d'_{m-i-1}\}$ and $\{e'_1, \dots, e'_{i+j-1}\}$ be inverses of $\{d_1, \dots, d_j\}$, $\{e_1, \dots, e_{m-i-j}\}$ respectively, and let $s_1, \dots, s_k, s_{k+1}, \dots, s_{m-i}$

(18) Dörnte solves this equation for $m > 2$ by his "querelement" \bar{a} , defined as the solution of the equation $c(a \cdots ax) = a$ for x . The very economy of this concept, however, helps obscure the concepts of our present section, so necessary for the basic coset theorem. It may be pointed out that actually our method of solution can be so presented as to be independent of the previous theorems on identities, and thus leads to that part of the footnote to §1 concerning the provability of the uniqueness of the solution. Indeed, in this primordial form, the same method is constantly used by Dörnte without specific formulation. The reader may be interested in noting that Dörnte's Theorems 3 and 4, §1, may be considered special cases of our identity results in that the definition of \bar{a} may now be restated: $\{a, \dots, a, \bar{a}\}$ is a right identity.

be arbitrary elements of G . Operating on both sides of the above equation by the extended operation $c(s_1 \cdots s_k d'_1 \cdots d'_{m-i-1} | e'_1 \cdots e'_{i+i-1} s_{k+1} \cdots s_{m-i})$ we obtain, after simplification,

$$c(s_1 \cdots s_k a_1 a_2 \cdots a_i s_{k+1} \cdots s_{m-i}) = c(s_1 \cdots s_k b_1 b_2 \cdots b_i s_{k+1} \cdots s_{m-i}).$$

A similar argument can be given when j , or $m-i-j$, is 0 or $m-1$. Hence, if the sequence $b_1 b_2 \cdots b_i$ can replace $a_1 a_2 \cdots a_i$ somewhere in one operation it can do so anywhere in any operation⁽¹⁹⁾. Clearly the same result holds good for extended operations as well. The i -ads $\{a_1, a_2, \dots, a_i\}$ and $\{b_1, b_2, \dots, b_i\}$ will then be said to be *equivalent*. Thus we may define an m -group G to be abelian if the dyads $\{s_1, s_2\}$ and $\{s_2, s_1\}$ are equivalent for every pair of elements s_1, s_2 of G . For then the value of $c(s_1 s_2 \cdots s_m)$, s 's in G , is unaltered by any interchange of adjacent s 's, and hence by any permutation of all the s 's.

Let $\{a_1, a_2, \dots, a_i\}$ and $\{b_1, b_2, \dots, b_i\}$ be equivalent i -ads, and let $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$ be an inverse of $\{a_1, a_2, \dots, a_i\}$. We have then $c(a'_1 a'_2 \cdots a'_{m-i-1} a_1 a_2 \cdots a_i s) = s$. Hence also $c(a'_1 a'_2 \cdots a'_{m-i-1} b_1 b_2 \cdots b_i s) = s$ so that $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$ is an inverse of $\{b_1, b_2, \dots, b_i\}$ as well. A similar argument applies when $i=m-1$. That is, *every inverse of one of a pair of equivalent i -ads is also an inverse of the other*. Again, let $\{a_1, a_2, \dots, a_i\}$ and $\{b_1, b_2, \dots, b_i\}$ both be inverses of $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$. Since we then have $c(a'_1 a'_2 \cdots a'_{m-i-1} a_1 a_2 \cdots a_i s) = s = c(a'_1 a'_2 \cdots a'_{m-i-1} b_1 b_2 \cdots b_i s)$, it follows that $\{a_1, a_2, \dots, a_i\}$ and $\{b_1, b_2, \dots, b_i\}$ are equivalent. That is, *inverses of the same polyad are equivalent*. It follows from these results that if $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$ is an inverse of $\{a_1, a_2, \dots, a_i\}$, the class of inverses of $\{a_1, a_2, \dots, a_i\}$ is the class of $(m-i-1)$ -ads equivalent to $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$. Conversely, the class of i -ads equivalent to $\{a_1, a_2, \dots, a_i\}$ is the class of inverses of $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$. Finally, the first class is the class of inverses of each member of the second, and conversely. This for $i < m-1$. For $i=m-1$ both classes consist of $(m-1)$ -ads.

We shall speak of the class of all i -ads equivalent to a given i -ad as a *class of equivalent i -ads*. As in the case of identities, to obtain all i -ads equivalent to a given i -ad we may assign arbitrary values to $i-1$ of the elements, the i th being then determined. We may therefore say that a class of equivalent i -ads has g^{i-1} members. If, on the other hand, we keep $i-1$ elements fixed, and let the remaining element run through G , 1 of §1 shows that no two of the resulting i -ads can be equivalent, while each class of equivalent i -ads thus finds a representative. We may therefore say that for each i there are exactly g classes of equivalent i -ads. These classes are, of course, mutually exclusive. For $i=1$ they are nothing more than the unit classes consisting of single elements of G . For $i=m-1$ one class of equivalent i -ads is singled out, that is, the class of identities.

⁽¹⁹⁾ This result is proved in part by Dörnte as Theorem 2, §1, but the corresponding concept is not formulated. Clearly this relationship between i -ads is an "equivalence relationship."

3. The coset theorem. We are now in a position to embed our m -adic group G in an ordinary group. Let C^* be the class of all classes of equivalent i -ads for $i = 1, 2, \dots, m-1$. Each element of C^* is thus a class of equivalent i -ads, and C^* may then be said to have $(m-1)g$ elements, g for each i . It is convenient to drop the distinction between a unit class and its sole member, so that we may consider C , the class of elements of G , a subclass of C^* . We proceed now to define a dyadic operation on the elements of C^* . But first we must remove the above tacit restriction $i < m$ in our discussion of equivalence. Clearly, by using extended operations, our results go over for $i \geq m$. Furthermore, we can extend the concept of equivalence to allow an i -ad to be equivalent to a j -ad. With only the basic operation c involved, we must clearly have $j-i$ a multiple of $m-1$. Without further elaboration, $\{b_1, b_2, \dots, b_{i+k(m-1)}\}$ will be equivalent to $\{a_1, a_2, \dots, a_i\}$ if $\{b_1, b_2, \dots, b_{i-1}, c(b_i \dots b_{i+k(m-1)})\}$ and $\{a_1, a_2, \dots, a_i\}$ are equivalent in the original sense⁽²⁰⁾.

We first prove the following: if two of the three polyads $\{a_1, a_2, \dots, a_i\}$, $\{b_1, b_2, \dots, b_j\}$, $\{a_1, a_2, \dots, a_i, b_1, b_2, \dots, b_j\}$ are respectively equivalent to the corresponding two of the three polyads $\{a'_1, a'_2, \dots, a'_i\}$, $\{b'_1, b'_2, \dots, b'_j\}$, $\{a'_1, a'_2, \dots, a'_i, b'_1, b'_2, \dots, b'_j\}$, the remaining polyads are equivalent. We shall prove this result for $i+j \leq m$, a corresponding proof with the use of extended operations serving for $i+j > m$. Consider then the operations $c(a_1 a_2 \dots a_i b_1 b_2 \dots b_j d_1 \dots d_{m-i-j})$ and $c(a'_1 a'_2 \dots a'_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j})$. If the first and second polyads of the first set of three are respectively equivalent to the first and second of the second set of three, we will have $c(a_1 a_2 \dots a_i b_1 b_2 \dots b_j d_1 \dots d_{m-i-j}) = c(a_1 a_2 \dots a_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j}) = c(a'_1 a'_2 \dots a'_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j})$, and the third polyads are equivalent. If the hypothesis concerns the first and third polyads, then $c(a_1 a_2 \dots a_i b_1 b_2 \dots b_j d_1 \dots d_{m-i-j}) = c(a'_1 a'_2 \dots a'_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j}) = c(a_1 a_2 \dots a_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j})$, whence the corresponding conclusion. Similarly for the second and third polyads.

Let then the dyadic operation $c^*(r_1 r_2)$ be defined as follows. If r_1 and r_2 are members of C^* , and if $\{a_1, a_2, \dots, a_i\}$ is in the class r_1 of equivalent i -ads, $\{b_1, b_2, \dots, b_j\}$ in the class r_2 of equivalent j -ads, then $c^*(r_1 r_2)$ is to be the class of $(i+j)$ -ads equivalent to $\{a_1, a_2, \dots, a_i, b_1, b_2, \dots, b_j\}$ when $i+j \leq m-1$, the class of $(i+j-(m-1))$ -ads equivalent to $\{a_1, a_2, \dots, a_i,$

⁽²⁰⁾ And, of course, our basic theorem on equivalent i -ads extends to equivalent polyads. It may then be noted that if we include a null sequence in this framework, an independent proof of the identity of left and right identities results. In fact, the about-to-be-proved coset theorem depends only on the concept of equivalence; and the properties of identity and inverse could therefore be derived with the help of that theorem. Their direct formulation in terms of the operation of the m -group, however, will be found indispensable for correct thinking on such topics as those of §5.

$b_1, b_2, \dots, b_i\}$ when $i+j > m-1$. When $i+j \leq m-1$, our previous results not only show that $c^*(r_1 r_2)$ is independent of the particular i -ad and j -ad chosen from r_1 and r_2 respectively, but that if any two symbols in the equation $c^*(r_1 r_2) = r_3$ are assigned values in C^* , the third is uniquely determined in C^* . The same is true when $i+j > m-1$ by the transitive property of equivalence. Hence, condition 1 of §1 for a dyadic group is satisfied by (C^*, c^*) ; likewise condition 2, that is, the associative law. For let $\{a_1, \dots, a_i\}, \{b_1, \dots, b_j\}, \{c_1, \dots, c_k\}$ be in r_1, r_2, r_3 respectively. Then, with equivalence extended as above, if $i+j+k = l+\lambda(m-1)$, $1 \leq l \leq m-1$, both $c^*(c^*(r_1 r_2)r_3)$ and $c^*(r_1 c^*(r_2 r_3))$ represent the class of l -ads equivalent to $\{a_1, \dots, a_i, b_1, \dots, b_j, c_1, \dots, c_k\}$, so that, for all members of C^* ,

$$c^*(c^*(r_1 r_2)r_3) = c^*(r_1 c^*(r_2 r_3)).$$

Hence, the members of C^* constitute an ordinary group under c^* . With G as the given m -adic group, this ordinary group will be symbolized G^* .

We have observed that we may consider the members of G to be members of G^* , that is, those classes of equivalent i -ads for which $i=1$. We now further observe that the operation $c(s_1 s_2 \dots s_m)$ can be identified with the extended operation $c^*(s_1 s_2 \dots s_m)$ when, of course, the s 's are in G . For $c^*(s_1 s_2 \dots s_m)$ is, indeed, the class of monads equivalent to $\{s_1, s_2, \dots, s_m\}$, and so consists of but the one monad $c(s_1 s_2 \dots s_m)$ ⁽²¹⁾. We shall therefore call G^* the abstract containing ordinary group of G , abstract by contrast with other possibilities to be discussed later. In fact, G^* is clearly determined by the abstract form of G . And while G^* as derived is not abstract, it may be made so by replacing the members of C^* by symbols formally obeying the rule of combination c^* as determined above.

To obtain a clearer view of the relationship between G and G^* , and thus, indeed, really to solve the problem of the essential nature of a polyadic group, let us consider those members of G^* which are classes of equivalent $(m-1)$ -ads. We have already observed that one of these g classes is the class of identities of G . Now if in the equation

$$c^*(r_1 r_2) = r_3$$

any two of the three symbols represent classes of equivalent $(m-1)$ -ads, so does the third. It follows that the g classes of equivalent $(m-1)$ -ads constitute an ordinary group under c^* , and hence a subgroup of G^* . We shall symbolize this ordinary group by G_0 , and call it the associated ordinary group of G . It is readily seen that G_0 is an invariant subgroup of G^* ⁽²²⁾. To prove that, it

(21) If then G has but a finite number of elements, Miller's theorem concerning perfect cosets can be applied immediately to give the coset theorem that follows. However we here make no such restriction on G .

(22) Provided $m > 2$. For $m=2$, $G^*=G=G_0$. If then we here allow the term subgroup to include the group itself, the results of the present section are also valid for ordinary groups, though in trivial fashion.

is sufficient to show that in the equation

$$c^*(tr_1) = c^*(r_1r_2)$$

if t is in G_0 , r_1 in G^* , then r_2 is in G_0 . But if r_1 is a class of equivalent i -ads, t being a class of equivalent $(m-1)$ -ads, then $c^*(tr_1)$, and hence $c^*(r_1r_2)$, is also a class of equivalent i -ads. r_2 , then, can only be a class of equivalent $(m-1)$ -ads, as was to be proved.

Let us now expand G^* in cosets as regards its invariant subgroup G_0 . As in the invariance proof, if a multiplier r represents a class of equivalent i -ads, the corresponding coset consists of classes of equivalent i -ads, and indeed, constitutes the class of all g classes of equivalent i -ads. While this is immediate when g is finite, in any case if r_1 is a class of equivalent i -ads, the equation $c^*(r_2r) = r_1$ demands that r_2 be in G_0 , so that r_1 is in the coset in question. Hence the expansion of G^* as regards G_0 consists of exactly $m-1$ augmented cosets, each being the class of all g classes of equivalent i -ads, for some $i=1, 2, \dots, m-1$. The elements of G itself therefore constitute one of these cosets, that is, that one for which $i=1$. Hence our basic theorem. *Every polyadic group is a coset of an ordinary group with respect to an invariant subgroup*, it being understood that the polyadic operation of the polyadic group is an extension of the dyadic operation of the ordinary group.

With the relationship between G , G_0 and G^* made thus precise, it becomes desirable to simplify our notation. Hence, when but a single m -adic operation c is involved, we shall write the corresponding dyadic operation $c^*(r_1r_2)$ simply as the product r_1r_2 of standard group theory. Our identification of $c(s_1s_2 \cdots s_m)$ with $c^*(s_1s_2 \cdots s_m)$ therefore enables us to write $c(s_1s_2 \cdots s_m)$, simply, $s_1s_2 \cdots s_m$. We now finally introduce the completely abstract view of G^* with symbols for elements. Clearly the element of G^* corresponding to the class of identities of G is the identity of G^* , and so will be symbolized by 1, as usual. With the elements of G^* as symbols, it will be convenient to call the symbol r , representing a class of equivalent i -ads, an i -ad. Thus $s_1s_2 \cdots s_i$ will be an i -ad when the s 's are elements of G . Conversely, every i -ad can be written thus. In particular, G_0 , itself, consists of all distinct products $s_1s_2 \cdots s_{m-1}$ of $m-1$ elements in G . To avoid duplication, of course, we may keep $m-2$ of these elements fixed, and let the remaining one run through G .

In particular, if s is an element of G , s^i is an i -ad, and so may correspondingly be used as multiplier in the expansion of G^* in cosets as regards G_0 . We may therefore write this expansion

$$G^* = G_0s + G_0s^2 + \cdots + G_0s^{m-2} + G_0 = sG_0 + s^2G_0 + \cdots + s^{m-2}G_0 + G_0.$$

Most significantly we may then also write

$$G = G_0s = sG_0.$$

Since G_0 consists of products of elements of G , we see that G^* itself is generated by the elements of G . The expansion of G^* shows the quotient group G^*/G_0 to be of order $m-1$, and, indeed, cyclic, with the element corresponding to the given polyadic group G as generator. Our *coset theorem* is thus more precise than its brief formulation, given above, would indicate.

By means of this theorem we shall be able to prove many results concerning polyadic groups by means of known results on ordinary groups. On the other hand, the following almost immediately obvious converse enables polyadic group theory to make contributions to a certain aspect of ordinary group theory. To wit, if a coset of an ordinary group with respect to an invariant subgroup is of finite order $m-1$ as element of the corresponding quotient group, then the elements of the coset constitute a polyadic group under the product of m elements as operation⁽²³⁾. Though easily proved directly, this result may be considered a consequence of the general theorem of §8. It will also be generalized at the end of the next section. Note that such a result cannot be true for a coset corresponding to an element of infinite order of the quotient group.

4. Subgroups and transforms; expansion in cosets. Dörnte has treated the subject of expansions of polyadic groups in cosets exhaustively. While not possessing identities and inverses to lead to a concept of transforms, he was enabled adequately to treat invariant subgroups by mere commutativity properties. He further introduced what we shall refer to as semi-invariant subgroups, a concept which the writer completely overlooked in his own development, and was thus led to a more general concept of polyadic quotient groups than is given by invariant subgroups. Nevertheless we shall reexamine these concepts from the point of view of the coset theorem, and a theory of transforms, since not only do they become clearer thereby, but indeed admit of a certain degree of generalization.

A proper subclass of the class of elements of an m -adic group G will be said to constitute a subgroup H of G if the elements of that subclass constitute a polyadic group under the polyadic operation of G . This is clearly equivalent to the following. If in an equation $c(s_1 s_2 \cdots s_m) = s_{m+1}$ any m elements are in the subclass, the $(m+1)$ -st is. For the rest of the definition of m -adic group follows from the elements of the subclass being in G . Where no confusion can result we shall occasionally allow G to be a subgroup (improper) of itself. We proceed first to investigate the relationship between H^* and G^* , H_0 and G_0 .

With H^* and G^* considered as being composed of classes of equivalent i -ads, only those members of H^* which are in H will also be members of G^* . For if $\{s_1, s_2, \dots, s_i\}$ is an i -ad of H , and hence also of G , the class of H i -ads equivalent to $\{s_1, s_2, \dots, s_i\}$ is but a proper subclass of the class of G i -ads equivalent to $\{s_1, s_2, \dots, s_i\}$ whenever $i > 1$. Nevertheless a 1-1 correspondence is thus set up between the members of H^* and the members of G^* .

⁽²³⁾ Already proved by Miller in equivalent form for finite groups.

containing them. For the latter are mutually exclusive. Hence, when G^* is treated abstractly with symbols as elements, we may symbolize the members of H^* correspondingly; and as the operation $c^*(s_1s_2)$, that is, s_1s_2 as explained above, when set up for G^* now serves also for H^* , H^* thereby becomes a subgroup of G^* .

The $(m-1)$ -ads of H^* are then also $(m-1)$ -ads of G^* , so that H_0 is a subgroup of G_0 . If s is any element of H , we can simultaneously expand H^* and G^* in the form

$$H^* = H_0s + H_0s^2 + \cdots + H_0s^{m-2} + H_0,$$

$$G^* = G_0s + G_0s^2 + \cdots + G_0s^{m-2} + G_0.$$

It follows that the $m-1$ augmented cosets of H^* as regards H_0 are respectively contained in the $m-1$ augmented cosets of G^* as regards G_0 . As an immediate consequence, we have *Lagrange's theorem holds for finite polyadic groups*. For, defining the order of a polyadic group as the number of its elements, the relations $G=G_0s$, $H=H_0s$ show that the order g of the polyadic group G , and the order h of its subgroup H , are respectively the same as the order of the ordinary group G_0 , and its subgroup H_0 ; and hence, h is a divisor of g .

Since H generates H^* , and in turn consists of the common elements of H^* and G , the correspondence between the subgroups H of G , and their abstract containing groups H^* , is 1-1. H_0 consists of the common elements of H^* and G_0 , and hence is also determined by H . In fact, we shall find useful the result that the products of $m-1$ elements chosen from a subgroup H of G constitute a subgroup of G_0 , namely H_0 . On the other hand, different subgroups of G may have the same associated ordinary group H_0 . Hence, in general, we can only say that the correspondence between the subgroups H of G , and their associated ordinary groups H_0 , is but many-one. Furthermore, not every subgroup H_0 of G_0 need be the associated ordinary group of a subgroup H of G . The coset theorem and its converse, indeed, show that *the necessary and sufficient condition that a subgroup H_0 of G_0 be the associated ordinary group of some subgroup H of G is that there exist an element s of G such that H_0 is invariant under s , while s^{m-1} is in H_0* . Indeed the subgroups of G are the distinct H_0s 's obtained from all H_0 's and s 's satisfying this condition.

As has been observed by Dörnte, two subgroups H and K of an m -adic group G need have no element in common. Thus, this will always be so if H and K are distinct subgroups of G with the same associated group. If, however, H and K do have an element in common, their common elements clearly constitute a subgroup of each of the subgroups, if they are not identical with one or the other. Moreover, if s be such a common element, by writing $H=H_0s$, $K=K_0s$, we see that the associated group of the "crosscut" of H and K is the crosscut of their associated groups.

We consider next the expansion of G in cosets as regards a subgroup H thereof. H_0 is clearly a subgroup of G^* . We may therefore expand G^* in say right cosets as regards H_0 . Now it is immediately seen that such a coset of H_0 either has no element in G , or is completely contained in G . For if this coset has an element s in common with G , then, since the coset can be written H_0s , and H_0 is contained in G_0 , H_0s will be wholly contained in $G = G_0s$. As all the elements of G must appear in the given expansion of G^* , we see that the cosets in question containing elements of G constitute a separation of the elements of G into mutually exclusive classes of elements. We may say then that G has thus been *expanded in right cosets as regards H* . A similar result holds for *left cosets*.

And now an immediate generalization. In the above discussion H served only to introduce the subgroup H_0 of G_0 . If then H_0 be any subgroup of G_0 , whether it corresponds to a subgroup H of G , or not, the above argument holds without change. Hence, *every subgroup of the associated ordinary group of a polyadic group leads to an expansion of the polyadic group in right cosets, and in left cosets, as regards that subgroup.*

Specifically, if in the expansion of G^* in right cosets as regards H_0 the corresponding multipliers which are in G are $s_\alpha, s_\beta, \dots, s_\kappa$, then the expansion of G in right cosets as regards H_0 can be written

$$G = H_0s_\alpha + H_0s_\beta + \dots + H_0s_\kappa.$$

Similarly for left cosets. A not easily proved theorem for ordinary finite groups is that the coset multipliers may be so selected that they are the same on the right as on the left. An immediate corollary of the preceding formulation is that the same is true of finite polyadic groups.

It is sometimes necessary to consider the intersections of cosets in the expansion of G in, say, right cosets as regards subgroups H_0 , and K_0 , of G_0 . We have then immediately that while a coset with respect to H_0 and a coset with respect to K_0 may have no elements in common, if they do have a common element s , then their common elements constitute the set L_0s where L_0 is the crosscut of H_0 and K_0 . In particular, if G is finite, all such intersecting pairs of cosets intersect in the same number of elements, namely, a number equal to the order of the crosscut of H_0 and K_0 .

Expansions of G in double cosets likewise admit of simple treatment. With H_0 and K_0 arbitrary subgroups of G_0 , we may expand G^* in double cosets H_0rK_0 . If any element of such a double coset is in G , the entire double coset is contained in G . Hence, if in the expansion of G^* we select those double cosets with r in G , the result will be a separation of the elements of G^* into mutually exclusive sets, that is, the expansion of G in double cosets as regards H_0 and K_0 . In particular, if G has subgroups H and K whose associated ordinary groups are H_0 and K_0 respectively, the resulting expansion may be

spoken of as the expansion of G in double cosets as regards H and K , the case considered by Dörnte⁽²⁴⁾.

We shall introduce the property of invariance through the more general concept of transform. To insure the fundamental correctness of our concept, we go back to first principles. Given an element s , and an i -ad $\{s_1, s_2, \dots, s_i\}$, both considered in the m -adic sense, we define the *transform* of s under $\{s_1, s_2, \dots, s_i\}$ to be the element

$$c(s'_1 s'_2 \cdots s'_{m-i-1} s s_1 s_2 \cdots s_i)$$

where $\{s'_1, s'_2, \dots, s'_{m-i-1}\}$ is an inverse of $\{s_1, s_2, \dots, s_i\}$. This for $i < m - 1$; a similar definition holds for $i = m - 1$. Since all inverses of a given polyad are equivalent, this transform is uniquely determined by s , and $\{s_1, s_2, \dots, s_i\}$. Since inverses of equivalent i -ads are also equivalent, it follows that equivalent i -ads yield identical transforms of a given element.

In saying s and $\{s_1, s_2, \dots, s_i\}$ are m -adic, we tacitly assume that there is some m -adic group to which s, s_1, s_2, \dots, s_i belong. Let us then consider the abstract containing ordinary group of this m -adic group, and treat it in abstract form, with simplified notation. If, then, i -ad $\{s_1, s_2, \dots, s_i\}$ corresponds to abstract i -ad r of the containing group, the $(m-i-1)$ -ad $\{s'_1, s'_2, \dots, s'_{m-i-1}\}$ will correspond to an abstract $(m-i-1)$ -ad r' such that if s be an element of the m -adic group, $r'rs = s$. Writing the identity of the containing group as usual, we thus have $r'r = 1$, and hence in customary notation, $r' = r^{-1}$. Consequently, if r represents a class of equivalent polyads of a polyadic group, r^{-1} represents the class of inverses of those polyads. The transform of s under $\{s_1, s_2, \dots, s_i\}$ can now be written $r^{-1}s r$. And so, the transform of an element by an i -ad is the ordinary transform of that element by the corresponding abstract i -ad in the abstract containing group.

We can now extend our concept of transform to that of the transform of a polyad by a polyad. In general, via the abstract containing group, the transform of r_1 by r_2 is $r_2^{-1}r_1r_2$. Had we resorted to our primitive concepts in this case, we would have, as with inverses, a class of equivalent transforms. We readily see that in all cases the transform of an i -ad, $i \leq m - 1$, is an i -ad.

Consider now an m -adic group G , and an i -ad r not necessarily an i -ad of G . Then, as with ordinary groups, if each element of G is transformed by r , there results an m -adic group G' which may be said to be simply isomorphic with G , and will be termed the transform of G under r . In fact, let s' be the transform under r of any element s of G . Since $r^{-1}s_1 r \cdot r^{-1}s_2 r \cdots r^{-1}s_m r = r^{-1}s_1 s_2 \cdots s_m r$, we see that the relationship $s_1 s_2 \cdots s_m = s_{m+1}$ is equivalent

⁽²⁴⁾ At first glance it would appear that Dörnte's expansions in cosets and double cosets, while depending on actual subgroups of G , are more general than we have stated them to be. However, it is readily seen that Dörnte's expansions with respect to a subgroup, or subgroups of G are our expansions of G with respect to transforms, in the sense defined below, of the given subgroup or subgroups by polyads of G . And since these transforms are again subgroups of G , the Dörnte expansions are no more general than we have stated them to be.

to $s'_1 s'_2 \cdots s'_m = s'_{m+1}$. The defining properties 1 and 2 for an m -adic group then follow immediately for the transform of G from the selfsame properties for G —hence the m -adic group G' . In general, two m -adic groups G and G' may be said to be *simply isomorphic* if a 1-1 correspondence can be set up between their elements such that if s' of G' is the correspondent of s in G , then we will have, for all elements of G ,

$$[c(s_1 s_2 \cdots s_m)]' = c'(s'_1 s'_2 \cdots s'_{m+1}),$$

c and c' designating the m -adic operations of G and G' respectively. For G' the transform of G this is immediate with c and c' the common unexpressed m -adic operation.

We reserve a more detailed treatment of transforms for our study of finite polyadic groups, and turn to the question of invariance. An m -adic element, polyad, or group will be said to be invariant under an i -ad if it is transformed into itself by that i -ad. It will then be said to be invariant under an m -adic group if it is invariant under every polyad of that group. Since G^* is generated by G , it follows that for K to be invariant under G , it is sufficient that it be invariant under every element of G . If such a K is an element (subgroup) of G it will then be said to be an invariant element (subgroup) of G . Clearly, the condition that an m -group G be abelian is equivalent to each of its elements being an invariant element of G . For, in the notation of the coset theorem, $\{s_1, s_2\}$ and $\{s_2, s_1\}$ being equivalent becomes $s_1 s_2 = s_2 s_1$, or, $s_2^{-1} s_1 s_2 = s_1$; and conversely.

Given an invariant subgroup H of G , the expansion of G in cosets as regards H immediately leads to an m -adic quotient group G/H . In fact, since H is invariant under G , it immediately follows that H_0 , the associated 2-group of H , is also invariant under G ; that is, H_0 , as subgroup of G^* , is invariant under each element of G considered as element of G^* . For H_0 consists of all products of $m-1$ elements chosen arbitrarily and independently from H . Hence the transform of H_0 under any element s of G consists of all products of $m-1$ elements chosen arbitrarily and independently from the transform of H under s , that is, from H all over again.

Consider then the expansion in cosets $G = H_0 s_\alpha + H_0 s_\beta + \cdots + H_0 s_\kappa$. Then, exactly as in ordinary group theory, the coset in which the element $s_1 s_2 \cdots s_m$ appears depends only on the cosets containing the elements s_1, s_2, \dots, s_m . If then $\sigma_1, \sigma_2, \dots, \sigma_m$ represent the cosets containing s_1, s_2, \dots, s_m respectively, we may write the coset containing $s_1 s_2 \cdots s_m$ in the form $\sigma_1 \sigma_2 \cdots \sigma_m$. An m -adic operation is thus determined on these cosets as elements; and, again as in classic theory, these cosets constitute an m -adic group under this operation. We may therefore call this group the quotient group G/H .

As we shall see later, m -adic quotient groups arising from invariant subgroups are very special kinds of polyadic groups. However, Dörnte has em-

phasized that m -adic quotient groups can arise in more general fashion. In our presentation, his argument reduces to the fact that the only use made of the invariance of subgroup H under G was to prove the invariance of H_0 under G . We shall call a subgroup H of G whose associated 2-group H_0 is invariant under G a *semi-invariant* subgroup of G . It follows that every semi-invariant subgroup of an m -adic group leads to an m -adic quotient group.

This result can be made still more general. For we observed earlier that any subgroup H_0 of the associated 2-group G_0 of G gives rise to expansions in cosets. It therefore follows that *every subgroup of the associated 2-group of an m -adic group which is invariant under the m -adic group leads to an m -adic quotient group*. In the absence of a subgroup H of G we shall write this quotient group G/H_0 .

It is immediately seen that with H_0 thus invariant under G , the right cosets of G as regards H_0 are identical with the left cosets. For $s^{-1}H_0s = H_0$ yields $H_0s = sH_0$. Conversely, if the right cosets of G as regards H_0 are identical with the left cosets, then, for each element s of G , $H_0s = sH_0$, so that H_0 is invariant under G . We thus see that the Dörnte concept of semi-invariance may be said to be the necessary and sufficient condition that a subgroup of a polyadic group give rise to a quotient group. Our extension, however, frees G from the need of possessing a subgroup H corresponding to the H_0 invariant under G .

In recent literature the concept of homomorphism appears as essentially equivalent to that of quotient group⁽²⁵⁾. By means of our coset theorem we readily show the same to be true for m -groups⁽²⁶⁾. As the analysis is not too immediate, we have refrained from explicitly using this concept except in the last section where it is especially needed.

An m -group G with operation c may be said to be *homomorphic* to an m -group \bar{G} with operation \bar{c} if there is a many-one correspondence between the elements of G and of \bar{G} such that whenever s_1, s_2, \dots, s_m of G respectively correspond to $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_m$ of \bar{G} , $c(s_1s_2 \cdots s_m)$ corresponds to $\bar{c}(\bar{s}_1\bar{s}_2 \cdots \bar{s}_m)$. We first show that such a homomorphism between G and \bar{G} determines a homomorphism between their abstract containing groups G^* and \bar{G}^* . In fact, let i -ad r of G^* be said to correspond to i -ad \bar{r} of \bar{G}^* if there exist elements s_1, s_2, \dots, s_i of G , and corresponding elements $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_i$ of \bar{G} , such that $r = c^*(s_1s_2 \cdots s_i)$, $\bar{r} = \bar{c}^*(\bar{s}_1\bar{s}_2 \cdots \bar{s}_i)$. It is readily seen that this sets up a correspondence between all the elements of G^* and all the elements of \bar{G}^* . Furthermore, this correspondence is many-one. For suppose r of G^* corresponds to \bar{r}_1 and \bar{r}_2 of \bar{G}^* . Then we must have $r = c^*(s_1s_2 \cdots s_i)$, $\bar{r}_1 = \bar{c}^*(\bar{s}_1\bar{s}_2 \cdots \bar{s}_i)$, and, also, $r = c^*(s'_1s'_2 \cdots s'_i)$, $\bar{r}_2 = \bar{c}^*(\bar{s}'_1, \bar{s}'_2, \dots, \bar{s}'_i)$, with $s_1, s_2, \dots, s_i, s'_1, s'_2, \dots, s'_i$ of G corresponding to $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_i, \bar{s}'_1, \bar{s}'_2, \dots, \bar{s}'_i$ respectively of \bar{G} . If then s of G corresponds to \bar{s} of \bar{G} , the equation

⁽²⁵⁾ See, for example, B. L. van der Waerden, *Moderne Algebra*, Berlin, 1930, vol. 1, §9.

⁽²⁶⁾ Dörnte's Theorem 8, §6, does the same for his more limited concept of m -adic quotient group under the assumption that the homomorph has at least one "first order element."

$c(s_1s_2 \cdots s_is \cdots s) = c(s'_1s'_2 \cdots s'_is \cdots s)$, obtained from the two forms of r , yields $\bar{c}(\bar{s}_1\bar{s}_2 \cdots \bar{s}_i\bar{s} \cdots \bar{s}) = \bar{c}(\bar{s}'_1\bar{s}'_2 \cdots \bar{s}'_i\bar{s} \cdots \bar{s})$ as a result of the homomorphism between G and \bar{G} . Hence $\bar{r}_1 = \bar{r}_2$. Finally, if r_1 and r_2 of G^* thus correspond to \bar{r}_1 and \bar{r}_2 of \bar{G}^* , $c^*(r_1r_2)$ corresponds to $\bar{c}^*(\bar{r}_1\bar{r}_2)$ —immediately, if r_1 and r_2 are an i -ad and j -ad respectively with $i+j \leq m-1$, and via the homomorphism between G and \bar{G} if $i+j > m-1$. The many-one correspondence between the elements of G^* and of \bar{G}^* is therefore a homomorphism.

The ordinary theorem on homomorphisms is therefore applicable, and we can state that the elements of G^* corresponding to the identity of \bar{G}^* constitute an invariant subgroup H_0 of G^* , while the elements of G^* corresponding to any element of \bar{G}^* constitute a coset in the expansion of G^* as regards H_0 , the quotient group G^*/H_0 being then simply isomorphic with \bar{G}^* . Since the identity of \bar{G}^* is an $(m-1)$ -ad, H_0 must consist of $(m-1)$ -ads in G^* , and is thus a subgroup of G_0 invariant under G . Those cosets of G^* as regards H_0 which involve elements of G therefore constitute an expansion of G as regards H_0 . Finally, the correspondence between G^* and \bar{G}^* is but the original correspondence for elements of G and \bar{G} . We thus have the following theorem.
If m -group G is homomorphic to m -group \bar{G} , there is an m -adic quotient group G/H_0 such that the correspondents of each element of \bar{G} constitute a coset in G/H_0 , this quotient group then being simply isomorphic with \bar{G} . Actually, as we have seen, H_0 consists of the elements of G_0 corresponding to the identity of \bar{G}_0 in the homomorphism between G^* and \bar{G}^* , and hence between G_0 and \bar{G}_0 determined by the given homomorphism. Since an m -group G is clearly homomorphic to any m -adic quotient group G/H_0 , the equivalence of the concepts of homomorphism and quotient group has been shown to hold also for m -groups.

A homomorphism between m -groups G and \bar{G} is thus always an $(N, 1)$ isomorphism with fixed N , N of course finite for finite m -groups. A more immediate consequence of the given homomorphism is that it sets up a many-one correspondence between the subgroups of G and the subgroups of \bar{G} , an m -group being considered now as a subgroup of itself. In fact, given a subgroup of G , the corresponding elements of \bar{G} are readily seen to satisfy the conditions for an m -group, and thus constitute the uniquely corresponding subgroup of \bar{G} . On the other hand, given a subgroup of \bar{G} , the set of all corresponding elements of G constitutes a subgroup of G with the given subgroup of \bar{G} as corresponding subgroup, and indeed, contains all such subgroups of G . Clearly this many-one correspondence between the subgroups of G and of \bar{G} is preserved under the relation "subgroup of"—subgroup, in the above sense of group or subgroup.

It is also readily verified that if the set \bar{G} is not known to be an m -group under operation \bar{c} , yet the remainder of the definition of homomorphism between G and \bar{G} is satisfied, then \bar{G} is an m -group under \bar{c} , and hence the given relation a genuine homomorphism. In fact, the only part of our defi-

nition of m -group not immediately given for \bar{G} under \bar{c} , as a consequence of its being satisfied by G under c , is the uniqueness of the solution of $\bar{c}(\bar{s}_1\bar{s}_2 \cdots \bar{s}_m) = \bar{s}_{m+1}$ for \bar{s}_i , $1 \leq i \leq m$. Passing by the considerations of the footnote of §1 and a special argument valid only for \bar{G} finite, we can in every case solve corresponding equations $c(s_1s_2 \cdots s_m) = s_{m+1}$ for s_i as in §2, with all s 's except s_i and s_{m+1} fixed, and thus find that all such s 's must correspond to the same, consequently unique, \bar{s}_i ⁽²⁷⁾.

Our converse of the coset theorem admits of immediate extension to the case of an m -adic quotient group. For the statement of this result we need the concept of order, when finite, of an element of an m -group as given in the beginning of §21. We may note now, however, that an element s may be said to be of first order if $c(ss \cdots s) = s$, the unit class with sole member s then being a subgroup of the given m -group. We see then immediately that *if an element of an m -adic quotient group is of the first order, the corresponding coset constitutes a subgroup of the given m -group*. For the isomorphism between the given m -group and the quotient group shows that if in an equation $c(s_1s_2 \cdots s_m) = s_{m+1}$ any m elements are in the coset, the $(m+1)$ -st element must also be in that coset. Now consider any element σ of finite order k of the quotient group. Anticipating a concept of the next section, we may note now that our given m -group will constitute a polyadic group under the extended operation $c(s_1s_2 \cdots s_\mu)$ with $\mu = k(m-1)+1$. Our m -adic quotient group likewise extends to a μ -group with the element σ now being a first order element of the μ -adic quotient group. The previous result therefore leads to the following. *If an element of an m -adic quotient group is of finite order k , then the elements of the corresponding coset constitute a polyadic group under the operation of the given group extended to $k(m-1)+1$ elements.*

5. **Reducibility.** Given any ordinary group with class of elements C and dyadic operation s_1s_2 , an m -adic group on the same elements will be determined if we set up the m -adic operation $c(s_1s_2 \cdots s_m) = s_1s_2 \cdots s_m$. We shall

(27) If a general isomorphism between m -groups G and \bar{G} be defined as a many-many correspondence between their elements in which m -adic products of corresponding elements correspond, then, for finite m -adic groups, as for finite ordinary groups, the correspondence is that of a simple isomorphism between m -adic quotient groups of G and \bar{G} . On the other hand, Dickson (these Transactions, vol. 6 (1905), pp. 205–208) has shown by an example that the finite group theorem does not hold for infinite groups, while Loewy (Festschrift Heinrich Weber, 1912, pp. 198–227) calls an isomorphism “vollständig” if inverses of corresponding elements also correspond—the case when the finite group theorem does hold for infinite groups—and derives a number of interesting conditions for a general isomorphism to be “vollständig.” In the case of infinite m -adic groups, the condition under which the finite m -adic group theorem goes over can be written in a variety of ways, but perhaps most symmetrically as follows. If in two equations $c(s_1s_2 \cdots s_m) = s_{m+1}$, $\bar{c}(s'_1s'_2 \cdots s'_{m'}) = s'_{m'+1}$, m of the $m+1$ symbols in the first equation, and the m corresponding symbols in the second equation, represent elements of G and \bar{G} respectively that correspond, then the elements represented by the remaining symbols must correspond. The writer is indebted to Reinhold Baer for the above references (as well as for the Neumann reference of §30).

call the m -group an extension of the 2-group, and say that it is reducible to that 2-group. Note that while the coset theorem presented an arbitrary polyadic group in a somewhat similar light, the elements of the polyadic group formed but a proper subclass of the class of elements of the 2-group; whereas, when a polyadic group is reducible to a 2-group, the classes of elements are identical.

More generally, given a μ -group with class of elements C and operation $c_\mu(s_1s_2 \cdots s_\mu)$, if m is any number in the form $k(\mu-1)+1$ we can form the extended operation $c_\mu(s_1s_2 \cdots s_m) = c_\mu(s_1s_2 \cdots s_{\mu-1}c_\mu(s_\mu s_{\mu+1} \cdots s_{2\mu-2} (\cdots c_\mu(s_{(k-1)(\mu-1)+1}s_{(k-1)(\mu-1)+2} \cdots s_{k(\mu-1)+1}) \cdots)))$. The members of C will then form an m -adic group under the operation $c_m(s_1s_2 \cdots s_m) = c_\mu(s_1s_2 \cdots s_m)$. As before, the m -group will be said to be an extension of the μ -group, and reducible to the μ -group.

An m -adic operation on a finite number of elements is most naturally exhibited by an m -dimensional table. We shall therefore say that an m -adic group is of dimension m . We then see that while a 2-group has an extension for each dimension $m > 2$, a μ -group has an extension for those and only those dimensions m for which $m-1$ is a multiple of $\mu-1$.

A given m -group will be said to be *reducible to a μ -group* if there exists a μ -group to which it is reducible. The m -group will be said to be irreducible if it is not reducible to a μ -group for any $\mu < m$ ⁽²⁸⁾. Dörnte has already given a necessary and sufficient condition that a polyadic group be reducible to a 2-group. We proceed to generalize this result to reducibility to a μ -group.

A $(\mu-1)$ -ad $\{a_1, a_2, \dots, a_{\mu-1}\}$ will be said to be commutative with an element a if the μ -ads $\{a_1, a_2, \dots, a_{\mu-1}, a\}$ and $\{a, a_1, a_2, \dots, a_{\mu-1}\}$ are equivalent. We then have the following basic theorem on reducibility. *A necessary and sufficient condition that a given m -group be reducible to a μ -group, $m = k(\mu-1)+1$, is that there be a $(\mu-1)$ -ad $\{a_1, a_2, \dots, a_{\mu-1}\}$ formed from elements of the m -group such that the $(\mu-1)$ -ad is commutative with every element of the m -group, and such that the $(m-1)$ -ad $\{a_1, a_2, \dots, a_{\mu-1}, a_1, a_2, \dots, a_{\mu-1}, \dots, a_1, a_2, \dots, a_{\mu-1}\}$ is an identity of the m -group.*

The necessity of this condition follows immediately from the existence and properties of identities. For, if the m -group is reducible to a μ -group, let $\{a_1, a_2, \dots, a_{\mu-1}\}$ be an identity of such a μ -group. If c_μ is the operation of the μ -group, $c_\mu(a_1a_2 \cdots a_{\mu-1}s) = s = c_\mu(sa_1a_2 \cdots a_{\mu-1})$ for every element s of the μ -group. Hence $\{a_1, a_2, \dots, a_{\mu-1}\}$ is commutative with every element of the μ -group, and hence, by the hypothesis of reducibility, with every element of the m -group. Furthermore, the $(m-1)$ -ad $\{a_1, a_2, \dots, a_{\mu-1}, a_1, a_2, \dots, a_{\mu-1}, \dots, a_1, a_2, \dots, a_{\mu-1}\}$ is an extended identity of the μ -group, and hence an identity of the m -group, as was to be proved.

As for the sufficiency of the condition, with $\{a_1, a_2, \dots, a_{\mu-1}\}$ as in the

⁽²⁸⁾ "Echt" in Dörnte. Otherwise, "unecht" or "ableitbar."

hypothesis, define the μ -adic operation

$$c_\mu(s_1s_2 \cdots s_\mu) = c_m(s_1s_2 \cdots s_\mu a_1a_2 \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1}).$$

We proceed to prove that the elements of the m -group constitute a μ -group under the operation c_μ , and that the given m -group is reducible to this μ -group. Of the two conditions defining a polyadic group, condition 1 is satisfied by the proposed μ -group as an immediate consequence of its being satisfied by the given m -group. On the other hand, condition 2 for the μ -group becomes

$$\begin{aligned}
& c_m(c_m(s_1s_2 \cdots s_\mu a_1a_2 \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1})s_{\mu+1} \\
& \quad \cdots s_{2\mu-1}a_1a_2 \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1}) \\
= & c_m(s_1c_m(s_2s_3 \cdots s_{\mu+1}a_1a_2 \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1})s_{\mu+2} \\
& \quad \cdots s_{2\mu-1}a_1a_2 \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1}) \\
& \cdots \\
= & c_m(s_1s_2 \cdots s_{\mu-1}c_m(s_\mu s_{\mu+1} \cdots s_{2\mu-1}a_1a_2 \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1})a_1a_2 \\
& \quad \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1}),
\end{aligned}$$

which follows from condition 2 for the m -group, and the commutativity of $\{a_1, a_2, \dots, a_{\mu-1}\}$ with each element of the m -group. Hence the existence of the μ -group. Finally, using extended operations, and applying the commutativity part of our hypothesis, we will have

$$c_\mu(s_1s_2 \cdots s_m) = c_m(s_1s_2 \cdots s_ma_1a_2 \cdots a_{\mu-1} \cdots a_1a_2 \cdots a_{\mu-1}) = c_m(s_1s_2 \cdots s_m),$$

the second expression involving a sequence consisting of $k(k-1)$ sequences $a_1a_2 \cdots a_{\mu-1}$, which sequence, therefore, constitutes an extended identity of the m -group—since by hypothesis k such sequences constitute an identity. Hence the reducibility of the m -group to the μ -group follows.

From the definition of c_μ , we see that the $(\mu-1)$ -ad $\{a_1, a_2, \dots, a_{\mu-1}\}$ is indeed an identity of the resulting μ -group.

The above theorem may be used to prove a polyadic group irreducible, as is shown by the following simple illustration. The class of integers constitutes an infinite m -adic group under the operation $s_1 + s_2 + \dots + s_m + 1$. Since the group is abelian, reducibility to a μ -group with $m = k(\mu - 1) + 1$ is equivalent to the existence of an integer a such that $ka + s + 1 = s$, that is, $ka = -1$, which is impossible for any integral $k > 1$. Hence the m -group is irreducible.

The commutativity condition can be restated to read $\{a_1, a_2, \dots, a_{\mu-1}\}$ is invariant under the m -group. Since the present multiplicity of basic operations makes us refrain from employing the simplifications of the coset theorem, the concept of invariance is preferable only for $\mu-1=1$. Our $(\mu-1)$ -ad is now a single element a ; and the further condition that the $(m-1)$ -ad

$\{a, a, \dots, a\}$ be an identity of the m -group may be restated to read: a is of first order. For this condition is equivalent to $c_m(aa \dots aa) = a$. We may therefore state the special result, a rewording only of Dörnte's, a necessary and sufficient condition that a given m -group be reducible to an ordinary group is that the m -group possess an invariant element of first order. Our succeeding development will reveal many general classes of polyadic groups that can be proved reducible to 2-groups. One such class is already at hand, that is, all m -adic quotient groups arising from invariant subgroups of m -adic groups are reducible to 2-groups. For the element of the quotient group corresponding to the invariant subgroup is immediately seen to be invariant under the quotient group, and of m -adic order one. In this connection we may observe that semi-invariant subgroups also lead to special kinds of polyadic quotient groups, for the element corresponding to that semi-invariant subgroup must again be of first order. On the other hand, any polyadic group can be a quotient group in our most general sense; for, with H_0 the identity of G_0 , G/H_0 is identical with G .

Given an m -adic group G , we may ask for the distribution of, and interrelations between, the polyadic groups to which it is reducible. Note immediately that if G is reducible to G' , and G' to G'' , G is reducible to G'' , so that the class of groups to which G' is reducible is a subclass of the class of groups to which G is reducible whenever G is reducible to G' . Our results are of two kinds, both derived from the above theorem.

The first type of result is not much more than a restatement of the condition of the theorem. We recall that, if G is reducible to G' , the class of elements of G is identical with the class of elements of G' , while the operation of G is an extended operation of G' . It follows that a class of equivalent i -ads of G is also a class of equivalent i -ads of G' , and conversely. In particular, the class of identities of G' is a class of equivalent polyads⁽²⁹⁾ of G , so that the classes of identities of two groups to which G may be reducible are either the same or mutually exclusive.

When the classes of identities are distinct, the two groups in question will be distinct, as their operations cannot then be identical⁽³⁰⁾. On the other hand, we easily see that when the classes of identities are the same, the groups are identical. For, if their operations are c' and c'' , then, with $\{a_1, a_2, \dots, a_{m-1}\}$ an identity of each, we have

$$c'(s_1 s_2 \dots s_\mu) = c(s_1 s_2 \dots s_\mu a_1 a_2 \dots a_{m-1} \dots a_1 a_2 \dots a_{m-1}) = c''(s_1 s_2 \dots s_\mu),$$

(29) By a class of equivalent polyads we mean a class of equivalent i -ads for some fixed i . While the elements of G^* as first written are classes of equivalent i -ads with $1 \leq i \leq m-1$, in general no such restriction is intended by the above phrase. As suggested in §2, by the use of extended operations the concept of equivalent i -ads becomes valid for $i > m-1$. This observation will be of greater importance later in the present section.

(30) They may however be "abstractly the same" in the sense of being simply isomorphic. See the opening paragraph of §23.

c being an extended operation of each group. Observe finally that in the sufficiency proof of our basic theorem, and in the succeeding observation, if $\{a_1, a_2, \dots, a_{\mu-1}\}$ satisfies the given condition of that theorem, each $(\mu-1)$ -ad equivalent to $\{a_1, a_2, \dots, a_{\mu-1}\}$ also does. We therefore can state the following result. *There is a 1-1 correspondence between the groups to which a given m -adic group is reducible and the classes of equivalent polyads satisfying the condition of the basic theorem, each such class of equivalent polyads being the class of identities of the corresponding group.*

In particular, there are as many 2-groups to which an m -adic group is reducible as there are invariant elements of order one in the m -group⁽³¹⁾. Thus, consider an ordinary abelian group of finite order g . If d is any divisor of g , there are at least d elements a in this 2-group with $a^d=1$. If this 2-group be extended to a $(d+1)$ -group, each such element a is of order one in the $(d+1)$ -group, and invariant therein. The $(d+1)$ -group is therefore reducible to at least d distinct 2-groups, each such a , in fact, being the identity of the corresponding 2-group.

Our second type of result concerns the possible dimensions of the groups to which a given polyadic group is reducible. The complete result is an immediate consequence of the following theorem. *If an m -group is reducible to a μ_1 -group and a μ_2 -group, it is reducible to a μ -group where $\mu-1$ is the highest common factor of μ_1-1 and μ_2-1 .* To prove this theorem let $\{a'_1, a'_2, \dots, a'_{\mu_1-1}\}$ and $\{a''_1, a''_2, \dots, a''_{\mu_2-1}\}$ be identities of the μ_1 -group and μ_2 -group respectively. They then satisfy the condition of our basic theorem. Furthermore, all but one of the letters in each can be chosen arbitrarily.

If then $\mu_1 > \mu_2$, we may assume $a'_1 = a''_1, \dots, a'_{\mu_2-1} = a''_{\mu_2-1}$. Consider then the sequence $\{a'_{\mu_2}, \dots, a'_{\mu_1-1}\}$ which we shall write $\{a'''_1, \dots, a'''_{\mu_3-1}\}$, with $\mu_3-1 = (\mu_1-1) - (\mu_2-1)$. Then all but one of the letters of this sequence are arbitrary. Inductively, we thus obtain the sequence $\{a^{(\lambda)}_1, \dots, a^{(\lambda)}_{\mu_{\lambda}-1}\}$, with all but one letter arbitrary, from the sequence $\{a^{(\lambda-1)}_1, \dots, a^{(\lambda-1)}_{\mu_{\lambda-1}-1}\}$ and the smallest preceding sequence, easily seen to be unique. Clearly the process terminates when and only when $\mu_{\lambda-1}$ is equal to the smallest preceding μ .

Now in terms of the $\mu_{\lambda}-1$'s, this process is nothing more than the Euclid algorithm for finding the highest common factor of μ_1-1 and μ_2-1 , where the process of division is replaced by the more primitive form of repeated subtractions. Hence, the above process terminates, and the last sequence found may be written $\{a_1, \dots, a_{\mu-1}\}$, where $\mu-1$ is the highest common factor of μ_1-1 and μ_2-1 . We now prove that such a $(\mu-1)$ -ad satisfies the condition of our basic theorem.

First, the sequence $\{a'''_1, \dots, a'''_{\mu_3-1}\}$ is commutative with every element of the given m -group. For we have $\{a'_1, \dots, a'_{\mu_1-1}\} = \{a''_1, \dots, a''_{\mu_2-1}, a'''_1, \dots, a'''_{\mu_3-1}\}$, so that $c(a''_1 \cdots a''_{\mu_2-1} a'''_2 \cdots a'''_{\mu_3-1} s_1 s_2 \cdots s_m) = c(s_1 a''_1 \cdots$

(31) In the case of abelian triadic groups this reduces to a theorem of Lehmer's.

$a''_{\mu_2-1}a'''_1 \cdots a'''_{\mu_3-1}s_{i+1} \cdots s_m) = c(a''_1 \cdots a''_{\mu_2-1}s_i a'''_1 \cdots a'''_{\mu_3-1}s_{i+1} \cdots s_m)$. Hence, by induction, each $\{a_1^{(\lambda)}, \dots, a_{\mu_\lambda-1}^{(\lambda)}\}$ is commutative with every element of the m -group, and so $\{a_1, \dots, a_{\mu-1}\}$ also is thus commutative.

As for the second part of the condition, clearly $m-1=k(\mu-1)$ with integral k . As in the commutativity argument, and with the commutativity property, we obtain from the extended identities consisting of k sequences $\{a'_1, \dots, a'_{\mu_1-1}\}$ and k sequences $\{a''_1, \dots, a''_{\mu_2-1}\}$ an extended identity consisting of k sequences $\{a'_1, \dots, a'_{\mu_3-1}\}$. By induction, k sequences $\{a_1^{(\lambda)}, \dots, a_{\mu_\lambda-1}^{(\lambda)}\}$ constitute an extended identity for every λ , and hence the same is true of k sequences $\{a_1, \dots, a_{\mu-1}\}$. But, since $k(\mu-1)=m-1$, the last is indeed an identity of our given m -group: $\{a_1, \dots, a_{\mu-1}\}$ therefore satisfies completely the condition of our basic theorem, whence the present result.

It follows that if μ_0 is the least dimension of the groups to which a given m -group is reducible, all other dimensions μ of such groups must be such that $\mu-1$ is a multiple of μ_0-1 . We shall call μ_0 the *real dimension* of the m -group, with, of course, $\mu_0=m$ if the group is irreducible. Since every $\mu-1$ must also be a divisor of $m-1$, we easily obtain the following solution of the problem of the distribution of the dimensions of the groups to which a given polyadic group is reducible. *If a group of dimension m has real dimension μ_0 , and we write $m-1=k_0(\mu_0-1)$, then the dimensions of the groups to which the m -group is reducible are those and only those numbers μ for which $\mu-1=k(\mu_0-1)$, k a proper divisor of k_0 .*

While this result justifies the term *real dimension* on the basis of a mere enumeration of distinct dimensions, other considerations show that an m -group in general, even if reducible, must still be considered an m -group. We have already given an example which shows that the same m -group may be reducible to different groups of the same dimension, and, indeed, of the real dimension of the m -group. We now further observe that an m -group may be reducible to an irreducible group of higher dimension than the real dimension of the m -group, that is, not every succession of reductions of a group need lead to the real dimension of the group. If we call the dimensions of the irreducible groups to which a polyadic group is reducible the *irreducible dimensions* of the given group, the real dimension of the group is only the smallest of its irreducible dimensions.

In contrast with the class of groups to which an m -group is reducible, the class of extensions of an m -group is of very simple structure, since it has one and only one group of each dimension μ with $\mu-1$ a multiple of $m-1$, and no others. Of course, the reason is that extension is the direct process, reduction indirect. We now combine these processes to yield the concept of derived group.

Given an m -group G , a polyadic group G' will be said to be *derivable from G* if it can be obtained from G by a finite succession of extensions and reductions.

The class of all polyadic groups derivable from a given polyadic group will be called a *net* of polyadic groups. From this definition we see that each group of a net yields that net. Furthermore, all groups of a given net have the same class of elements; only the operations differ.

The concept of a net of polyadic groups is considerably simplified by the following result. *Any group of a net can be obtained from any other by a single extension followed by a single reduction.* A single extension or a single reduction can obviously be replaced by an extension followed by a reduction. Since two successive extensions are equivalent to a single extension, two successive reductions to a single reduction, our result will follow if we can show that a reduction followed by an extension is equivalent to an extension followed by a reduction. Let then G' with operation $c'_{m'}$ be reducible to G'' with operation $c''_{m''}$, and let G'' be extended to G''' with operation $c'''_{m'''}$. With the above subscripts designating dimensionality, we have $m'-1=k'(m''-1)$, $m'''-1=k''(m''-1)$. Now $c'_{m'}$ and $c'''_{m'''}$ are both extensions of operation $c''_{m''}$. If then we extend $c''_{m''}$ to an operation $c^{IV}_{m^{IV}}$ with $m^{IV}-1=k'k''(m''-1)$, $c^{IV}_{m^{IV}}$ will be an extension of both $c'_{m'}$ and $c'''_{m'''}$. The corresponding group G^{IV} is then reducible to both G' and G'' , whence our result.

Stated otherwise, *given any two groups of a net there is a third group of the net reducible to each of the given groups.* We could therefore redefine a net as the class of groups to which the extensions of a given group are reducible, though the conclusion that a net does not depend on the particular group in it chosen as the given group is then not immediate.

The two types of results referred to in the case of the groups to which a given group is reducible now easily lead to corresponding results for the net of groups derivable from a given group. In this connection, a $(\mu-1)$ -ad $\{a_1, a_2, \dots, a_{\mu-1}\}$ of an m -group will be said to be of finite order if some polyad of the form $\{a_1, a_2, \dots, a_{\mu-1}, a_1, a_2, \dots, a_{\mu-1}, \dots, a_1, a_2, \dots, a_{\mu-1}\}$ is an extended identity of the m -group. We then easily prove the following. *There is a 1-1 correspondence between the groups of the net of groups derivable from a given group and the classes of equivalent polyads of finite order which are commutative with every element of the given group, each such class of equivalent polyads then being the class of identities of the corresponding group*⁽³²⁾. In fact, the above redefinition of a net immediately yields a many-one correspondence of the above type, which is then seen to be one-one due to any pair of groups of a net being in the class of groups to which a third is reducible.

Actually, it is easily verified that each of the concepts: class of equivalent polyads, commutative with every element, and even polyad of finite order, is independent of the particular group of the net chosen as given group, so that the above result can be restated in terms of the net alone. It is also easily proved that for finite polyadic groups every polyad is of finite order, so that

(32) Here, as elsewhere, "group" unqualified means polyadic group.

in such cases the corresponding condition need not be explicitly stated. In particular, there are as many 2-groups in the net as there are invariant elements of finite order, and hence, for finite polyadic groups, as many as there are invariant elements.

We pause to prove explicitly that the transform of one element of a group of a net by another is independent of the particular group employed. This will be so if true of any pair of groups, one reducible to the other. Since the operation of one of these groups is an extended operation of the other, an identity of the first group is an extended identity of the second; hence an inverse of an element in the first, an extended inverse of that element in the second, whence the identical transforms.

The second type of result is obtained still more easily. We shall call the least dimension of the groups of a net their *outer real dimension*. The outer real dimension of a group is then always less than or equal to its real dimension. Given an m -group G of outer real dimension μ^0 , some third group G' of the net will be reducible both to the m -group, and a group of dimension μ^0 . The real dimension of G' will therefore exactly equal μ^0 . As G' is reducible to G , we see that $m - 1$ is a multiple of $\mu^0 - 1$. That is, if the outer real dimension of an m -group is μ^0 , then $\mu^0 - 1$ must be a divisor of $m - 1$.

Hence, also, all the groups of the net have dimensions μ with $\mu - 1$ a multiple of $\mu^0 - 1$. Since, from a group of dimension μ^0 , mere extensions yield groups of all such dimensions, we have the following main result. *If the outer real dimension of the groups of a net is μ^0 , their dimensions are those and only those numbers μ for which $\mu - 1 = k(\mu^0 - 1)$.*

The first type of result is easily restated to yield a criterion for determining the outer real dimension of a group. In particular, *the outer real dimension of a group is 2 when and only when it contains an invariant element of finite order*. Thus, a finite abelian polyadic group is always of outer real dimension 2, and so is derivable from a 2-group, while a group having no invariant element is always of outer real dimension greater than 2. The existence of the latter type of group is peculiar to polyadic theory. A simple example is furnished by the class of odd substitutions of the symmetric group of degree three. By the converse of the coset theorem they form a triadic group of order three under the product of three substitutions as operation, and yet involve no invariant element. The three elements, incidentally, are all of first order in the triadic group.

As in the case of mere reducibility, we shall call the dimensions of the irreducible groups of a net the *outer irreducible dimensions* of each group in the net. By contrast, a dimension will be said to be a *reducible dimension* of the groups of the net if there is at least one group of the net of that dimension, while all such groups are reducible. While we have no general theorem giving the distribution of these dimensions, the following special results lend a certain insight into the possibilities involved.

First, a group may have its real dimension as its only outer irreducible dimension. This is readily proved to be so for any 2-group which has no invariant element other than the identity. In this case, in fact, the net of groups consists only of the 2-group, and its extensions.

By contrast, a group may have an infinite number of outer irreducible dimensions. Thus it can be shown that for the ordinary cyclic group of order two the outer irreducible dimensions are the infinite set of numbers of the form $2^n + 1$, $n = 0, 1, 2, \dots$.

Finally, it can be shown that every finite polyadic group has an infinite number of reducible dimensions. To be specific, if an m -group has g elements, there is, of course, at least one group of the net of dimension $(kg+1)(m-1)+1$, for each $k=1, 2, 3, \dots$, and every group of the net of such a dimension is reducible, reducible to dimension m , in fact.

We append a brief discussion of the generalization of the concept of a net of groups that arises from a consideration of the subgroups of a group. Let the *complex* of groups obtainable from a given polyadic group be the class of all polyadic groups obtainable from the given group by finite successions of the three operations "extension of," "reduction of," and "subgroup of." It is readily verified by means of the very concepts involved that an extension of a subgroup of a group is also a subgroup of an extension of a group; and that a subgroup of a reduction of a group is also a reduction of a subgroup of the group. It follows that *any group in a complex can be obtained from the given group by an operation of the single form "extension of" followed by "subgroup of" followed by "reduction of" if not merely by "extension of" followed by "reduction of."*

In the case of abelian groups we further have that a reduction of a subgroup of a group is also a subgroup of a reduction of the group, a result obtainable with the help of our criterion of reducibility. It follows that *the complex of groups obtainable from an abelian polyadic group consists of the groups in the corresponding net of groups, and their subgroups.* That this is not true for all complexes can be seen from the case of a group with a first order element, but no invariant element. For the first order element constitutes a subgroup of the given group reducible to a 2-group; while, the outer real dimension of the given group being greater than 2, the dimensions of all the groups in the net, and hence of their subgroups, is greater than 2.

It is readily seen that the groups of a complex whose classes of elements are the same as that of the original group constitute the net of that group, or, as we shall now phrase it, the net of the complex. Clearly the net of a complex also consists of all of its groups from which that complex is obtainable. On the other hand, a group of a complex with class of elements a proper subclass of that of the original group will yield a complex which is a proper subclass of the given complex, and may be called a subcomplex thereof. If we call the nets of the subcomplexes of a complex the subnets of that complex,

then it is clear that the net and subnets of a complex constitute a separation of the groups of the complex into mutually exclusive sets.

The relationship between the subcomplexes of a complex is in part furnished by the following result. *If of two groups in a complex the class of elements of the first group is contained in the class of elements of the second, then the first group is in the complex obtained from the second.* For consider the two groups to be obtained from an initial group according to our first result. Using (c_m, C) to designate a group with m -adic operation c_m and class of elements C , we may indicate the process as follows:

$$(c_m, C) \rightarrow (c'_m, C) \rightarrow (c''_m, C') \rightarrow (c'''_m, C''),$$

$$(c_m, C) \rightarrow (c''''_m, C) \rightarrow (c''''_m, C'') \rightarrow (c^{IV}_m, C'').$$

The two groups in the second column are also reductions of a third group (c^v_{mv}, C) . Since the third column symbolizes groups, it follows that (c^v_{mv}, C') and (c^v_{mv}, C'') are groups; and as C' is contained in C'' by hypothesis, (c^v_{mv}, C') is a subgroup of (c^v_{mv}, C'') , if not identical with it. Now (c^v_{mv}, C') , (c'_m, C') and (c''_m, C') are in a single net of groups, as are also (c^v_{mv}, C'') , (c'''_m, C'') and (c^{IV}_m, C'') . Hence (c''_m, C') is in the complex obtainable from (c^{IV}_m, C'') , as was to be proved.

A particular application of the above result is the following. *Any two groups of a complex which have the same class of elements are derivable from each other, that is, belong to one and the same net.* It follows that there is a 1-1 correspondence between the subnets, including the net, into which the groups of a complex were separated, and the different classes of elements of the groups in the complex.

Hence also, or directly from our general result, there is a 1-1 correspondence between the subcomplexes, including the complex, of a complex, and the different classes of elements of the groups in the complex, each complex being obtainable from those and only those groups whose classes of elements are identical with the class of elements corresponding to the complex. Moreover, our general result shows that one subcomplex contains a second when and only when the class of elements corresponding to the first contains the class of elements corresponding to the second. We now complete this picture by proving the following. *If two subcomplexes K' and K'' of a complex correspond to the classes of elements C' and C'' , then the logical product of K' and K'' , null when the logical product of C' and C'' is null, is otherwise a complex, namely the complex corresponding to the logical product of C' and C'' .* For C' and C'' must be the classes of elements of two groups (c'_m, C') and (c^{IV}_m, C'') of the complex. In the notation of the previous proof, (c^v_{mv}, C') and (c^v_{mv}, C'') are then groups of the complex. If then C''' , the logical product of C' and C'' , is not null, (c^v_{mv}, C''') is a group of the complex. The case C''' null is immediate. Otherwise, then, there will be a subcomplex K''' corresponding to C''' .

Our earlier result then shows immediately that a group G is common to K' and K'' when and only when it is in K''' .

Further results on the subcomplexes of a complex obtained from a finite polyadic group, and more particularly a finite abelian polyadic group, will be found at the end of §22, our second section on cyclic polyadic groups⁽³³⁾.

6. Arbitrary containing ordinary groups. The coset theorem led to the abstract containing ordinary group G^* of an m -group G merely by a consideration of G treated abstractly. Often, however, the elements of G may immediately be given in such a form that the m -adic operation is but an extension of a more primitive dyadic operation, as when G is an m -adic group of ordinary substitutions. In such a case a containing 2-group arises directly, and may be more useful than the abstract containing group.

A 2-group G^{**} will be called a *containing group* of an m -group G if the elements of G are among the elements of G^{**} , the operation of G an extension of the operation of G^{**} , while G^{**} is generated by the elements of G . In what follows we simultaneously investigate the possible structure of G^{**} , and its relationship to G^* . We must therefore explicitly distinguish between their operations c^{**} and c^* respectively⁽³⁴⁾.

Let two polyads $\{s_1, s_2, \dots, s_i\}$ and $\{s'_1, s'_2, \dots, s'_{i'}\}$ of G lead to identical products in G^* ; that is, let $c^*(s_1s_2 \cdots s_i) = c^*(s'_1s'_2 \cdots s'_{i'})$. Since $i' - i$ must then be a multiple of $m - 1$, we can annex elements $s''_1, \dots, s''_{i''}$ of G , if need be, so that the resulting equation $c^*(s_1s_2 \cdots s_i s''_1 \cdots s''_{i''}) = c^*(s'_1s'_2 \cdots s'_{i'} s''_1 \cdots s''_{i''})$ can be rewritten $c(s_1s_2 \cdots s_i s''_1 \cdots s''_{i''}) = c(s'_1s'_2 \cdots s'_{i'} s''_1 \cdots s''_{i''})$ in, perhaps, extended notation. But this equation can now be written $c^{**}(s_1s_2 \cdots s_i s''_1 \cdots s''_{i''}) = c^{**}(s'_1s'_2 \cdots s'_{i'} s''_1 \cdots s''_{i''})$, whence we obtain $c^{**}(s_1s_2 \cdots s_i) = c^{**}(s'_1s'_2 \cdots s'_{i'})$. That is, if two polyads of G lead to identical products in G^* they lead to identical products in G^{**} . If then we let every element of the form $c^{**}(s_1s_2 \cdots s_i)$ in G^{**} correspond to element $c^*(s_1s_2 \cdots s_i)$ of G^* , a one-many correspondence is set up between those elements of G^{**} and of G^* which are obtainable as products of elements of G .

This correspondence is clearly preserved under the respective operations of these groups. For if r_1 and r_2 of G^* correspond to r'_1 and r'_2 respectively

(33) The development of the section just ended, lengthy as it is, is probably but one of many possible developments leading to sets of related polyadic groups. Dörnte's Theorem 7, §2, can probably be made the starting point for such a different development. The possibilities are further widened if a theory is contemplated which would include the relationship between a polyadic group and the corresponding "schar."

(34) It might be thought that now, when the ordinary group demanded by Miller's theorem is immediately given, at least the structure of G^{**} requires no further investigation. But, apart from the fact that Miller's theorem is given for finite groups, his hypothesis that for some integer n the products of any n but no fewer elements of G is in G is not immediately given, but is replaced by G 's being an m -group. As we also need the relationship between G^* and G^{**} , we make our development entirely independent of Miller's.

of G^* , by writing these elements as corresponding products of elements in G we see immediately that $c^*(r_1 r_2)$ corresponds to $c^{**}(r'_1 r'_2)$. Since G^* consists of the products of elements in G , it easily follows that the products in G^{**} of elements of G themselves constitute a group which can then be none other than G^{**} ; for G^{**} is generated by G . Furthermore our one-many correspondence, which is therefore a correspondence between all the elements of G^{**} and of G^* , is indeed a one-many isomorphism between G^{**} and G^* .

For fixed i we shall call the set of elements of G^{**} which are the products of i elements of G the i th coset of G^{**} . For these elements the above set of equations can be reversed so that our one-many correspondence between G^{**} and G^* becomes a 1-1 correspondence between the elements of the i th cosets of G^{**} and of G^* for each $i \geq 1$. From the corresponding result for G^* , it follows that the elements of the i th coset of G^{**} will be obtained in 1-1 fashion if in the expression $c^{**}(s_1 \dots s_{i-1} s)$ we let s_1, \dots, s_{i-1} be arbitrary fixed elements of G , and let s run through G .

Let now k designate the least i for which the corresponding coset of G^{**} contains the identity I' of G^{**} . It follows, first, that the first k cosets of G^{**} are mutually exclusive. For if we could have $c^{**}(s_1 \dots s_i) = c^{**}(s'_1 \dots s'_i)$ with $1 \leq i < j \leq k$, then, by rewriting $c^{**}(s'_1 \dots s'_i)$ in the form $c^{**}(s_1 \dots s_i s'_{i+1} \dots s'_j)$, we would have $c^{**}(s'_{i+1} \dots s'_j) = I'$, in contradiction to our definition of k . On the other hand, the $(k+1)$ -st coset of G^{**} is identical with the first, that is, with G , for we can write its elements in the form $c^{**}(s_1 \dots s_k s)$ with $c^{**}(s_1 \dots s_k) = I'$. Hence also the $(k+2)$ -nd coset is identical with the 2d, and so on. G^{**} therefore consists of the elements of its first k cosets, while succeeding cosets are cyclic repetitions of these. In particular, the $(m-1)$ -st coset must be identical with the k th coset. For if $\{s_1, s_2, \dots, s_{m-1}\}$ is an identity of G , $c^{**}(s_1 s_2 \dots s_{m-1}) = I'$, so that the $(m-1)$ -st and k th cosets have an element in common. Hence k is a divisor of $m-1$.

Returning to our correspondence between the elements of G^{**} and of G^* we see that it is 1-1 between the elements of G^{**} and the elements of the first k cosets of G^* , and of each succeeding set of k cosets of G^* . Our one-many correspondence is thus actually $[1, (m-1)/k]$, and we therefore have a $[1, (m-1)/k]$ isomorphism between G^{**} and G^* . To complete our analysis we consider the analogue in G^{**} of the associated 2-group G_0 of G in G^* .

Our $[1, (m-1)/k]$ correspondence is clearly 1-1 between the elements of the k th coset of G^{**} , and of G_0 , the $(m-1)$ -st coset of G^* . Since the product of two elements of the k th coset of G^{**} is in the $2k$ th coset, and hence also in the k th coset, of G^{**} , the previous $[1, (m-1)/k]$ isomorphism between G^{**} and G^* is simple between the k th coset of G^{**} , and G_0 . It follows that the k th coset of G^{**} constitutes a group with operation c^{**} simply isomorphic with G_0 . We shall call it the associated ordinary group of G in G^{**} , and symbolize it G'_0 . The same argument used in proving G_0 invariant under G^* shows G'_0 to be invariant under G^{**} .

Since the i th coset of $G^{* \prime}$ is given by $c^{* \prime}(s_1 \cdots s_{i-1}s)$, with s_1, \dots, s_{i-1} fixed elements of G , s running through G , we can let s_1, \dots, s_{i-1} be the same element s_0 of G , and write that i th coset $s_0^{i-1}G$ in ordinary notation. It can likewise be written Gs_0^{i-1} . We thus obtain the expansion $G^{* \prime} = G + Gs_0 + Gs_0^2 + \cdots + Gs_0^{k-1}$. Since $Gs_0^{k-1} = G'_0$, and $Gs_0^k = G$, we therefore have

$$G = G'_0 s_0,$$

while the above expansion becomes

$$G^{* \prime} = G'_0 s_0 + G'_0 s_0^2 + \cdots + G'_0 s_0^{k-1} + G'_0.$$

But this is the expansion of $G^{* \prime}$ in augmented cosets as regards the invariant subgroup G'_0 , assuming G'_0 is not itself $G^{* \prime}$. It follows that the quotient group $G^{* \prime}/G'_0$ is of index k , while the element in that quotient group corresponding to G generates $G^{* \prime}/G'_0$.

This concludes our discussion of the structure of $G^{* \prime}$. As for its isomorphism with G^* , observe first that in that isomorphism elements of G correspond to themselves. We then see that the isomorphism between $G^{* \prime}$ and G^* is determined by this partial correspondence provided k , and the element of the k th coset of $G^{* \prime}$ which serves as the identity of $G^{* \prime}$, are specified. For the correspondence between elements of G and themselves determines the 1-1 correspondence between the elements of the i th cosets of $G^{* \prime}$ and of G^* for every i . And given k , and $c^{* \prime}(s_1^0 s_2^0 \cdots s_k^0) = I'$, s 's in G , if $j = kk+l$, $1 \leq l \leq k$, the equation $c^{* \prime}(s_1^0 s_2^0 \cdots s_k^0 \cdots s_1^0 s_2^0 \cdots s_k^0 s_1 s_2 \cdots s_i) = c^{* \prime}(s_1 s_2 \cdots s_i)$ serves to identify each symbolized element of the j th coset of $G^{* \prime}$ with a unique element of the l th coset, and thus completes the correspondence between the elements of $G^{* \prime}$ and G^* . In particular, the simple isomorphism between G'_0 and G_0 is also thus determined. We therefore have the following comprehensive theorem:

Every containing 2-group $G^{ \prime}$ of an m -group G , if not itself a 2-group G'_0 to which G is reducible, contains an invariant subgroup G'_0 of index k , with k a divisor of $m-1$, G a coset of $G^{* \prime}$ as regards G'_0 , and the quotient group $G^{* \prime}/G'_0$ generated by the element corresponding to G . Furthermore, $G^{* \prime}$ admits a $[1, (m-1)/k]$ isomorphism with G^* , the abstract containing 2-group of G , which reduces to a simple isomorphism between G'_0 and G_0 , the associated 2-group of G . This isomorphism makes each element of G correspond to itself, and is, in fact, determined by this correspondence when k , which is the smallest i for which an i -ad of G yields the identity of $G^{* \prime}$, as well as the class of equivalent k -ads of G thus yielding the identity of $G^{* \prime}$, are specified.*

We shall call k the index of the containing 2-group. We have then, in particular, that any two containing groups of index $m-1$ of an m -group are simply isomorphic, the isomorphism in question making each element of the m -group correspond to itself, and being in turn determined by this correspondence. Hence,

any containing group of index $m-1$ of an m -group G may be considered to be the abstract containing group G^* of G .

We further have that *any two containing groups of index 1 of an m -group are simply isomorphic*. For the G^{**} 's are then also the G'_0 's which are both simply isomorphic with G_0 . Observe, however, that the simple isomorphism now no longer makes elements of G correspond to themselves, or the G^{**} 's would be identical. In fact, a different element of G serves as identity in each G^{**} . Since G is now reducible to G^{**} , and conversely, we have as a corollary the following result on the 2-groups to which an m -group is reducible, and hence also on the 2-groups in a net. *All 2-groups in a net of groups are simply isomorphic.*

Before considering the same question for two containing groups of index k , $1 < k < m-1$, we ask when an m -group will admit a containing 2-group of index k . We then easily obtain the following theorem. *A necessary and sufficient condition that an m -group admit a containing group of index k , $k < m-1$, is that the m -group be reducible to a $(k+1)$ -group.* In fact, the observation that in a containing group G^{**} of index k the products of $k+1$ elements of G must be in G is easily extended to show that the elements of G constitute a $(k+1)$ -group under the operation $c^{**}(s_1s_2 \cdots s_{k+1})$. As k is a divisor of $m-1$, the operation $c(s_1s_2 \cdots s_m) = c^{**}(s_1s_2 \cdots s_m)$ is an extension of $c^{**}(s_1s_2 \cdots s_{k+1})$, and, consequently, G is reducible to the corresponding $(k+1)$ -group. Conversely, if G is reducible to a $(k+1)$ -group, the abstract containing group of the $(k+1)$ -group is of index k . But this group is clearly also a containing group of G , and of index k . In particular, *an irreducible m -group admits containing groups of index $m-1$ only, and conversely.* Hence, the abstract containing group of an irreducible polyadic group may be said to be its only containing group.

This relation to reducibility shows that there are as many essentially different containing groups of index $k < m-1$ of an m -group G as there are $(k+1)$ -groups to which G is reducible. Hence when $1 < k < m-1$, as when $k=1$, two essentially different containing groups of index k will not admit a simple isomorphism which makes each element of G correspond to itself, since the classes of equivalent k -ads yielding their identities will be different. Moreover, unlike the case $k=1$, they need not even admit a simple isomorphism which transforms the class of elements of G into itself. For our example of a group having an infinite number of outer irreducible dimensions easily leads to a group G reducible to two groups G_1 and G_2 of the same dimension, one reducible, the other irreducible. The abstract containing groups of G_1 and G_2 are containing groups of G of the same index; and did they admit a simple isomorphism of the type in question, G_1 and G_2 would be simply isomorphic, and hence could not be one reducible, the other irreducible.

Finally, a word about the application of arbitrary containing groups of an m -group to the study of the m -group. With the containing group G^{**} specified,

we may use ordinary notation for its operation, and write the operation of the m -group G , $s_1s_2 \cdots s_m$. If s is an element of G , and $\{s', s'', \dots, s^{(m-2)}\}$ an inverse of s , we shall still have $s's'' \cdots s^{(m-2)} = s^{-1}$. It follows that the transform of an element s_1 of G by an element s_2 will be given by $s_2^{-1}s_1s_2$ no matter what the containing group. Likewise, if $s's'' \cdots s^{(i)} = r$ in G^* , the transform of element s of G by the i -ad $\{s', s'', \dots, s^{(i)}\}$ will be given by $r^{-1}sr$. On the other hand, the fact that, for an element s of G , s^{-1} , in a containing group of index $k < m - 1$, can also be written as a product of $k - 1$ elements of G , or as an element of G when $k = 1$, gives but spurious information about the corresponding $(k - 1)$ -ad, or element, of G .

7. Determination of all types of semi-abelianisms. In the notation of the abstract containing group of an m -group G we may write the condition that G be abelian in the form $s_1s_2 = s_2s_1$ for all elements s_1, s_2 of G . Dörnte discovered that an m -group, $m > 2$, may satisfy the weaker type of commutativity property $s_1s_2 \cdots s_{m-1}s_m = s_ms_2 \cdots s_{m-1}s_1$ without necessarily being abelian, and termed such groups semi-abelian. An immediate generalization of the Dörnte type of semi-abelianism is that given by any relation

$$s_1s_2 \cdots s_{\mu-1}s_\mu = s_\mu s_2 \cdots s_{\mu-1}s_1,$$

where $\mu - 1$ is a divisor of $m - 1$, s 's arbitrary elements of G . We shall then say that G is μ -semi-abelian. At least a trivial example of an m -group that is μ -semi-abelian, but not abelian, would be given by the extension to an m -group of a non-abelian μ -group semi-abelian in Dörnte's sense.

In general we shall say that an m -group G is semi-abelian according to the corresponding formal type if for all choices of s 's as elements in G a set of relations of the following form are satisfied:

$$\begin{aligned} s_1s_2 \cdots s_{l'} &= s_{i'_1}s_{i'_2} \cdots s_{i'_{l'}}, \\ s_1s_2 \cdots s_{l''} &= s_{i''_1}s_{i''_2} \cdots s_{i''_{l''}}, \\ &\vdots \\ s_1s_2 \cdots s_{l^{(\lambda)}} &= s_{i^{(\lambda)}_1}s_{i^{(\lambda)}_2} \cdots s_{i^{(\lambda)}_{l^{(\lambda)}}}, \end{aligned}$$

each right-hand member being a specific permutation of the left, not all the permutations being the identity. Given m , two such formulations will then be said to be equivalent, or to define the *same type of semi-abelianism* if all m -groups which are semi-abelian according to one formal type are also semi-abelian according to the other. We proceed to prove that the above extensions of the Dörnte type of semi-abelianism constitute all possible semi-abelianisms. More specifically, by the *displacement* of a letter in a given equation of a set of the above type we shall mean the number of places, right or left, it has to be moved in passing from the left side of the equation to the right. We then prove that *every formal type of semi-abelianism, for given dimension m , is equivalent to μ -semi-abelianism with $\mu - 1$ the highest common factor of $m - 1$*

and all the displacements of the letters in the equations defining the semi-abelianism.

Observe immediately that for $m=2$ there is no semi-abelianism distinct from abelianism. For in some equation a pair of letters s_i, s_j will appear in different orders on opposite sides of the equation; and by replacing all other letters by the identity we obtain the condition for abelianism $s_i s_j = s_j s_i$. This serves to make plausible our general result, and to give a hint of its proof.

In the general case, then, let G be any m -group semi-abelian according to a given formal type, and let some letter s_i have a nonzero displacement k in one of the equations defining that semi-abelianism. Since re-symbolization allows either member of the equation to be written first, we may write the equation

$$s_1 \cdots s_{j-1} s_j s_{j+1} \cdots s_l = s_{i_1} \cdots s_{i_{j+k-1}} s_i s_{i_{j+k+1}} \cdots s_{i_l},$$

so that

$$s_j = [(s_1 \cdots s_{j-1})^{-1} s_{i_1} \cdots s_{i_{j+k-1}}] s_i [s_{i_{j+k+1}} \cdots s_{i_l} (s_{j+1} \cdots s_l)^{-1}].$$

The first bracket is equivalent to some k -ad $s' s'' \cdots s^{(k)}$. Since at least one letter inside that bracket and outside the parenthesis must be different from all the letters in the parenthesis, that k -ad, and hence $s', s'', \dots, s^{(k)}$, can be arbitrary. The second bracket is equivalent to some κ -ad $\bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa)}$. We can always assume $\kappa > 1$, by introducing an identity if need be, and hence at least $\bar{s}^{(\kappa)}$ is arbitrary. That is, for every $s', s'', \dots, s^{(k)}, \bar{s}^{(\kappa)}$, we can find $\bar{s}', \bar{s}'', \dots, \bar{s}^{(\kappa-1)}$ so that

$$s_j = s' s'' \cdots s^{(k)} s_i \bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} \bar{s}^{(\kappa)}$$

for every s_i . Letting $s_i = s'$, we find that $s'' \cdots s^{(k)} s' \bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} \bar{s}^{(\kappa)} = 1$, whence

$$\bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} \bar{s}^{(\kappa)} s'' \cdots s^{(k)} s' = 1.$$

Letting $s_i = \bar{s}^{(\kappa)}$, we find $s' s'' \cdots s^{(k)} \bar{s}^{(\kappa)} \bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} = 1$, whence

$$\bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} \bar{s}^{(\kappa)} s'' \cdots s^{(k)} \bar{s}^{(\kappa)} = 1.$$

It follows that for every $s', s'', \dots, s^{(k)}, \bar{s}^{(\kappa)}$ in G ,

$$s' s'' \cdots s^{(k)} \bar{s}^{(\kappa)} = \bar{s}^{(\kappa)} s'' \cdots s^{(k)} s'.$$

Dropping momentarily the condition $\mu - 1$ a divisor of $m - 1$ in our definition of μ -semi-abelianism, we have therefore proved that for each displacement $k > 0$, G is $(k+1)$ -semi-abelian.

Let now G be (k_1+1) -semi-abelian and (k_2+1) -semi-abelian. We then prove that G is $(k+1)$ -semi-abelian with $k = \text{H.C.F.}(k_1, k_2)$. This will follow if for every such k_1 and k_2 with $k_2 > k_1$, G is (k_2+1) -semi-abelian with $k_2 = k_2 - k_1$. But under our hypothesis, with all other letters unmoved, we have

$s_1 \cdots s_{k_1+1} \cdots s_{k_2+1} = s_{k_2+1} \cdots s_{k_1+1} \cdots s_1 = s_{k_1+1} \cdots s_{k_2+1} \cdots s_1 = s_1 \cdots s_{k_2+1} \cdots s_{k_1+1}$. Hence $s_{k_1+1} \cdots s_{k_2+1} = s_{k_2+1} \cdots s_{k_1+1}$ as desired.

Finally, we show that if the m -group G is $(k+1)$ -semi-abelian, it is also $(k'+1)$ -semi-abelian with $k' = \text{H.C.F.}(k, m-1)$. Since G is $(k+1)$ -semi-abelian, it is also $(\kappa k+1)$ -semi-abelian for every positive integral κ . It is therefore also $(\kappa k'+1)$ -semi-abelian with k' any positive integer in the form $\kappa k - \lambda(m-1)$. For in the equation defining the $(\kappa k+1)$ -semi-abelianism there are at least $\lambda(m-1)$ letters between the first and last letters of each member; and by choosing $\lambda(m-1)$ of these letters consecutively to form an extended identity the desired $(\kappa k'+1)$ -semi-abelianism is revealed. As positive integers κ and λ can always be chosen so that $\kappa k - \lambda(m-1) = \text{H.C.F.}(k, m-1)$, our result follows.

From these three special results it follows that every m -group possessing a given formal type of semi-abelianism is μ -semi-abelian with μ as in the statement of our theorem. It remains to be shown that every m -group that is μ -semi-abelian also satisfies the given formal semi-abelianism. For each of the given equations separates the letters in the left side of the equation into $\mu-1$ mutually exclusive sets such that each set consists of all letters whose "distance" from a given letter is a multiple of $\mu-1$. Since in passing from the left side to the right side of the equation each letter suffers a displacement itself a multiple of $\mu-1$, the result is to permute the letters of each set among themselves. Now a single application of our hypothesis of μ -semi-abelianism to the left side of the equation in question constitutes a transposition of two letters in the same set. As μ -semi-abelianism implies $[\kappa(\mu-1)+1]$ -semi-abelianism, every such transposition can be effected. And, as any substitution is the product of transpositions, successive applications of our hypothesis of semi-abelianism will transform the left side of each equation so that each of its $\mu-1$ sets assumes the form it has on the right. That is, each equation of the given formal semi-abelianism will be satisfied by the elements of any m -group that is μ -semi-abelian. The equivalence in question has therefore been demonstrated.

That μ -semi-abelianism is a different type of semi-abelianism for different divisors $\mu-1$ of $m-1$ is readily proved by examples. By the theorem of the next section, an m -group $G = G_0 s_0$ will be determined by the following hypothesis: G_0 an ordinary cyclic group of order $2^{m-1}-1$ generated by t , $s_0^{m-1} = 1$, $s_0^{-1} t s_0 = t^2$. Since G_0 is abelian, the first result of the next paragraph shows G to be m -semi-abelian. Now a similar argument shows an m -group G to be μ -semi-abelian, $\mu-1$ a divisor of $m-1$, when and only when the $(\mu-1)$ -ads of G are commutative with the $(m-1)$ -ads of G . Since s_0^{m-1} is the first ordinary positive power of s_0 commutative with t , it follows that G is not μ -semi-abelian for any divisor $\mu-1$ of $m-1$ other than $m-1$. Now let μ_1-1 , μ_2-1 be any two distinct divisors of $m-1$ with, say, $\mu_1 > \mu_2$. By the preceding method construct a μ_1 -group G' which is μ_1 -semi-abelian, but not μ_2 -semi-

abelian for any divisor $\mu_3 - 1$ of $\mu_1 - 1$ other than $\mu_1 - 1$. The extension of G' to an m -group G'' then has the same property. It then follows that the m -group G'' while μ_1 -semi-abelian is not μ_2 -semi-abelian, since otherwise it would be μ_3 -semi-abelian with $\mu_3 - 1 = \text{H.C.F.}(\mu_1 - 1, \mu_2 - 1)$, and thus a divisor of $\mu_1 - 1$ other than $\mu_1 - 1$. The m -group G'' thus shows μ_1 -semi-abelianism to be not equivalent to μ_2 -semi-abelianism whenever $\mu_1 \neq \mu_2$. Coupled with our previous theorem it yields the following result. *There are as many distinct types of semi-abelianism for m -adic groups as there are distinct divisors of $m - 1$.*

In what follows we restrict our attention to ordinary, that is, m -semi-abelianism, a property implied by any type of semi-abelianism. Since the associated ordinary group G_0 of an m -group G consists of the products of $m - 1$ arbitrary elements of G , the condition that G_0 is abelian is a condition of semi-abelianism on G of formal type

$$s_1 s_2 \cdots s_{m-1} s_m s_{m+1} \cdots s_{2m-2} = s_m s_{m+1} \cdots s_{2m-2} s_1 s_2 \cdots s_{m-1}.$$

As each letter suffers a displacement $m - 1$, by our general result this type of semi-abelianism is equivalent to m -semi-abelianism. Hence, *every semi-abelian m -group has an abelian associated group, and conversely*. If an element s of a semi-abelian group G is invariant under G , it is also invariant under G_0 , and hence $G = G_0 s$ is abelian. That is, *if a semi-abelian m -group is non-abelian, it has no invariant element*. If s_1 and s_2 are any two elements of semi-abelian G , t any element of G_0 , then, since $s_1 = t's_2$, with t' in G_0 , and since t and t' are commutative, we have $s_1^{-1}ts_1 = s_2^{-1}ts_2$. Hence, *all the elements of a semi-abelian m -group G transform an arbitrary given element of the associated group G_0 into the same element*. Now let H be any subgroup of semi-abelian G . Its associated subgroup H_0 is then invariant under any element s_0 of H . But every element s of G transforms the elements of H_0 as does s_0 . Hence H_0 is invariant under G . That is, *every subgroup of a semi-abelian group is semi-invariant*⁽³⁵⁾.

8. On the construction of polyadic groups. We proceed to prove the following general theorem on the construction of abstract polyadic groups referred to in connection with the converse of the coset theorem. *Given any abstract 2-group G_0 to serve as associated group, an abstract element s_0 subject to the condition $s_0^{m-1} = t_0$, t_0 in G_0 , and any automorphism T of G_0 , which carries t_0 into itself, and whose $(m - 1)$ -st power is the automorphism of G_0 under t_0 , to serve as the automorphism of G_0 under s_0 , then there is one and only one corresponding abstract m -group G ; conversely every m -group can be thus determined*⁽³⁶⁾.

(35) See Dörnte's §7 for quite a different set of properties of semi-abelian groups. Dörnte's result that a triadic group consisting of first order elements only must be semi-abelian is equivalent for finite groups to a result of Miller's as a consequence of the above equivalence of the semi-abelianism of G , and abelianism of G_0 . By introducing the polyadic groups G_i of our §34 to take the place of G_0 in the discussion of the last paragraph, the results of that paragraph can be specifically generalized to μ -semi-abelianism.

(36) After this theorem was obtained by the writer, a closely related result was published by Turing as an illustration of a more general theorem in the theory of group extensions. (Not

For the second part of this theorem note that given an m -group G , and any s_0 in G , G_0 , t_0 , and T are determined, and obviously satisfy the conditions of the theorem. It follows from the first part of the succeeding proof that G is determinable as stated.

We turn then to the first part of the theorem. For purposes of analysis, consider the coset representation of a hypothetical G satisfying the given conditions. We would then have $G = G_0 s_0$. If we write the elements of G_0 as t_i , we may correspondingly symbolize the elements of G by s_i , with $s_i = t_i s_0$. Of course s_0 must then be identified with that s_i for which t_i is the identity of G_0 , while t_0 will appear as some t_k . We must then have, for the operation of G ,

$$\begin{aligned} c(s_{i_1} s_{i_2} \cdots s_{i_m}) &= t_{i_1} s_0 t_{i_2} s_0 \cdots t_{i_m} s_0 = t_{i_1} (s_0 t_{i_2} s_0^{-1}) \cdots (s_0^{m-1} t_{i_m} s_0^{-m+1}) s_0^m \\ &= (t_{i_1} \cdot T^{-1} t_{i_2} \cdots T^{-(m-1)} t_{i_m} \cdot t_0) s_0, \end{aligned}$$

so that $c(s_{i_1} s_{i_2} \cdots s_{i_m})$, and with it G , if it exists, is completely determined by our hypothesis.

We next prove that the elements $s_i = t_i s_0$ actually constitute an m -group under this operation. As to condition 1 of the definition of an m -group, given $c(s_{i_1} s_{i_2} \cdots s_{i_m}) = s_{i_{m+1}}$ with all s 's but s_{ij} specified members of G , we correspondingly have $t_{i_1} \cdot T^{-1} t_{i_2} \cdots T^{-(m-1)} t_{i_m} \cdot t_0 = t_{i_{m+1}}$, with all elements specified members of G_0 with the exception of $t_{i_{m+1}}$, when $j = m+1$, $T^{-(i-1)} t_{ij}$, when $j \neq m+1$. In the first case, a unique $t_{i_{m+1}}$ in G_0 , and, hence $s_{i_{m+1}}$ in G , are immediately determined. In the second case, a unique $T^{-(i-1)} t_{ij}$ in G_0 is determined, hence again $t_{i_{m+1}}$ in G_0 , and s_{ij} in G . As for condition 2, we have

$$\begin{aligned} c(s_{i_1} \cdots s_{i_{j-1}} c(s_{ij} \cdots s_{i_{j+m-1}}) s_{i_{j+m}} \cdots s_{i_{2m-1}}) \\ &= (t_{i_1} \cdots T^{-(i-2)} t_{i_{j-1}} \cdot T^{-(i-1)} (t_{ij} \cdots T^{-(m-1)} t_{i_{j+m-1}} \cdot t_0) \\ &\quad \cdot T^{-i} t_{i_{j+m}} \cdots T^{-(m-1)} t_{i_{2m-1}} \cdot t_0) s_0 \\ &= (t_{i_1} \cdots T^{-(i-2)} t_{i_{j-1}} \cdot T^{-(i-1)} t_{ij} \cdots T^{-(i+m-2)} t_{i_{j+m-1}} \\ &\quad \cdot T^{-(i+m-1)} t_{i_{j+m}} \cdots T^{-(2m-2)} t_{i_{2m-1}} \cdot t_0^2) s_0, \end{aligned}$$

the last since $T^{-(i-1)} t_0 = t_0$, and $t_0 \cdot t = T^{-(m-1)} t \cdot t_0$, by our hypothesis. The result is thus independent of j , whence follows condition 2.

It remains to be shown that the m -group G thus obtained actually redetermines, via s_0 , the G_0 , t_0 , T of the given hypothesis⁽³⁷⁾. From the operation c

to be confused with our polyadic concept of §5. See A. M. Turing, *The extensions of a group*, Compositio Mathematica, vol. 5 (1938), pp. 357-367.) From this point of view, the abstract containing groups of m -groups with given G_0 are the extensions of G_0 by the cyclic group of order $m-1$. Our theorem on the determination of G could then have been based on the determination of G^* as cyclic extension of G_0 . The theorem on cyclic extensions thus envisaged would be not quite Turing's (Theorem 5, loc. cit.), but equivalent thereto by the identification of our T with his ξ , t_0 with $\zeta^{-1} r^*$.

⁽³⁷⁾ In connection with the preceding footnote it must be mentioned that this part of the proof was overlooked by the writer until the final check-up on the entire paper.

as given, and again with the aid of the relation $T^{-(m-1)}t_{i_m} \cdot t_0 = t_0 \cdot t_{i_m}$, we see that equivalent $(m-1)$ -ads $\{s_{i_1}, s_{i_2}, \dots, s_{i_{m-1}}\}$ are those for which the corresponding elements $t_{i_1} \cdot T^{-1}t_{i_2} \cdots T^{-(m-2)}t_{i_{m-1}} \cdot t_0$ of the given 2-group G_0 are the same. If then we represent the elements of the associated 2-group of G thus by the elements of the given group G_0 , and determine the operation of this associated group via $[\{s_{i_1}, s_{i_2}, \dots, s_{i_{m-1}}\}] \cdot [\{s_{j_1}, s_{j_2}, \dots, s_{j_{m-1}}\}] = [\{c(s_{i_1}s_{i_2} \cdots s_{i_{m-1}}s_{j_1}), s_{j_2}, \dots, s_{j_{m-1}}\}]$, bracket meaning class of $(m-1)$ -ads equivalent to the specified $(m-1)$ -ad, we find this operation, again with the help of the above relation, to be identical with the operation of the given 2-group. That is, abstractly, the given G_0 is the associated ordinary group of G . Since, for the $s_i = s_0$, t_i is the identity, we immediately have $s_0^{m-1} = [\{s_0, s_0, \dots, s_0\}] = t_0$ in the above representation. Finally, by introducing identity, hence inverse, and thus transform, in their original polyadic form, it can likewise be shown that if the elements of G_0 are transformed by s_0 the resulting automorphism of G_0 is T . Therefore, the proof has been completed.

We have already used the converse of the coset theorem in giving an example of a 3-group of order three having no variant element. This 3-group can now be given abstractly in accordance with the above theorem. For G_0 , take the cyclic group $(1, t, t^2)$. Let $s_0^2 = 1$, and let the automorphism T of G_0 be $T(1, t, t^2) = (1, t^2, t)$. Our hypothesis is verified, thus giving us a 3-group (s_0, ts_0, t^2s_0) of order three. We obtain directly $(ts_0)^{-1}s_0(ts_0) = t^2s_0$, $s_0^{-1}ts_0s_0 = t^2s_0$, $s_0^{-1}t^2s_0s_0 = ts_0$, proving that none of the three elements of the 3-group are invariant under the 3-group.

This theorem may be used to determine all finite abstract polyadic groups of given small order. In this connection we have as an immediate consequence of the preceding theorem the following. *A necessary and sufficient condition that two m -groups G' and G'' be simply isomorphic is that a simple isomorphism can be set up between their associated 2-groups G'_0 and G''_0 , and an element s'_0 of G' made to correspond to an element s''_0 of G'' , so that $(s'_0)^{m-1}$ in G'_0 corresponds to $(s''_0)^{m-1}$ in G''_0 , and s'_0 and s''_0 transform G'_0 and G''_0 respectively so that corresponding elements go over into corresponding elements.* We postpone the application of these theorems even to our modest determination of the polyadic groups of the first three orders until our detailed study of cyclic polyadic groups of finite order gives us some basis for comparison of polyadic groups.

However, one result of some theoretical interest emerges immediately. From our general determination theorem, it follows that the number of m -adic groups with g given symbols as elements is no greater than the number of 2-groups on g other given symbols as elements times g times the largest number of automorphisms a 2-group of order g can have. We may therefore conclude that *the number of abstract m -adic groups of given finite order g is a bounded function of m* .

II. FINITE POLYADIC GROUPS

A. *m*-ADIC SUBSTITUTIONS AND SUBSTITUTION GROUPS

9. The symmetric *m*-adic substitution group of degree *n*. An ordinary substitution, finite or infinite, may be considered to be a 1-1 correspondence between the members of a class Γ and the members of the same class. Let now $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ be an ordered sequence of $m-1$ equivalent classes. By an *m*-adic substitution on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ we shall mean a transformation which in 1-1 fashion carries the members of Γ_1 into those of Γ_2 , of Γ_2 into those of Γ_3, \dots , of Γ_{m-1} into those of Γ_1 ⁽³⁸⁾. Symbolically we shall write $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$. Intrinsically, therefore, the Γ 's really enter into an *m*-adic substitution as a cycle, with Γ_1 following Γ_{m-1} . If s_1 and s_2 represent two *m*-adic substitutions on the same sequence of Γ 's, we may as usual refer to $s_1 s_2$, the product of s_1 and s_2 , that is, the transformation equivalent to performing s_1 followed by s_2 . But in general, for $m > 2$, the product of two *m*-adic substitutions will not be an *m*-adic substitution on the given sequence of Γ 's, for it will transform Γ_1 into Γ_3 , instead of Γ_2 . On the other hand, the product of m *m*-adic substitutions on the $m-1$ Γ 's will again transform $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$, and hence we can expect to have *m*-adic groups of *m*-adic substitutions⁽³⁹⁾. We can likewise expect to have *m*-adic groups of μ -adic substitutions provided $\mu-1$ is a divisor of $m-1$. However, by *m*-adic substitution group we shall understand the former, that is, a set of *m*-adic substitutions, all on the same sequence of Γ 's, and forming an *m*-adic group under the product of m substitutions as operation⁽⁴⁰⁾.

When the Γ 's are mutually exclusive, an *m*-adic substitution can be given by an ordinary substitution where the one class Γ is the logical sum of the given Γ 's. On the other hand, when the Γ 's have common elements, an *m*-adic substitution cannot in general be thus considered, since one and the same element may be transformed into different elements according to the Γ_i of which it is considered to be a member. We shall restrict our attention to the former case⁽⁴¹⁾. But our results will be foreshadowed not by considering the resulting

(38) Our language is that of transformation; that is, we shall say "*a* is carried into *b*" where the language of substitution would say "*a* is replaced by *b*".

(39) On the other hand, the product of m *m*-adic substitutions not all on the same sequence of Γ 's will "usually" fail to be an *m*-adic substitution for any sequence of Γ 's. Hence the straight-laced definition following.

(40) The following generalization of ordinary substitution likewise suggests itself in connection with the schar concept. For but two equivalent classes Γ_1, Γ_2 , consider transformations which in 1-1 fashion carry the elements of Γ_1 into those of Γ_2 . If A, B, C are three such transformations, then $AB^{-1}C$ is also such a transformation. Note that here the product of two such transformations does not, in general, even exist.

(41) For simplicity. If each member *a* of Γ_i is replaced by the couple (i, a) , Γ 's not mutually exclusive become mutually exclusive, and it is then readily seen when results obtained for mutually exclusive Γ 's hold for arbitrary Γ 's. Actually, our results were first obtained for arbitrary Γ 's. But that they are so little affected by the overlapping or nonoverlapping of the Γ 's indicates that we have left wholly unexplored the more interesting part of the complete theory.

m -adic substitutions special types of ordinary substitutions, but generalizations of ordinary substitutions, reducing to the latter when $m=2$.

We further restrict our attention to the case where the Γ 's are finite classes, and hence consist each of the same finite number of members n . The analogy with an ordinary substitution will be furthered by saying that the m -adic substitution is then of *degree n*. Let then the members of Γ_1 be symbolized $a_{11}, a_{12}, \dots, a_{1n}$, of $\Gamma_2, a_{21}, a_{22}, \dots, a_{2n}, \dots$ of $\Gamma_{m-1}, a_{(m-1)1}, a_{(m-1)2}, \dots, a_{(m-1)n}$. Corresponding to the primitive mode of writing ordinary substitutions we have the following form for any m -adic substitution on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$:

$$\begin{array}{ccccccc} a_{11} & & a_{12} & & \cdots & a_{1n} \\ a_{2j'_1} & & a_{2j'_2} & & \cdots & a_{2j'_n} \\ & \cdot & & \cdot & & \cdot & \\ a_{(m-1)j^{(m-1)}_1} & & a_{(m-1)j^{(m-1)}_2} & & \cdots & a_{(m-1)j^{(m-1)}_n} \\ a_{1j^{(m)}_1} & & a_{1j^{(m)}_2} & & \cdots & a_{1j^{(m)}_n} \end{array}$$

where the i th row is some permutation of $(a_{11} a_{12} \dots a_{1n})$ except for $i=m$, when it is a permutation of the first row, and each letter is carried into the one immediately below it by the substitution. If, as suggested above, we consider our m -adic substitution an ordinary substitution on all the letters a_{ij} , it can also be written in standard form as a product of cycles on different letters. In that case, each cycle will have a multiple of $m-1$ letters, these letters cyclically running through the $m-1$ Γ 's.

Since an m -adic substitution of degree n is thus determined by $m-1$ independent permutations of n elements each, we thus see that there are $(n!)^{m-1}$ m -adic substitutions of degree n , the sequence of Γ 's being understood given. Observe again that if s_1, s_2, \dots, s_m are m -adic substitutions on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, their products $s_1 s_2 \dots s_m$ is also an m -adic substitution on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$. In detail, s_1 will carry a_{ij} into some $a_{(i+1)j'}$, s_2 will carry $a_{(i+1)j'}$ into some $a_{(i+2)j''}$, \dots , and s_m a resulting $a_{ij^{(m-1)}}$ into $a_{ij^{(m)}}$. Hence $s_1 s_2 \dots s_m$ carries a_{ij} into $a_{ij^{(m)}}$ as required. It then easily follows that the $(n!)^{m-1}$ m -adic substitutions of degree n constitute an m -group under the operation $s_1 s_2 \dots s_m$. While the corresponding result holds good apart from our hypothesis of finite mutually exclusive Γ 's, for the present case it suffices to reinterpret our m -adic substitutions as ordinary substitutions. Condition 2 for an m -group then follows from the associative law for the multiplication of ordinary substitutions. As for condition 1, the case where all s 's but s_{m+1} in $s_1 s_2 \dots s_m = s_{m+1}$ are given m -adic substitutions has been taken care of. And if all but s_i are given m -adic substitutions, $1 \leq i \leq m$, by letting s_i run through the $(n!)^{m-1}$ possible m -adic substitutions, $s_1 s_2 \dots s_m$ must do the same, and hence equals s_{m+1} for one and only one m -adic substitution s_i .

We shall call this m -group of order $(n!)^{m-1}$ the *m -adic symmetric group* of

degree n . It clearly becomes the ordinary symmetric group of degree n when $m=2$. As in the case of ordinary substitution groups, every m -adic substitution group on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, or briefly of degree n , will be a subgroup of the m -adic symmetric group of degree n . It readily follows that the necessary and sufficient condition that a finite set of m -adic substitutions all on the same sequence of Γ 's form an m -adic substitution group is that the product of any m substitutions in the set be in the set.

Of special interest are those m -adic substitutions of degree n in which the last row is an exact repetition of the first row. There are clearly $(n!)^{m-2}$ such substitutions. If s be such a substitution, s^{m-1} clearly carries each letter into itself, and hence $s^m = s$. Conversely, if $s^m = s$, s must be such a substitution. According to a definition already given, s is then of m -adic order one. The unit class with s as sole member therefore itself constitutes an m -adic substitution group of order one. Hence the m -adic symmetric group of degree n has $(n!)^{m-2}$ first order elements, and correspondingly $(n!)^{m-2}$ subgroups of order one. For $m=2$ these become the sole identity of the group.

10. 2^{m-1} -fold classification of m -adic substitutions; the m -adic alternating groups. The classic theory of positive and negative substitutions involves the use of the determinant

$$\Delta = \begin{vmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ \cdot & \cdot & \cdot & \ddots & \cdot \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{vmatrix}$$

which is left invariant under every positive substitution on the letters a_1, a_2, \dots, a_n , and is transformed into its negative under every negative substitution on those letters. We generalize this theory by the same means.

We now form the $m-1$ determinants $\Delta_1, \Delta_2, \dots, \Delta_{m-1}$, where Δ_i is the determinant Δ for the n letters $a_{i1}, a_{i2}, \dots, a_{in}$ of Γ_i , and transform them accordingly to a given m -adic substitution

$$\begin{array}{ccccccccc} a_{11} & & a_{12} & & \cdots & & a_{1n} & & \\ a_{2j'_1} & & a_{2j'_2} & & \cdots & & a_{2j'_n} & & \\ \cdot & & \cdot & & \ddots & & \cdot & & \\ & & & & & & & & \\ a_{(m-1)j_1^{(m-1)}} & & a_{(m-1)j_2^{(m-1)}} & & \cdots & & a_{(m-1)j_n^{(m-1)}} & & \\ a_{1j_1^{(m)}} & & a_{1j_2^{(m)}} & & \cdots & & a_{1j_n^{(m)}} & & \end{array}$$

If in the i th row each letter a_{ij} is rewritten $a_{(i+1)j}$ ⁽⁴²⁾, then the new i th row together with the old $(i+1)$ -st row defines an ordinary substitution on the letters of the $(i+1)$ -st row. The transform of Δ_i under the m -adic substi-

(42) a_{1j} , when $i=m-1$. Likewise, below, Δ_{i+1} is Δ_1 when $i=m-1$.

tion is clearly the transform of Δ_{i+1} under this ordinary substitution, and hence is Δ_{i+1} , or its negative, according as this ordinary substitution is positive or negative. We therefore have under the m -adic substitution

$$\Delta_1 \rightarrow \delta_1 \Delta_2, \Delta_2 \rightarrow \delta_2 \Delta_3, \dots, \Delta_{m-1} \rightarrow \delta_{m-1} \Delta_1, \quad \delta_1, \delta_2, \dots, \delta_{m-1} = \pm 1.$$

With each m -adic substitution there is thus associated a sequence of $m-1$ numbers $[\delta_1, \delta_2, \dots, \delta_{m-1}]$ whose values are $+1$ or -1 . Clearly, when $n > 1$, an m -adic substitution of degree n can be written down for every possible assignment of values to the δ 's. The m -adic substitutions of degree n , $n > 1$, thus fall into 2^{m-1} mutually exclusive classes corresponding to the 2^{m-1} possible δ -sequences $[\delta_1, \delta_2, \dots, \delta_{m-1}]$.

Given m m -adic substitutions s_i , with the corresponding δ -sequences $[\delta_{i1}, \delta_{i2}, \dots, \delta_{i(m-1)}]$, the m -adic substitution $s_1 s_2 \dots s_m$ has a δ -sequence $[\delta_1, \delta_2, \dots, \delta_{m-1}]$ which depends only on the δ -sequences of the s_i 's. In fact, by following through the effect of the succession of substitutions s_1, s_2, \dots, s_m on the determinants $\Delta_1, \Delta_2, \dots, \Delta_{m-1}$, we obtain the following equations for determining $[\delta_1, \delta_2, \dots, \delta_{m-1}]$:

$$\begin{aligned}\delta_1 &= \delta_{11} \delta_{22} \dots \delta_{(m-1)(m-1)} \delta_{m1}, \\ \delta_2 &= \delta_{12} \delta_{23} \dots \delta_{(m-1)1} \delta_{m2}, \\ &\dots \dots \dots \dots \dots \dots, \\ \delta_{m-1} &= \delta_{1(m-1)} \delta_{21} \dots \delta_{(m-1)(m-2)} \delta_{m(m-1)}.\end{aligned}$$

Now let K be the class of the 2^{m-1} possible δ -sequences. If $\sigma_1, \sigma_2, \dots, \sigma_m$ are any m such δ -sequences, and σ is the δ -sequence obtained from $\sigma_1, \sigma_2, \dots, \sigma_m$ in accordance with the above equations, an m -adic operation $k(\sigma_1 \sigma_2 \dots \sigma_m)$ is determined such that $\sigma = k(\sigma_1 \sigma_2 \dots \sigma_m)$. It is then readily shown that K constitutes an m -group under k . In fact, condition 1 for an m -group is immediately verified by referring to the above equations. And condition 2 follows from the associative law for m -adic substitutions, and the fact that if s_1, s_2, \dots, s_m are substitutions corresponding to $\sigma_1, \sigma_2, \dots, \sigma_m$ respectively, $s_1 s_2 \dots s_m$ corresponds to $k(\sigma_1 \sigma_2 \dots \sigma_m)$. We shall call this m -group of order 2^{m-1} the *complete m-adic δ-group*⁽⁴³⁾.

Consider now any m -adic substitution group of degree n and form the class K' of δ -sequences corresponding to its members. Since the product of any m substitutions of the group is in the group, the k product of any m δ -sequences in K' will be in K' . As K' is a subclass of the class of members of the complete m -adic δ -group, and the latter is finite, this suffices to prove

(43) Actually, then, we have established a homomorphism between the symmetric m -adic substitution group of degree n , $n > 1$, and this complete m -adic δ -group. The rest of this section could then largely have been given as a consequence of our general results on homomorphisms between m -adic groups, as could indeed the very fact that K is an m -group under k . In the generalization of this section occurring in the last section of our paper full use will be made of the concept of homomorphism.

the following. *The δ -sequences corresponding to the members of any m -adic substitution group of degree n form the complete m -adic δ -group, or a subgroup thereof.*

By means of the above equations for the k operation we readily prove, as for ordinary substitutions, that *every m -adic substitution group of degree n has the same number of substitutions for each δ -sequence in the corresponding "δ-subgroup"*⁽⁴⁴⁾. In fact, let s_m and s_{m+1} be any two substitutions in the group corresponding to any two given δ -sequences σ_m and σ_{m+1} of the corresponding δ -subgroup, and choose s_1, \dots, s_{m-1} so that $s_1 \cdots s_{m-1} s_m = s_{m+1}$. If now we let s_m run through all the substitutions in the group corresponding to σ_m , s_{m+1} assumes an equal number of values in the group all corresponding to σ_{m+1} . Hence there are at least as many substitutions in the group corresponding to one δ -sequence as to another, and consequently, by reciprocal reasoning, the same number. Since the order of the complete m -adic δ -group is 2^{m-1} , that of a subgroup thereof must be of the form 2^{μ} ⁽⁴⁵⁾. From the above result it follows that the order of an m -adic substitution group is a multiple of the order of its δ -subgroup. We therefore have as a corollary of the above result *every m -adic substitution group of odd order has a δ-subgroup of order one*, that is, all of its substitutions correspond to one and the same δ -sequence.

Applied to the symmetric group itself, the above result shows that the 2^{m-1} mutually exclusive classes into which the m -adic symmetric group of degree n is divided all have the same number of members. Now given any subgroup of the complete m -adic δ -group, form the class C' of all the m -adic substitutions of degree n corresponding to each δ -sequence in the given δ -subgroup. The product of any m substitutions in C' will therefore be in C' . Hence the members of C' form a subgroup of the symmetric group. By analogy with ordinary groups we shall call it an *m -adic alternating group*. Consequently, *there are as many m -adic alternating groups of degree n , $n > 1$, as there are subgroups of the complete m -adic δ -group, each alternating group consisting of all the substitutions of the symmetric group with δ -sequences in the corresponding δ-subgroup.* We may now further state that there is a one-many correspondence between the m -adic δ -subgroups and m -adic substitution groups of degree n , $n > 1$, that is, between the class consisting of the complete m -adic δ -group and its subgroups, and the m -adic symmetric group of degree n and its subgroups; and this correspondence is preserved under the relation "group or subgroup of."

For $m=2$ the complete δ -group is the cyclic group of order 2, and its sole subgroup, the identity, corresponds to the sole ordinary alternating group of degree n ⁽⁴⁶⁾. For $m=3$ the complete δ -group is of order 4. By direct calcula-

⁽⁴⁴⁾ We shall use the phrase δ -subgroup to cover the complete δ -group as well.

⁽⁴⁵⁾ By Lagrange's theorem for polyadic groups—proved in §4.

⁽⁴⁶⁾ van der Waerden has already noted the homomorphism between any substitution group having at least one odd substitution and this cyclic group of order two.

tion we find it to possess exactly four subgroups, that is, with classes of elements $([+1, +1]), ([-1, -1]), ([+1, +1], [-1, -1]), ([+1, -1], [-1, +1])$. Hence, there are exactly four triadic alternating groups of degree $n, n > 1$.

Thanks to B. P. Gill, we are able to determine the m -adic alternating groups of degree n for arbitrary m . For this purpose it is essential to obtain a suitable representation of the associated ordinary group of the complete m -adic δ -group. The ideas leading up to this are of more general application, and hence at least part of the following digression.

11. Associated and containing ordinary groups; commutative m -adic substitutions. The substitutions of an m -adic substitution group G of degree n , considered as ordinary substitutions on $(m-1)n$ letters, generate an ordinary substitution group which satisfies our definition of a containing group of G . With G thus an m -adic group of m -adic substitutions, this containing group will be of index $m-1$, and hence simply isomorphic with the abstract containing group G^* of G . We shall therefore use it throughout to represent G^* , and for simplicity symbolize it G^* . We may likewise refer to the associated group of G with respect to this containing group as G_0 .

In the terminology of §6, the i th coset of G^* consists of the products of i elements of G . To avoid duplication, it will be convenient henceforth to assume that $1 \leq i \leq m-1$. Since each substitution in G transforms $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$, it follows that the i th coset of G^* consists of transformations which in 1-1 fashion carry the members of each Γ_j into those of $\Gamma_{j+i}, j+i$ reduced modulo $m-1$ if need be. We may therefore call these substitutions of G^* the i -ads of G . In particular, G_0 , which consists of the $(m-1)$ -ads of G in G^* , consists of transformations which transform each Γ_j into itself. Each $(m-1)$ -ad of G thus appears in G^* as the product of $m-1$ ordinary substitutions, each of these ordinary substitutions being on the letters of a single Γ . We have incidentally verified that G^* is of index $m-1$.

Considered as ordinary substitution groups on $(m-1)n$ letters we see that for $m > 2$, G^* is imprimitive with systems of imprimitivity $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, while G_0 is intransitive with the letters in each Γ carried into letters of the same Γ only, by every substitution of G . If then for each Γ we separate from each substitution in G_0 the substitution involving only the letters of that Γ , there results an ordinary substitution group on the letters of that Γ . We shall symbolize these $m-1$ groups on the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ by $G'_1, G''_1, \dots, G^{(m-1)}_1$ respectively, and call them the *associated constituent groups* of the m -adic substitution group G . It is then significant that the *associated constituent groups of an m -adic substitution group are conjugate ordinary groups*. In fact, recall that G_0 is an invariant subgroup of G^* , and hence is invariant under every m -adic substitution s in G . Now s carries the letters of each Γ_i into those of Γ_{i+1} . Hence, when the substitutions of G_0 are transformed by s , the components of these substitutions on the letters of Γ_i be-

come the components of the same class of substitutions on the letters of Γ_{i+1} . We thus have specifically

$$s^{-1}G'_0s = G''_0, \quad s^{-1}G''_0s = G'''_0, \quad \dots, \quad s^{-1}G^{(m-1)}_0s = G'_0,$$

for every s in G .

If s_1 and s_2 are m -adic substitutions on the same sequence of Γ 's, we may consider them as elements of the corresponding m -adic symmetric group. The transform of s_2 under s_1 is then $s_1^{-1}s_2s_1$ in the notation of the containing group of the symmetric group, and hence may be obtained by the ordinary rule for transforming substitutions. Restated for our primitive mode of representing m -adic substitutions, this rule becomes the following. Replace each letter in s_2 by the letter immediately under it in s_1 and rewrite in standard form. Thus, to illustrate, let

$$\begin{array}{cccc} a_{11}a_{12}a_{13} & a_{11}a_{12}a_{13} & a_{22}a_{23}a_{21} & a_{11}a_{12}a_{13} \\ s_2 = a_{22}a_{21}a_{23}, & s_1 = a_{22}a_{23}a_{21}; & s_1^{-1}s_2s_1 = a_{13}a_{12}a_{11} = a_{23}a_{21}a_{22}. \\ a_{11}a_{13}a_{12} & a_{13}a_{11}a_{12} & a_{22}a_{21}a_{23} & a_{12}a_{11}a_{13} \end{array}$$

Actually, the result before it is rewritten defines the transform equally well; for, as stated before, it is really the cycle, rather than the sequence, of Γ 's that is significant.

If s_2 is invariant under s_1 , then s_1 and s_2 are commutative; and conversely. The problem of determining all m -adic substitutions s , commutative with a given m -adic substitution s_1 of degree n , and on the same Γ 's, is best treated by writing the substitutions in ordinary cycle form. We recall that the number of letters in each cycle is then a multiple of $m-1$. If s_1 consists of a single cycle, the ordinary substitutions r , on the $(m-1)n$ letters of s_1 , which are commutative with s_1 , are the $(m-1)n$ ordinary powers of s_1 . Of these exactly n , i.e., those of the form $s_1^{k(m-1)+1}$, are m -adic substitutions on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$. We shall later call these the m -adic powers of s_1 , i.e., the elements of the m -adic group generated by s_1 . Hence, the only m -adic substitutions on the Γ 's of s_1 , commutative with the single cycle m -adic substitution s_1 of degree n , are the n m -adic powers of s_1 . We now have no difficulty in paraphrasing the corresponding argument for ordinary substitutions, and obtain the following results. If s_1 consists of λ cycles with numbers of letters $(m-1)n_1, (m-1)n_2, \dots, (m-1)n_\lambda$ no two of which are equal, the m -adic substitutions s , on the Γ 's of s_1 , commutative with s_1 , are the $n_1n_2 \dots n_\lambda$ products of the m -adic powers of the several cycles. And, if s_1 consists of k equal cycles of $(m-1)\nu$ letters each, the m -adic substitutions s on the Γ 's of s_1 commutative with s_1 are $\nu^k k!$ in number, there being ν^k such m -adic substitutions for each of the $k!$ possible permutations of the k cycles. Clearly in any case, the m -adic substitutions s , commutative with an m -adic substitution s_1 , and on the Γ 's of s_1 , constitute an m -adic substitution group.

12. Further study of the complete m -adic δ -group and m -adic alternating groups. The ideas of the preceding section enable us to clear up a certain difficulty in our presentation of the 2^{m-1} -fold classification of m -adic substitutions and in its consequences. Observe that whereas the Γ_i 's are mere classes, the determinants Δ_i assume the letters in each Γ_i arranged in a sequence. The δ -sequence associated with a given m -adic substitution s will therefore in general depend not only on s but also on the original ordering of the letters in the Γ_i 's. However, we shall see that the same 2^{m-1} classes are obtained no matter what ordering is assumed, only their description by δ -sequences being thus affected.

Actually, this ordering is equivalent to a first order m -adic substitution s_0 which carries each a_{ii} into $a_{(i+1)i}$, $i+1$ being replaced by 1 when $i=m-1$. Let us then write the m -adic symmetric group in coset form with arbitrary element $s=ts_0$. t is then in the form $t't''\cdots t^{(m-1)}$ where $t^{(i)}$ is an ordinary substitution on the letters of Γ_i . If now we associate with t the ϵ -sequence $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$, where ϵ_i is +1 or -1 according as $t^{(i)}$ is a positive or negative substitution, we see from the effect on the determinants Δ_i that the ϵ -sequence of t is identical with the δ -sequence of s . Let then $s_1=t_1s_0$ and $s_2=t_2s_0$ have the same δ -sequence, and hence t_1 and t_2 the same ϵ -sequence. Then $s_1s_2^{-1}=t_1t_2^{-1}$ will have an ϵ -sequence (+1, +1, ..., +1), i.e., will be the product of positive substitutions only. Conversely, if $s_1s_2^{-1}$ is the product of positive substitutions only, the corresponding t_1 and t_2 must have the same ϵ -sequence, and s_1 and s_2 the same δ -sequence. Hence, s_1 and s_2 belong to the same one of the 2^{m-1} classes of m -adic substitutions when and only when $s_1s_2^{-1}$ is the product of positive substitutions on the letters of the several Γ 's. As this criterion is independent of s_0 , the intrinsic character of our classification has been demonstrated.

The ϵ -sequences may be used to obtain a concrete representation of the associated ordinary group of the complete m -adic δ -group. More generally, consider the containing group of the m -adic symmetric group of degree n , $n > 1$. Since each i -ad R thereof is the product of i m -adic substitutions, R will transform the Δ 's according to some scheme

$$\Delta_1 \rightarrow \eta_1 \Delta_{i+1}, \Delta_2 \rightarrow \eta_2 \Delta_{i+2}, \dots, \Delta_{m-1} \rightarrow \eta_{m-1} \Delta_i, \quad \eta_1, \eta_2, \dots, \eta_{m-1} = \pm 1.$$

With R we may thus associate the η -sequence (with subscript) $\{\eta_1, \eta_2, \dots, \eta_{m-1}\}_i$. If R_1 thus corresponds to $\{\eta'_1, \eta'_2, \dots, \eta'_{m-1}\}_{i_1}$, R_2 to $\{\eta''_1, \eta''_2, \dots, \eta''_{m-1}\}_{i_2}$, R_1R_2 will correspond to $\{\eta'_1 \eta''_{i_1+1}, \eta'_2 \eta''_{i_2+2}, \dots, \eta'_{m-1} \eta''_{i_1}\}_{i_1+i_2}$, subscripts being reduced modulo $m-1$ if need be. It follows that the containing group of the m -adic symmetric group is homomorphic to the resulting complete η -group (with subscript). Now with $i=1$, the η -sequence is nothing more than the δ -sequence of the corresponding m -adic substitution. From the way in which our operations were obtained it follows that the complete η -group may be considered a containing group, of index $m-1$,

indeed, of the complete m -adic δ -group. The associated group of the complete m -adic δ -group will then be composed of the η -sequences whose subscript is $m-1$. But going back to the Δ 's we see that these η -sequences are then actually the ϵ -sequences of the corresponding $(m-1)$ -ads of m -adic substitutions. Under this representation, therefore, the operation of the associated group of the complete m -adic δ -group, i.e., of the *complete ϵ -group* as we shall call it, becomes

$$(\epsilon'_1, \epsilon'_2, \dots, \epsilon'_{m-1})(\epsilon''_1, \epsilon''_2, \dots, \epsilon''_{m-1}) = (\epsilon'_1\epsilon''_1, \epsilon'_2\epsilon''_2, \dots, \epsilon'_{m-1}\epsilon''_{m-1})^{(47)}.$$

We therefore see that the complete ϵ -group is an ordinary abelian group of order 2^{m-1} . Since each ϵ is ± 1 , its elements other than the identity are all of order two, so that it is indeed of type $(1, 1, \dots, 1)$.

The complete m -adic δ -group is therefore semi-abelian. As it is readily seen to be non-abelian whenever $m > 2$, it follows that it then has no invariant element. More specifically, the transform of $[\delta_1, \delta_2, \delta_3, \dots, \delta_{m-1}]$ by $[\delta'_1, \delta'_2, \delta'_3, \dots, \delta'_{m-1}]$ is easily found, via the complete η -group, to be

$$[\delta'_{m-1}\delta_{m-1}\delta'_1, \delta'_1\delta_1\delta'_2, \delta'_2\delta_2\delta'_3, \dots, \delta'_{m-2}\delta_{m-2}\delta'_{m-1}].$$

The condition for invariance is then easily rewritten $\delta_1\delta'_1 = \delta_2\delta'_2 = \delta_3\delta'_3 = \dots = \delta_{m-1}\delta'_{m-1}$. It follows that there are exactly two δ -sequences leaving any given δ -sequence $[\delta_1, \delta_2, \dots, \delta_{m-1}]$ invariant, namely, $[\delta_1, \delta_2, \dots, \delta_{m-1}]$ and $[-\delta_1, -\delta_2, \dots, -\delta_{m-1}]$.

The present and succeeding paragraph presuppose a partial reading of the later §21 and §22. We have observed that except for the identity the elements of the complete ϵ -group are all of order two. While it follows therefrom that the elements of the complete m -adic δ -group are of no other m -adic orders than one or two, we find directly that exactly half of them are of order one, half of order two. Thus, if σ is the δ -sequence $[\delta_1, \delta_2, \dots, \delta_{m-1}]$, and $\delta_0 = \delta_1\delta_2 \dots \delta_{m-1}$, then, with k as in §10, we find $k(\sigma\sigma \dots \sigma) = [\delta_0\delta_1, \delta_0\delta_2, \dots, \delta_0\delta_{m-1}]$, $k(\sigma\sigma \dots k(\sigma\sigma \dots \sigma)) = [\delta_1, \delta_2, \dots, \delta_{m-1}]$. Hence, the m -adic order of a δ -sequence is one or two according as the product of its δ 's is $+1$ or -1 .

The cyclic subgroups of the complete m -adic δ -group are therefore of orders one or two, there being 2^{m-2} first order subgroups, and, for $m > 2$, 2^{m-2} or 2^{m-3} cyclic second order subgroups according as m is even or odd. Our result on the δ -sequences leaving a given δ -sequence invariant, coupled with the easily verified fact that an m -group of order two must be abelian, leads to the result that the complete m -adic δ -group has exactly 2^{m-2} second

⁽⁴⁷⁾ Actually, by a slight change in point of view, the transformation of the Δ 's resulting from an m -adic substitution can be considered an m -adic linear transformation in one variable in the sense of our later §35. The present and several other formulas, derived independently in the present section, would then become special cases of the formulas of §35.

order subgroups for $m > 2$. Hence, when m is odd, half of them are non-cyclic⁽⁴⁸⁾.

We turn now to the determination of all the subgroups of the complete m -adic δ -group, and consequently, the determination of all m -adic alternating groups. Since the complete m -adic δ -group is semi-abelian, all of its elements transform a given ϵ -sequence $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$ into the same ϵ -sequence. As before, we can employ the operation of the complete η -group, and thus find the unique transform of $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$ under every δ -sequence to be $(\epsilon_{m-1}, \epsilon_1, \epsilon_2, \dots, \epsilon_{m-2})$. Now if H is a subgroup of the complete m -adic δ -group, its associated ordinary group H_0 must be a subgroup of the complete ϵ -group invariant under H . Hence H_0 can only be such a subgroup of the complete ϵ -group that if $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-2}, \epsilon_{m-1})$ is in the subgroup, $(\epsilon_{m-1}, \epsilon_1, \epsilon_2, \dots, \epsilon_{m-2})$ also is in the subgroup. The determination of these "admissible" subgroups of the complete ϵ -group is the only difficult part of our problem. It was carried through independently by Gill; but he later found that his solution followed essentially the lines of the general theory of the "Verallgemeinerte Abelsche Gruppen," abbreviated V.A.G., as given by Otto Haupt in the second volume of his *Algebra*⁽⁴⁹⁾.

Following Gill we replace the two values $+1, -1$ by $0, 1$ respectively. If an ϵ -sequence be thus rewritten, the dyadic operation of our complete ϵ -group is best written in additive form, and we have

$$(\epsilon_{11}, \epsilon_{12}, \dots, \epsilon_{1(m-1)}) + (\epsilon_{21}, \epsilon_{22}, \dots, \epsilon_{2(m-1)}) \\ = (\epsilon_{11} + \epsilon_{21}, \epsilon_{12} + \epsilon_{22}, \dots, \epsilon_{1(m-1)} + \epsilon_{2(m-1)}),$$

where addition within the parentheses is modulo 2. Now let $\phi(x)$ be any polynomial in x with coefficients 0 or 1. With a any ϵ -sequence, a unique ϵ -sequence $\phi(x) \cdot a$ is determined as follows. If a is the ϵ -sequence $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-2}, \epsilon_{m-1})$, let $x \cdot a$ be the ϵ -sequence $(\epsilon_{m-1}, \epsilon_1, \epsilon_2, \dots, \epsilon_{m-2})$. With $1 \cdot a = a$, and $x^n \cdot a$ defined inductively through $x^n \cdot a = x \cdot (x^{n-1} \cdot a)$, we can define $\phi(x) \cdot a$ as the sum of the ϵ -sequences obtained by operating on a by the several terms of $\phi(x)$. We now observe two things. First, every ϵ -sequence can be written $\phi(x) \cdot (1, 0, \dots, 0)$. In fact, to obtain $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$, we need merely let $\phi(x) = \epsilon_1 + \epsilon_2 x + \dots + \epsilon_{m-1} x^{m-2}$. Secondly, with $(0, 0, \dots, 0)$ abbreviated 0, we see that $(1, 0, \dots, 0)$ satisfies the equation $(x^{m-1} + 1)(1, 0, \dots, 0) = 0$, but fails to satisfy any equation $\phi(x) \cdot (1, 0, \dots, 0) = 0$ with $\phi(x)$ of degree less than $m-1$, and not identically zero. For we have directly that $x^{m-1} \cdot (1, 0, \dots, 0) = (1, 0, \dots, 0)$; while with $\phi(x)$ of degree less than $m-1$ our previous expression for $\phi(x) \cdot (1, 0, \dots, 0)$ applies. Note finally that 0 and 1 constitute a field K under addition modulo 2, and multiplication. The entire theory of V.A.G.'s in general, and Theorem 3 of Haupt

(48) See §23 for the consequent structure of these second order subgroups.

(49) Otto Haupt, *Einführung in die Algebra*, Leipzig, 1929, vol. 2, pp. 617-621. The result we need is the Theorem 3 of page 620.

in particular, can then be shown to be applicable, and yield the following result.

The admissible subgroups of the complete m -adic ϵ -group are in 1-1 correspondence with the polynomial divisors, other than unity, of $x^{m-1}+1$ relative to the field of coefficients K . If $\tau(x)$ be such a divisor, and $a=\tau(x) \cdot (1, 0, \dots, 0)$, then the corresponding subgroup consists of all distinct ϵ -sequences $\phi(x) \cdot a$.

Actually, if μ is the degree of $(x^{m-1}+1)/\tau(x)$, then $\phi(x)$ can be restricted to degrees less than μ , different $\phi(x)$'s then also giving different ϵ -sequences. It follows that the order⁽⁵⁰⁾ of the corresponding subgroup is 2^μ . The subgroup corresponding to $\tau(x)$ can also be described as consisting of all ϵ -sequences b such that $(x^{m-1}+1)/\tau(x) \cdot b = 0$. It follows that these subgroups satisfy the same properties with respect to the relation of inclusion as do the subgroups of an ordinary cyclic group, $(x^{m-1}+1)/\tau(x)$ taking the place of the order of the subgroup. Note that the unique factorization theorem applies to polynomials with coefficients in a given field. If then $x^{m-1}+1$ is thus completely factored, the distinct divisors $\tau(x)$ can immediately be written down. Since $x^{m-1}+1=(x+1)(x^{m-2}+\dots+x+1)$ relative to K , $x+1$ is always one of the prime divisors of $x^{m-1}+1$. It can readily be shown that it is the only distinct prime divisor of $x^{m-1}+1$, that is, that $x^{m-1}+1=(x+1)^{m-1}$ relative to K , when and only when $m-1$ is itself a power of 2. The different $\tau(x)$'s are then $(x+1), (x+1)^2, \dots, (x+1)^{m-1}$, and each corresponding subgroup contains the next.

Having determined the admissible subgroups of the complete ϵ -group in accordance with the above theorem, it is a simple matter to find the subgroups of the complete δ -group. We return here to our original notation. Each δ -subgroup H , if written in coset form, will be given by $H=H_0\sigma$, with H_0 an admissible ϵ -subgroup, σ a δ -sequence. Hence, if $H_0\sigma$ is known to be a δ -subgroup, its elements can immediately be found from H_0 and σ by the relation

$$(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1}) [\delta_1, \delta_2, \dots, \delta_{m-1}] = [\epsilon_1\delta_1, \epsilon_2\delta_2, \dots, \epsilon_{m-1}\delta_{m-1}],$$

a mere specialization of the dyadic operation of the complete η -group.

Since every admissible ϵ -subgroup H_0 is invariant under every δ -sequence σ , it follows from an early theorem of §4 that $H_0\sigma$ will be a δ -subgroup for every first order σ , and for those second order σ 's for which σ^{m-1} is in H_0 . The distinct δ -subgroups thus arising will then be all the δ -subgroups for a given H_0 . Now when m is even, every δ -subgroup must have at least one first order element. Hence in this case, the δ -subgroups corresponding to H_0 will be all the distinct $H_0\sigma$'s with σ a first order element. Now it is readily proved that if $\tau=(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$, the order of $\tau\sigma$ is the same as that of σ , or opposite, according as $\epsilon_0=\epsilon_1\epsilon_2\dots\epsilon_{m-1}$ is +1 or -1, and furthermore, that the elements of H_0 either all have ϵ_0 equal to +1, or exactly half have $\epsilon_0=+1$,

(50) In the ordinary sense, not that of V.A.G.'s.

half -1 . Hence, if H_0 is of order 2^μ , $H_0\sigma$, with σ of first order, has 2^μ or $2^{\mu-1}$ first order elements according as the elements of H_0 have or have not ϵ_0 's all $+1$. Since the distinct $H_0\sigma$'s with given H_0 are mutually exclusive, while each of the 2^{m-2} first order elements of the complete m -adic δ -group is in some $H_0\sigma$, it follows that when m is even, for each admissible ϵ -subgroup H_0 of order 2^μ there are exactly $2^{m-\mu-2}$ or $2^{m-\mu-1}$ corresponding δ -subgroups according as the ϵ_0 's of the elements of H_0 are, or are not, all $+1$.

For m odd, and given H_0 , we also have these subgroups. But now there may be additional subgroups $H_0\sigma$ with all elements of order two. Now if σ is of second order, the ϵ -sequence of σ^{m-1} is readily seen to be $(-1, -1, \dots, -1)$. It follows that these additional subgroups can arise only when H_0 has the element $(-1, -1, \dots, -1)$, while the ϵ_0 's of all its elements are $+1$. But then each of the 2^{m-2} second order δ -sequences will be in one of these additional subgroups. For such an H_0 , therefore, in addition to the now $2^{m-\mu-2}$ δ -subgroups consisting wholly of first order elements, there will be $2^{m-\mu-2}$ additional δ -subgroups each, indeed, consisting wholly of second order elements.

Actually, the number of δ -subgroups with given associated ordinary group H_0 can be determined without explicitly writing out the elements of H_0 , but merely by an inspection of the corresponding $\tau(x)$. Thus, we have already seen that the order of H_0 is 2^μ , where μ is the degree of $(x^{m-1}+1)/\tau(x)$. By means of the second description given for the subgroup H_0 , it can further be shown that $(-1, -1, \dots, -1)$ is in H_0 when and only when $(x^{m-1}+1)/\tau(x)$ has $x+1$ for divisor; while from the first description it can be shown that the ϵ_0 's of H_0 are all $+1$ when and only when $\tau(x)$ has $x+1$ for divisor. This covers all we need to know about H_0 .

In particular, for $m > 3$, we always have the three distinct divisors of $x^{m-1}+1$ equal to $x^{m-1}+1$, $x^{m-2}+\dots+x+1$, $x+1$. In the first case H_0 is of order one, and consists of but $(+1, +1, \dots, +1)$, the identity. The corresponding δ -subgroups are the first order δ -subgroups listed above. In the second case H_0 is of order two, and consists of $(+1, +1, \dots, +1)$ and $(-1, -1, \dots, -1)$. It is obviously the only admissible second order ϵ -subgroup, and hence the corresponding δ -subgroups are all of the second order δ -subgroups as first listed. The third subgroup, of order 2^{m-2} , is again the only admissible ϵ -subgroup of that order, and consists of all ϵ -sequences with ϵ_0 equal to $+1$. Our general solution then shows that as a result there is but one δ -subgroup of order 2^{m-2} for m even, two for m odd.

Actually, the equations of §10 for the m -adic operation on δ -sequences directly show that we always have the subgroup of order 2^{m-2} consisting of all δ -sequences with $\delta_0 = +1$, and for m odd also the subgroup of order 2^{m-2} consisting of all δ -sequences with $\delta_0 = -1$. Since the complete m -adic δ -group is semi-abelian, all of its subgroups are semi-invariant. It is then of interest to note that the above one, or two, subgroups of order 2^{m-2} are its only invariant subgroups. In fact, our formula for the transform of one δ -sequence by an-

other shows that δ_0 is always thus left invariant; and it also shows that a δ -sequence can always be found which transforms a given δ -sequence into any other with the same δ_0 .

These results are immediately applicable to the corresponding alternating groups, assuming $n > 1$. There are thus always 2^{m-2} alternating groups with substitutions forming one of the 2^{m-1} classes of §10 and, for $m > 2$, 2^{m-2} alternating groups with substitutions forming two such classes. Passing by the general solution, we note that the conditions $\delta_0 = +1$, and $\delta_0 = -1$, correspond to an m -adic substitution considered as an ordinary substitution being positive, or negative. Hence, the only alternating groups invariant under the symmetric group are the alternating group of all positive substitutions, and, for m odd, also the alternating group of all negative substitutions. On the other hand, every alternating group is a semi-invariant subgroup of the symmetric group.

The last observation restricts the possible simplicity of m -adic alternating groups. Regarding the nonexistence of a quotient group of lower order than itself as the distinguishing mark of an ordinary simple group, we are led to define a *simple* m -group as one whose associated group has no subgroup other than the identity invariant under the m -group. It follows that for $n > 2$ only alternating groups corresponding to first order δ -subgroups can be simple. For in any other case, $(+1, +1, \dots, +1)$, the identity of the associated ϵ -subgroup, is a subgroup thereof invariant under the δ -subgroup. Hence the elements of the associated group of the alternating group with ϵ -sequence $(+1, +1, \dots, +1)$ then constitute a subgroup of the associated group invariant under the alternating group. We now proceed to show, on the strength of the corresponding result for ordinary groups, that when $n > 4$ every alternating group H corresponding to a first order δ -subgroup is a simple. Since the associated ϵ -subgroup has but the sole ϵ -sequence $(+1, +1, \dots, +1)$, the associated ordinary group H_0 of the alternating group H consists of all elements $t = t't'' \dots t^{(m-1)}$ where $t^{(i)}$ is any positive substitution on the letters of Γ_i ; and is thus the direct product of the ordinary alternating groups A_1, A_2, \dots, A_{m-1} on the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ respectively. Let there K_0 be any subgroup of H_0 invariant under H . If there could be more than one t in K_0 with the same components $t', t'', \dots, t^{(m-2)}$, then there would be more than one t in K_0 with each component $t', t'', \dots, t^{(m-2)}$ the identity. Now these t 's must constitute a subgroup of K_0 , and this subgroup will be invariant under H_0 as a consequence of the invariance of K_0 under H_0 . The corresponding $t^{(m-1)}$'s must then constitute an invariant subgroup of A_{m-1} , if not A_m itself. Under the present supposition the last would be true; for with $n > 4$, the alternating group A_{m-1} is simple. But then K_0 would coincide with H_0 , instead of being a subgroup of H_0 . For, K_0 being invariant under any s in H , if we transform the above elements of K_0 by $s, s^2, \dots, s^{(m-2)}$, we would have in K_0 every element of H_0 any $m-2$ of whose components are the identity; and th

products of these elements constitute H_0 . We have therefore proved that an element $t = t't'' \dots t^{(m-1)}$ of K_0 is uniquely determined by its first $m-2$ components. If then we transform t by any element of H_0 , the first $m-2$ of whose components are the identity, the first $m-2$ components of t , and hence t itself, will be unchanged. $t^{(m-1)}$ is then always an invariant element of A_{m-1} , and, again with $n > 4$, can only be the identity. The same argument would show each component of an element of K_0 to be the identity, so that K_0 is the identity.

We have therefore proved that for $n > 4$ the 2^{m-2} m -adic alternating groups of degree n corresponding to first order δ -subgroups are simple, the others not. For $n=4$ no m -adic alternating group is simple, since we can let K_0 be the direct product of the axial groups on the letters of the several Γ 's. Again, for $n=3$, no m -adic alternating group is simple for any $m > 2$. The preceding argument breaks down at the one point where the invariance of $t^{(m-1)}$ under A_{m-1} is used to prove $t^{(m-1)}$ the identity. K_0 may now be the third order group obtained from the simple isomorphism between A_1, A_2, \dots, A_{m-1} that results when A_i is transformed into A_{i+1} by a fixed element s of H . Finally, when $n=2$, the very first step of our argument breaks down. The m -adic alternating groups can now be identified with the δ -subgroups themselves. The simple δ -subgroups are those whose associated ϵ -subgroups have no admissible ϵ -subgroup for subgroup other than the identity. Hence, in terms of the above general determination of admissible ϵ -subgroups, the simple δ -subgroups are those whose associated ϵ -subgroups have $(x^{m-1} + 1)/\tau(x)$ prime.

13. Transitive m -adic substitution groups. Since an m -adic substitution group G can carry the letters of Γ_i only into those of Γ_{i+1} , we are led to define a transitive m -adic group G as one whose substitutions will carry each letter of each Γ into every letter of the succeeding Γ . Clearly, the m -adic symmetric group of arbitrary degree n , and the m -adic alternating groups of degree $n > 2$ are then transitive. Our analysis in the next section shows that G will be transitive if the above condition is true for any one Γ , and indeed for any one letter of a Γ , i.e., if the substitutions of G carry one letter of one Γ into every letter of the succeeding Γ , the same is true of every letter of every Γ , and G is transitive.

It is readily proved that the containing 2-group G^* of a transitive m -group G is transitive. In fact, let a_{ij} and $a_{(i+k)j'}$ be any two letters of the Γ 's. Considering $i+k$ reduced modulo $m-1$, we may assume $1 \leq k \leq m-1$. We need not consider $k=1$. For $k > 1$ let r be any $(k-1)$ -ad. It will carry a_{ij} into some $a_{(i+k-1)j''}$. Some s will carry $a_{(i+k-1)j''}$ into $a_{(i+k)j'}$. Hence the k -ad rs , which is a substitution in G^* , carries a_{ij} into $a_{(i+k)j'}$ as required. Conversely, if G^* is transitive, a substitution in G^* carrying a_{ij} into $a_{(i+k)j'}$ belongs to G , and hence G is transitive.

In terms of G_0 , the associated 2-group of G , we likewise see that G is transitive when and only when the substitutions of G_0 carry each letter of each Γ .

into every letter of the same Γ . Recalling our definition of the associated constituent groups $G'_0, G''_0, \dots, G^{(m-1)}_0$ of G , we thus have that G is transitive when and only when its associated constituent groups are transitive. As the latter are conjugate, it follows that G is transitive if any one of its associated constituent groups is known to be transitive.

Let $(G_0)_{ij}$ be the subgroup of G_0 which consists of all substitutions in G_0 that carry a_{ij} into itself. If we expand G in right cosets as regards $(G_0)_{ij}$, the members of each single coset carry a_{ij} into one and the same letter of Γ_{i+1} . Also, if s_1 and s_2 of G carry a_{ij} into the same letter, $s_1 s_2^{-1}$ will be in $(G_0)_{ij}$, so that s_1 and s_2 are in the same coset. Each coset therefore consists of all the substitutions of G carrying a_{ij} into the corresponding letter. If then G is transitive of degree n , there will be exactly n such cosets, one for each letter of Γ_{i+1} . Hence, the order of $(G_0)_{ij}$ is equal to g/n if G is a transitive group of order g , and degree n . *The order of a transitive m -adic substitution group is therefore a multiple of its degree.* Furthermore, *the number of substitutions of a transitive m -adic substitution group that carry any letter a_{ij} into any letter $a_{(i+1)}$ is, for all such pairs of letters, equal to the order of the group divided by its degree.*

Since for $m > 2$ an m -adic substitution group cannot carry a letter into itself, we have to turn to the associated group of a transitive m -adic substitution group for an average number of letters theorem. For this purpose we write the substitutions of the associated group in standard cycle form. Observe first that each associated constituent group $G_0^{(i)}$ being transitive, and of degree n , the average number of its letters appearing in its substitutions is $n-1$. Fixing our attention on $G_0^{(i)}$, we consider the subgroup $H_0^{(i)}$ of G_0 consisting of all the substitutions of G_0 whose component in $G_0^{(i)}$ is the identity of $G_0^{(i)}$. If we expand G_0 in cosets as regards $H_0^{(i)}$, each coset is easily seen to consist of all the substitutions of G_0 which have a fixed component in $G_0^{(i)}$. Each substitution of $G_0^{(i)}$ therefore occurs the same number of times in G_0 . It follows that *the average number of letters of each Γ_i occurring in the substitutions of the associated group of a transitive group of degree n is $n-1$.* This is our strongest result. From it, or from our discussion of $(G_0)_{ij}$, we also have that *the average number of all letters appearing in the substitutions of the associated group of a transitive m -adic substitution group of degree n is $(m-1)(n-1)$.* This may also be seen as follows. Since the containing group G^* of the transitive m -adic group G is transitive, and of degree $(m-1)n$, the average number of letters in its substitutions is $(m-1)n-1$. The total number of letters in its substitutions is then $(m-1)g[(m-1)n-1]$, g being the order of G . Of the $(m-1)g$ substitutions in G^* , the $(m-2)g$ substitutions not in G_0 each has its full complement of $(m-1)n$ letters. The total number of letters in the substitutions of G_0 is thus the remaining $(m-1)(n-1)g$ letters, whence the result.

14. Intransitive m -adic substitution groups. Let G be any m -adic substitution group on the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, and let $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ be the subclasses of the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ respectively into which $a_{(m-1)}$

is carried by the elements, dyads, \dots , $(m-1)$ -ads of G . If s is any substitution in G , then as r ranges through the i -ads of G , rs ranges through the $(i+1)$ -ads of G . Hence s transforms the letters of Γ'_i in 1-1 fashion into the letters of Γ'_{i+1} for each i , and thus determines an m -adic substitution on $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$. Furthermore, if a_{ij} and $a_{(i+1)k}$ are any two letters of Γ'_i and Γ'_{i+1} respectively, some s of G will carry a_{ij} into $a_{(i+1)k}$. For some i -ad r_1 of G carries $a_{(m-1)1}$ into a_{ij} , and some $(i+1)$ -ad r_2 of G carries $a_{(m-1)1}$ into $a_{(i+1)k}$. Hence element $r_1^{-1}r_2$ of G carries a_{ij} into $a_{(i+1)k}$. The m -adic substitutions on $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ obtained from all the substitutions of G therefore constitute a transitive m -adic substitution group on $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$. If G is transitive, this group is identical with G . If G is not transitive, we may call this group a *transitive constituent group* of the *intransitive group* G . In that case, by accounting for all the letters of Γ_{m-1} , we obtain a number of transitive constituent groups of G such that every substitution in G is the product of a selection of substitutions from the transitive constituent groups of G .

This result can also be obtained by analysing the containing group G^* of G , whence it also appears that the transitive constituent groups of G^* are the containing groups of the transitive constituent groups of G .

The direct product and simple isomorphism methods for obtaining intransitive ordinary groups admit of immediate extension to m -adic groups. In the latter case, let G_1 and G_2 be the same m -adic substitution group written on different letters. If the letters of G_1 form the sets $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$, of $G_2, \Gamma''_1, \Gamma''_2, \dots, \Gamma''_{m-1}$, the products of corresponding substitutions in G_1 and G_2 will be m -adic substitutions on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, where Γ_i consists of all the letters of Γ'_i and Γ''_i . Clearly, an m -adic substitution group is thus formed simply isomorphic with G_1 and G_2 , but of twice their degree. Similarly for any number of groups obtained by writing a given m -adic substitution group on different letters.

As for the direct product method, let H_1 and H_2 be m -adic substitution groups on $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ and $\Gamma''_1, \Gamma''_2, \dots, \Gamma''_{m-1}$ with all letters distinct. As before, form $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$. Then, if s'_i and s''_i be any two substitutions in H_1 and H_2 respectively, $s'_i s''_i$ will be an m -adic substitution on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$. The set of all such products clearly constitutes an m -adic substitution group G of order equal to the product of the orders of H_1 and H_2 , and degree equal to the sum of their degrees. When $m=2$, the existence of an identical element, coupled with the ambiguity of the cycle notation, allows us to consider H_1 and H_2 subgroups of G which can then be said to be generated by H_1 and H_2 . When $m>2$ this is no longer possible. We shall therefore refrain from calling G the direct product of H_1 and H_2 , reserving that phrase for a more special concept found useful in the sequel⁽⁵¹⁾.

15. Substitutions which are commutative with each of the substitutions

⁽⁵¹⁾ In fact, while G is an m -adic substitution group, the m -group "generated" by H_1 and H_2 is, for $m>2$, a hybrid sort of an affair of order $m-1$ times the order of G . On the other hand,

of a transitive m -adic substitution group. Recalling that the order of a transitive m -adic substitution group is a multiple of its degree, we may most briefly define a *regular* m -adic substitution group as a transitive m -adic substitution group whose order is equal to its degree. In view of the corresponding general result for transitive groups, this is equivalent to defining a regular m -adic substitution group as an m -adic substitution group, which, for any pair of letters in consecutive Γ 's, has one and only one substitution carrying the first letter into the second. Other transitive group results, coupled with the order criterion of regularity, show that an m -adic substitution group is regular if and only if its containing group is regular; also, if and only if its associated constituent groups are regular. The orders of the associated group, and the associated constituent groups, then being the same, it also follows that a regular m -adic substitution group is a transitive group whose associated group has no substitutions other than the identity omitting a letter. Regular m -adic substitution groups play the same role in polyadic as in ordinary group theory, since we later show that every finite abstract m -adic group can be represented as a regular m -adic substitution group.

According to a theorem of Jordan, the substitutions on the letters of a regular group commutative with each of its substitutions constitute a group conjugate to the regular group and known as its conjoint. We extend this theorem to a regular m -adic substitution group G by directly applying it to the containing group G^* , which is known to be regular. In fact, since G^* is generated by G , the ordinary substitutions on the letters of G commutative with each of its substitutions are the same as those commutative with each of the substitutions of G^* . Hence, to find the m -adic substitutions on the letters of G commutative which each of its substitutions we need merely pick out those substitutions in the conjoint of G^* which are m -adic substitutions.

To do this we must re-examine the standard proof of Jordan's theorem. In this proof the letters on which the given regular group is written are replaced by the symbols s_i used for the substitutions in the group. Then, in the simple isomorphism established between the group and its conjoint, the j th substitution of the given group in its new form replaces each symbol s_i by the symbol for the substitution $s_i s_j$, while the corresponding substitution in the conjoint replaces each s_i by $s_j s_i$. Finally, it is shown that the given group is transformed into its simply isomorphic conjoint by the substitution which carries the second letter of each substitution of the given group into the second letter of the corresponding substitution of the conjoint when all the substitutions⁽⁵²⁾ are written in cycle form with the same first letter.

if either H_1 or H_2 has a first order element, G will contain a subgroup H_2' or H_1' simply isomorphic to H_2 or H_1 respectively; and if both H_1 and H_2 possess an invariant first order element, the corresponding H_2' and H_1' will generate G , and G will then be the direct product of H_1' and H_2' in the sense later defined (§25).

(52) All except the identity, that is. Likewise later in the proof.

For our purpose it suffices to determine the nature of this substitution in the case of the regular group G^* . With G a regular m -adic substitution group on the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, the substitutions of G^* can be grouped into corresponding classes $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ according as they are elements, dyads, $\dots, (m-1)$ -ads of G . When G^* is rewritten in accordance with the proof of Jordan's theorem, $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ take the place of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$. In fact, if s_i is a k -ad in G^* , s_j an l -ad, $s_i s_j$ is a $(k+l)$ -ad. Hence, in the above description applied to G^* , if s_i is an l -ad, the j th substitution of G^* transforms each Γ'_k into Γ'_{k+l} , and hence is an l -ad on $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$. But the same reasoning shows the corresponding substitution in the conjoint of G^* also to be an l -ad on $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$. Or, returning to $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, we have that in the simple isomorphism between G^* and its conjoint the correspondant of an i -ad in G^* is an i -ad. If then we write the substitutions of G^* and its conjoint with the same first letter, say a_{11} , if the corresponding substitutions in G^* and its conjoint are both i -ads, their second letters will both be in Γ_{i+1} . Hence, the substitution which transforms G^* into its conjoint transforms each Γ into itself, and consequently is an $(m-1)$ -ad on the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$.

Our result now immediately follows. The $(m-1)$ -ad will transform only the m -adic substitutions of G^* into m -adic substitutions. Hence the m -adic substitutions in the conjoint of G^* are the transforms of the m -adic substitutions in G^* by the $(m-1)$ -ad, i.e., the transforms of the substitutions in G . Since the transform of an m -adic group is a simply isomorphic m -adic group, we thus have the following extension of Jordan's theorem. *The m -adic substitutions on the sequence of Γ 's of a regular m -adic substitution group commutative with all the substitutions of the group constitute a regular m -adic substitution group of the same order, and this group is the transform of the given group by an $(m-1)$ -ad of m -adic substitutions.* Clearly, the relationship between the two groups is a reciprocal one, and we may call each the *conjoint* of the other. Either directly, or as a consequence of a later general result on transforms, it may be verified that each group can be transformed into the other by an m -adic substitution, and hence, according to any m -adic definition, are conjugate.

We further have, as a result of the above discussion, that every ordinary substitution on the letters of a regular m -adic substitution group of degree n commutative with all the substitutions of the group are polyads of m -adic substitutions, there being n such i -ads for every i . Together they of course constitute the conjoint of the containing group of the given group; and this conjoint is now seen to be the containing group of the conjoint of the given group.

In passing from regular m -adic substitution groups to arbitrary transitive m -adic substitution groups for the purpose of extending Kuhn's theorem to m -adic groups, we shall adopt the viewpoint of the last paragraph, and seek

all substitutions on the letters of the transitive m -adic group commutative with each of its substitutions; for now m -adic substitutions of this kind will exist only if the given group satisfies a special condition. If G is a transitive m -adic substitution group, G^* is transitive. Again, the substitutions on the letters of G commutative with every substitution in G are the substitutions on the letters of G^* commutative with each of its substitutions, and hence can be found by applying Kuhn's theorem to G^* . As before, we assume all substitutions written in cycle form.

According to Kuhn's generalization of Jordan's theorem the number of substitutions on the letters of G^* commutative with each of its substitutions is the same as the number of letters omitted in all substitutions of G^* which omit a given letter. Actually, such substitutions will be in G_0 , the associated group of G . Let $\{a_{ij}\}$ designate the set of letters omitted by all substitutions of G^* that omit a_{ij} . Since G^* is transitive, it follows that if $a_{i_1j_1}$ is in $\{a_{ij}\}$, then $\{a_{i_1j_1}\} = \{a_{ij}\}$, and a substitution r of G^* , carrying a_{ij} into $a_{i_1j_1}$, carries the set of letters $\{a_{ij}\}$ into itself. But r carries all the letters of Γ_i into all those of Γ_{i_1} , and hence all the letters of $\{a_{ij}\}$ that are in Γ_i into all the letters of $\{a_{ij}\}$ that are in Γ_{i_1} . Hence, if there are α letters of $\{a_{ij}\}$ in one Γ , there are α letters of $\{a_{ij}\}$ in every Γ that has at least one of them. Now with the Γ 's arranged in a cycle, let δ be the least difference between the subscripts of consecutive Γ 's that have letters of $\{a_{ij}\}$. Then some δ -ad r in G^* will transform the set $\{a_{ij}\}$ into itself. As r will then carry the letters of $\{a_{ij}\}$ which are in any Γ_i into letters of $\{a_{ij}\}$ which are in $\Gamma_{i+\delta}$, it follows that the Γ 's having letters in $\{a_{ij}\}$ have subscripts which are in arithmetic progression, with the common difference, indeed, a divisor of $m-1$. Finally, the known properties of transitive groups show the different sets $\{a_{ij}\}$ to be mutually exclusive, and transformable into each other by the substitutions of G^* . It follows that the numbers α and δ are the same for all such sets; and since together they exhaust the letters of G^* , that α is a divisor of n . Hence the following result. *If G is a transitive m -adic substitution group of degree n , then the number of letters omitted by all substitutions of the containing group G^* that omit a given letter is of the form $\kappa\alpha$, where κ is a divisor of $m-1$, α a divisor of n ; furthermore, there are α of these letters in every Γ that has at least one, the subscripts of these Γ 's forming an arithmetic progression.*

According to the proof of Kuhn's theorem the resulting $\kappa\alpha$ substitutions on the letters of G^* commutative with each of its substitutions are obtained as follows. Let H_{11} be the subgroup of G^* composed of the substitutions of G^* which leave the set of letters $\{a_{11}\}$ unchanged, and let C_{11} be the conjoint of the regular group K_{11} , on the letters $\{a_{11}\}$, formed by the components on those letters of the substitutions in H_{11} . For each substitution in C_{11} , form the product of all the distinct transforms of that substitution under G^* . These products are the $\kappa\alpha$ substitutions on the letters of G^* commutative with each of its substitutions. According to our distribution result

for the set of letters $\{a_{11}\}$, there are α of these letters in each of the κ Γ 's, $\Gamma_1, \Gamma_{1+\delta}, \dots, \Gamma_{1+(\kappa-1)\delta}$, where $\kappa\delta = m - 1$. Clearly, the substitutions of H_{11} can only be i -ads with $i = \delta, 2\delta, \dots, \kappa\delta$. Since K_{11} is regular on the letters $\{a_{11}\}$, it will have, for each of the above i 's, α substitutions which are components of i -ads in H_{11} . Our proof of the extended Jordan theorem applies sufficiently to the relationship between K_{11} and its conjoint C_{11} to show that C_{11} also consists of α "components of i -ads" for each $i = \delta, 2\delta, \dots, \kappa\delta$. Now when a substitution of C_{11} is transformed by the substitutions of G^* , the set of letters $\{a_{11}\}$ will go over into all the mutually exclusive distinct sets $\{a_{ij}\}$, there being one and only one distinct transform of the substitution of C_{11} for each set $\{a_{ij}\}$. If the substitution in question is a component of an i -ad, each transform will also be a component of an i -ad. As the sets $\{a_{ij}\}$ are mutually exclusive, and exhaust the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, the product of these transforms will exactly constitute an i -ad. We thus have the following extension of Kuhn's theorem. *The only substitutions on the letters of the Γ 's of a transitive m -adic substitution group commutative with each of its substitutions are polyads of m -adic substitutions on the same sequence of Γ 's; in the notation of the distribution theorem, if $\delta = (m-1)/\kappa$, these polyads can only be i -ads with $i = \delta, 2\delta, \dots, \kappa\delta$, there being exactly α such i -ads for each admissible i .*

In particular, if we restrict our attention to m -adic substitutions, we have the following result. *The necessary and sufficient condition that there be at least one m -adic substitution on the sequence of Γ 's of a transitive m -adic substitution group commutative with each of its substitutions is that the subgroup of the associated group consisting of all its substitutions omitting a given letter in one Γ omits a fixed letter in the following Γ ; if then that subgroup omits exactly α letters from one Γ , it will omit α letters from every Γ , and there will be exactly α such m -adic substitutions.*

16. Holomorphs of a regular m -adic substitution group. The concept of holomorph of a regular group admits both of an immediate extension to regular m -adic substitution groups, as well as of a further generalization peculiar to polyadic theory. For the immediate extension, let G be a regular m -adic substitution group of order n on the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$. Then all the m -adic substitutions on $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ which transform G into itself constitute an m -adic substitution group of degree n which we shall call the *principal holomorph* of G . Clearly, the principal holomorph of G not only contains G , but also the conjoint of G . Since the transforms of commutative substitutions are commutative, it follows that the principal holomorph of G is in fact also the principal holomorph of its conjoint.

If K is the principal holomorph of G , then $(K_0)_{11}$, the subgroup of the associated group K_0 of K consisting of all the substitutions of K_0 omitting a_{11} , may be identified as the *group of isomorphisms* of G . That is, $(K_0)_{11}$ transforms G into all of its possible automorphisms, each automorphism being given by but one substitution of $(K_0)_{11}$. In fact, the argument used in extending

Jordan's theorem shows that the substitutions of one of two simply isomorphic regular m -adic substitution groups on the same sequence of Γ 's can be transformed into the corresponding substitutions of the other by an $(m-1)$ -ad which omits, say, a_{11} . Hence, every automorphism of G can be obtained by transforming G by the substitutions in $(K_0)_{11}$. Furthermore, if two distinct substitutions of $(K_0)_{11}$ yielded the same automorphism of G , a substitution of $(K_0)_{11}$ other than the identity would transform each member of G into itself. But this substitution would have to be in the associated group of the conjoint of G , and, as this conjoint is regular, the substitution in question, which omits a_{11} , could only be the identity.

We can now prove, as in the ordinary case, that *the order of the principal holomorph of a regular m -adic substitution group is equal to the product of the order of the group and the order of its group of isomorphisms*. In fact, if \bar{G} is the conjoint of the regular group G , K its holomorph, by expanding K in cosets as regards its invariant subgroup \bar{G} , we see that the substitutions of K transform G in k/n different ways, k being the order of K . But K as well as $(K_0)_{11}$ must transform G into all of its possible automorphisms. For if s is in K , t in $(K_0)_{11}$, as t runs through $(K_0)_{11}$ giving all the automorphisms of G , ts in K yields an equal number of automorphisms of G . Hence the order of $(K_0)_{11}$ is k/n , whence the above result.

To illustrate this result, consider the cyclic triadic group of degree and order two generated by the triadic substitution $s_1 = (a_{11}a_{21}a_{12}a_{22})$ given in cycle form. The letters of Γ_1 are a_{11}, a_{12} , of Γ_2 , a_{21}, a_{22} . The sole other triadic substitution on Γ_1, Γ_2 generated by s_1 is $s_2 = (a_{11}a_{22}a_{12}a_{21})$, so that the group is seen to be regular. s_2 also is a generator of the group, whence it follows that the group admits exactly two automorphisms. Hence the order of its principal holomorph is four. We find directly that $s_3 = (a_{11}a_{21})(a_{12}a_{22})$ and $s_4 = (a_{11}a_{22})(a_{12}a_{21})$ interchange s_1 and s_2 , so that the principal holomorph consists of s_1, s_2, s_3, s_4 . It is actually the entire triadic symmetric group of degree two. This example serves to answer the question whether some subgroup of the principal holomorph itself, instead of its associated ordinary group, can be identified with the group of isomorphisms of the given group. The answer in the present instance is no. For such a subgroup would have to possess as element s_1 or s_2 to yield the identical automorphism, but would then have for elements both s_1 and s_2 , each yielding that one automorphism.

The immediate extension of the concept of complete group to m -adic groups turns out to be rather trivial. Defining an m -adic group G to be *complete in the narrow sense* if its own elements transform it in 1-1 fashion into all of its possible automorphisms, we obtain the following result. *An m -group is complete in the narrow sense when and only when it is reducible to a complete ordinary group.* In fact, its sole element yielding the identical automorphism must be of first order, and invariant under the group—hence the reducibility.

The rest of the theorem follows from the easily demonstrated facts that if G is reducible to G' , every automorphism of one group is also an automorphism of the other, while the automorphisms induced by any element of either is the same for both. Since G can have but one invariant element, it also follows that *the net of derived groups of an m -adic group complete in the narrow sense consists of a single complete 2-group, and its extensions, which are then also complete in the narrow sense*. If G is regular, and complete in the narrow sense, we may use its elements as multipliers in the expansion of K in cosets as regards \bar{G} . We shall express this fact by saying that *the principal holomorph of an m -group complete in the narrow sense is the direct product of the group and its conjoint*. A precise abstract definition of this rather narrow concept of direct product will be given in §25.

We do not obtain a less restrictive concept of completeness by asking that the elements of G_0 transform G into all of its possible automorphisms in 1-1 fashion; for the coset theorem shows that G and G_0 transform G according to the same number of distinct automorphisms. We therefore define G to be *complete in the wide sense* if the elements of its abstract containing group G^* transform it in 1-1 fashion according to all of its possible automorphisms. Since only the identity of G^* is now invariant under G , it follows that an m -group complete in the wide sense is irreducible. If this m -group G is of order g , and is expressed as a regular m -adic substitution group, the order of its principal holomorph K will be $(m-1)g^2$. We now turn to the containing groups for a direct product theorem, and easily find that *the containing group of the principal holomorph of an m -group complete in the wide sense is the direct product of the containing groups of the group and its conjoint*.

Actually, a type of completeness can be defined for each divisor k of $m-1$, an m -group G being said to be complete in the k -sense if it admits some containing group of index k whose elements yield in 1-1 fashion all the automorphisms of G . We then have that an m -group is complete in the k -sense when and only when it is reducible to a $(k+1)$ -group complete in the wide sense. Furthermore, the net of derived groups of an m -group complete in the k -sense consists of a single $(k+1)$ -group complete in the wide sense, and its extensions. With G written as a regular m -adic substitution group, its principal holomorph will of course be of order kg^2 . But there does not then seem to be a direct product theorem in terms of groups.

We turn now to the purely m -adic generalization of holomorph. In ordinary group theory, due to the presence of the identity, if all of the elements of a group H transform a group G into one and the same group, that group must be G itself. Hence if G is a regular substitution group, H a substitution group on the letters of G , H will be the holomorph of G , or a subgroup thereof. This need not be so for m -adic groups with $m > 2$. Let then G be a regular m -adic substitution group with $m > 2$, H an m -adic substitution group on the

sequence of Γ 's of G such that all of the substitutions of H transform G into one and the same group G'' ⁽⁵³⁾. It follows that all of the substitutions of H transform G'' into one and the same group G''' , and so on. Since the m -ads of H must transform G as do its elements, it follows that there will be a cycle of $\mu - 1$ distinct, though not necessarily mutually exclusive, m -adic groups $(G', G'', \dots, G^{(\mu-1)})$, such that $G' = G$, $\mu - 1$ is a divisor of $m - 1$, and all the elements of H transform each G into the cyclically following G . Now all the m -adic substitutions on the sequence of Γ 's of G which transform $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(\mu-1)} \rightarrow G$ constitute an m -adic substitution group K containing H . We shall then call K a *holomorph* of G , and the *holomorph* of the cycle $(G', G'', \dots, G^{(\mu-1)})$. When $\mu - 1 = 1$, K becomes the principal holomorph of G .

Given the regular G , an m -adic substitution s on the sequence of Γ 's of G will be said to be *holomorphic* if it belongs to some holomorph of G . We then readily see that the necessary and sufficient condition that s be *holomorphic* is that s^{m-1} is in the associated group of the principal holomorph of G . For that associated group consists of all the $(m - 1)$ -ads on the sequence of Γ 's of G which transform G into itself. The necessity of the condition then follows from the fact that s^m must transform G into the same group that s does, the sufficiency from the fact that all the elements of the cyclic m -group generated by s will then transform G into one and the same group. In particular, the $(n!)^{m-2}$ first order substitutions of degree n are all holomorphic for the regular G of degree n . Hence, when the order of the principal holomorph of G is less than $(n!)^{m-2}$, as must be so, for example, in the case of cyclic m -groups of order greater than three, we are assured of the existence of a holomorph other than the principal holomorph. Clearly, any element s of a holomorph of G determines the corresponding cycle $(G', G'', \dots, G^{(\mu-1)})$, and hence the holomorph. It follows that all the holomorphs of a given G are mutually exclusive.

Our next result shows that the order of any holomorph of G is no greater than that of the principal holomorph of G . In fact, let $K', K'', \dots, K^{(\mu-1)}$ be the principal holomorphs of $G', G'', \dots, G^{(\mu-1)}$, K the holomorph of the cycle $(G', G'', \dots, G^{(\mu-1)})$. By writing an element t of K_0 as the product of $m - 1$ elements of K , we see that t must leave each $G^{(i)}$ invariant, and hence be in each $K_0^{(i)}$. Conversely, if t is in each $K_0^{(i)}$, it will transform each $G^{(i)}$ into itself. If then s is in K , ts will also be in K , so that t must be in K_0 . That is, the associated group of the holomorph of $(G', G'', \dots, G^{(\mu-1)})$ is the logical product of the associated groups of the principal holomorphs of $G', G'', \dots, G^{(\mu-1)}$. It is readily verified that an s in K actually transforms $K' \rightarrow K'' \rightarrow \dots \rightarrow K^{(\mu-1)} \rightarrow K'$, and hence also $K'_0 \rightarrow K''_0 \rightarrow \dots \rightarrow K^{(\mu-1)}_0 \rightarrow K'_0$. Hence, a subgroup of K'_0 invariant under s must be contained in K_0 . We thus have, in terms of G alone, the associated group of the holomorph of G correspond-

(53) The reader will note the marked analogy with Corral's concept of a function pertaining to a brigade, that is, one carried into the same function by all the substitutions of the brigade.

ing to a holomorphic s is the largest group or subgroup of the associated group of the principal holomorph of G invariant under s .

On turning to an order theorem for these m -adic holomorphs, we observe first that the holomorph of a cycle $(G', G'', \dots, G^{(\mu-1)})$ is also the holomorph of the "conjoint cycle" $(\bar{G}', \bar{G}'', \dots, \bar{G}^{(\mu-1)})$. For, inasmuch as transforms of commutative substitutions are commutative, transforms of conjoint regular groups are conjoint. Hence, if s transforms $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(\mu-1)} \rightarrow G'$, it must transform $\bar{G}' \rightarrow \bar{G}'' \rightarrow \dots \rightarrow \bar{G}^{(\mu-1)} \rightarrow \bar{G}'$, and conversely. Now let K be the holomorph of the cycle $(G', G'', \dots, G^{(\mu-1)})$. Each s in K transforms in 1-1 fashion the elements of $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(\mu-1)} \rightarrow G'$, and hence determines a μ -adic substitution on the $\mu - 1$ classes of elements $G', G'', \dots, G^{(\mu-1)}$ which may be termed a μ -adic automorphism of the cycle $(G', G'', \dots, G^{(\mu-1)})$. The class of all such μ -adic substitutions on $G', G'', \dots, G^{(\mu-1)}$ obtained through substitutions in K clearly constitutes an m -adic group which we shall term the *restricted m -adic group of isomorphisms* of the cycle $(G', G'', \dots, G^{(\mu-1)})$, restricted, both by the possible narrowness of K , and by the fact that while an m -adic substitution will transform any one $G^{(i)}$ into $G^{(i+1)}$ according to any simple isomorphism, it need not be able to do this arbitrarily and simultaneously for each i . Now s_1 and s_2 of K will yield the same μ -adic automorphism of the cycle $(G', G'', \dots, G^{(\mu-1)})$ when and only when the $(m-1)$ -ad $s_2 s_1^{-1}$ transforms each element of each $G^{(i)}$ into itself, and hence, when and only when $s_2 s_1^{-1}$ is in $\bar{G}_0 = \bar{G}' \bar{G}'' \dots \bar{G}^{(\mu-1)}$ ⁽⁵⁴⁾. Note that K_0 consists of all $(m-1)$ -ads which transform each $G^{(i)}$ into itself, and hence has \bar{G}_0 for subgroup, one, indeed, invariant under K . By expanding K in cosets as regards \bar{G}_0 , we then obtain the following result. *The order of the holomorph of $(G', G'', \dots, G^{(\mu-1)})$ is the product of the order of the crosscut of the associated groups of the conjoints of $G', G'', \dots, G^{(\mu-1)}$ and the order of the restricted m -adic group of isomorphisms of $(G', G'', \dots, G^{(\mu-1)})$.*

This result is weaker than the result for the principal holomorph of G in two ways. On the one hand, the order of \bar{G}_0 replaces the order of G itself. More significantly, in the case of the principal holomorph, we identified $(K_0)_{11}$ with the group of isomorphisms of G . Note that there both K and K_0 yielded every possible automorphism of G . In the present case K_0 transforms each $G^{(i)}$ into itself, the distinct transformations being $(\mu-1)$ -ads of μ -adic substitutions on $G', G'', \dots, G^{(\mu-1)}$, and constituting the associated group of the restricted m -adic group of isomorphisms of the cycle $(G', G'', \dots, G^{(\mu-1)})$. If then we ask whether $(K_0)_{11}$ can be identified with this associated restricted group of isomorphisms, we find that while no two members of $(K_0)_{11}$ can transform the $G^{(i)}$'s in the same way, for $(K_0)_{11}$ to transform the G 's in every way that K_0 does, it is necessary and sufficient that K_0 carry a_{11} into no other letters than does its subgroups \bar{G}_0 . We have not succeeded in answering the

⁽⁵⁴⁾ Product here is logical product.

question thus posed; and hence, whether $(K_0)_{11}$, or any other subgroup of K_0 , can be identified as the associated restricted group of isomorphisms of the cycle $(G', G'', \dots, G^{(\mu-1)})$ remains one of our unsolved problems⁽⁵⁵⁾.

17. *m*-adic groups of μ -adic substitutions. The present extension of the concept of *m*-adic substitution group is indispensable for a self-contained theory of primitivity, our next topic. This extension has the advantage of including *m*-adic groups of ordinary substitutions in its scope. However, the fact that any abstract *m*-adic group can be represented as a regular *m*-adic substitution group is perhaps sufficient reason for our restricting the explicit study of this wider class of substitution groups to the next section.

Given a cycle of $\mu - 1$ equivalent classes $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$, not only will the product of μ μ -adic substitutions on these Γ 's be a substitution of the same kind, but also the product of any *m* such substitutions, provided *m* is in the form $k(\mu - 1) + 1$. We are thus led to the concept of an *m*-adic group of μ -adic substitutions, or (m, μ) substitution group, with *m* and μ subject to the sole condition that $\mu - 1$ be a divisor of *m* - 1. We have already met this concept in the last section where the corresponding Γ 's, $G', G'', \dots, G^{(\mu-1)}$, while distinct, were probably not necessarily mutually exclusive⁽⁵⁶⁾. In what follows, for simplicity, as in our previous development, we shall assume the Γ 's to be mutually exclusive.

It is not difficult to review our previous work to see how much goes over to (m, μ) substitution groups. The chief failure turns out to be the extension of Jordan's theorem on regular groups. Particular mention must be made of the structure of the containing group of an (m, μ) group *G*. Letting, for simplicity, *G** symbolize the containing ordinary group of *G* generated by the elements of *G*, *G** will now be of some index *k* which is a divisor of *m* - 1 and a multiple of $\mu - 1$. We must now distinguish between *i*-ads of *G* and *i*-ads in *G**, the former being the products of any *i* substitutions in *G*, the latter all products in *G** of *i* μ -adic substitutions. In particular, there will be $k/(\mu - 1)$ cosets in *G** consisting of $(\mu - 1)$ -ads, one and only one of these cosets being *G*o.

In connection with the next section, the extension of the concept of transitivity to (m, μ) substitution groups is of most importance. Actually, our definition of transitivity as applied to *m*-adic substitution groups can be restated

⁽⁵⁵⁾ The above theory of *m*-adic holomorphs can be paraphrased for ordinary groups. Thus, if *s* is a substitution on the letters of an ordinary regular group *G*', but not in the holomorph of *G*', and if s^{m-1} is the first positive power of *s* in the holomorph of *G*', then a cycle of regular groups *G'*, *G''*, \dots , $G^{(m-1)}$ is determined such that, under *s*, $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(m-1)} \rightarrow G'$. The set of all substitutions on the letters of *G*' thus transforming this now given cycle of *G*'s will then constitute an *m*-adic group of ordinary substitutions, which may then be called an *m*-adic holomorph of *G*. The above theory, in somewhat simpler form, will then go over.

⁽⁵⁶⁾ On the other hand, the most general possibility is still not there illustrated; for a given element is carried into a single element independently of the $G^{(i)}$ of which it is an element. Note that our last footnote further introduced an $(m, 2)$ substitution group as *m*-adic holomorph of an ordinary regular group.

verbatim for (m, μ) substitution groups. It is then readily verified that all of the results of §13 go over, with the possible replacement of m by μ , with one exception. And that is that the transitivity of G^* no longer assures the transitivity of G ⁽⁵⁷⁾.

18. Primitive and imprimitive (m, μ) substitution groups. The distinct sets $\{a_{ij}\}$ of §15 are transformed as units under all the substitutions of the containing group G^* of the transitive m -adic substitution group G , and hence under the substitutions of G . We recall that each set $\{a_{ij}\}$ had α letters in each of κ Γ 's whose subscripts formed an arithmetic progression, α being a divisor of n , the degree of G , κ of $m-1$. Let $\nu = n/\alpha$, $m'-1 = (m-1)/\kappa$. Each set $\{a_{ij}\}$ then has letters in one and only one of the first $(m'-1)$ Γ 's, there being ν sets for each such Γ . The $(m'-1)\nu$ distinct sets $\{a_{ij}\}$ thus fall into $m'-1$ mutually exclusive classes $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m'-1}$ of ν members each. As any m -adic substitution s in G transforms each Γ_i into Γ_{i+1} , it will in 1-1 fashion transform the members of $\Gamma'_1 \rightarrow \Gamma'_2, \Gamma'_2 \rightarrow \Gamma'_3, \dots, \Gamma'_{m'-1} \rightarrow \Gamma'_1$, and so define an m' -adic substitution on $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m'-1}$. The totality of these m' -adic substitutions will then constitute an (m, m') substitution group G' of degree ν . As G is transitive, it follows that G' is transitive, there being, as in the ordinary case, a $(1, N)$ isomorphism between G' and G . With a restriction to be noted later, G will be said to be imprimitive with systems of imprimitivity $\{a_{ij}\}$ whenever $1 < (m'-1)\nu < (m-1)n$, this however, as in the ordinary case, being but an example of the general concept of imprimitivity.

We thus see that even if we start with transitive m -adic substitution groups, i.e., (m, m) groups, we are led to (m, μ) groups. This extension is however sufficient for our purpose. For if we start with a transitive (m, μ) substitution group G , and define the sets $\{a_{ij}\}$ as before, we obtain, by the same argument, an analogous distribution theorem, and then, as above, a transitive (m, μ') substitution group G' with $\mu'-1$ a divisor of $\mu-1$.

In general, then, let G be a transitive (m, μ) substitution group of degree n on $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$, with, of course, $\mu-1$ a divisor of $m-1$, and let there be some separation of the $(\mu-1)n$ letters of the Γ 's into mutually exclusive classes such that these classes are transformed as units under all the substitutions of G , and hence of G^* . An entirely analogous argument to the one used in determining the distribution of the letters in the sets $\{a_{ij}\}$ leads to a corresponding conclusion here. That is, each class consists of the same number $\kappa\alpha$ of letters, with α a divisor of n , κ of $\mu-1$, there being α letters in each of κ Γ 's whose subscripts are in arithmetic progression. As with the $\{a_{ij}\}$'s, each such

(57) On the other hand, if G^* is transitive, we may form a sequence of mutually exclusive Γ 's, $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{\mu'-1}$, such that G is a transitive (m, μ') group on the Γ' 's. Here $\mu'-1$ is a multiple of $\mu-1$, and a divisor of $m-1$; while the Γ' 's are successively subclasses of the Γ 's run through cyclically $(\mu'-1)/(\mu-1)$ times, and together exhausting the Γ 's. If we call such an (m, μ) -group G semi-transitive, then the main result of §14 goes over for an arbitrary (m, μ) -group if we replace the transitive constituent groups by semi-transitive constituent groups.

separation of the letters of the Γ 's leads to a transitive (m, μ') substitution group G' of degree ν , where $\nu = n/\alpha$, $\mu' - 1 = (\mu - 1)/\kappa$. The numbers α and κ , of course, depend on the separation in question.

In accordance with the standard definition we would then define G to be imprimitive if some such separation into mutually exclusive classes is possible with $1 < (\mu' - 1)\nu < (\mu - 1)n$, the classes then being the corresponding systems of imprimitivity of G . This restriction is equivalent to κ and α not being both one, or simultaneously equal to $\mu - 1$ and n respectively. But then G would always be imprimitive for $\mu > 2$, since its substitutions transform the Γ 's themselves as units. We therefore exclude the case $\alpha = n$, and thus have the following definition. G will be said to be imprimitive if it admits systems of imprimitivity for which $\alpha < n$, κ and α not both unity; otherwise G will be said to be primitive.

The rather artificial restriction $\alpha < n$ is entirely natural in the case $\mu = 2$, and, indeed, we have here the only complete generalizations of the primitivity theorems of ordinary groups. G is now an m -adic group of ordinary substitutions on letters which may then be written a_1, a_2, \dots, a_n . It is easily seen that if G is transitive, so are G^* and G_0 , the converse however holding only for G_0 . On the other hand, with G transitive, if G is imprimitive, so are G^* and G_0 , the converse holding only for G^* . Thus, with $m = n + 1$, G can be the intransitive group consisting of the single substitution $(a_1 a_2 \dots a_n)$, while G^* is transitive. And the following is an example of a transitive primitive G for which G_0 is imprimitive. Let G_0 be the transitive imprimitive group of order four: 1, $(a_1 a_2)(a_3 a_4)$, $(a_1 a_3)(a_2 a_4)$, $(a_1 a_4)(a_2 a_3)$. Then $s = (a_1 a_2 a_3)$ transforms G_0 into itself, while $s^3 = 1$ is in G_0 . Hence $G = G_0 s$ is a transitive tetradic group of ordinary substitutions, and is easily seen to be primitive.

Turning to the ordinary theorems on primitivity, with G thus a transitive $(m, 2)$ group, there will be at least one substitution in G carrying a_1 into itself, and the totality of these substitutions will constitute a subgroup G_1 of G . We then have the complete analogue of the corresponding theorem for ordinary substitution groups, i.e., a necessary and sufficient condition that a transitive m -adic group G of ordinary substitutions is imprimitive is that G_1 is contained in a larger subgroup of G . While this can be proved by applying the ordinary theorem to G^* , the ordinary proof⁽⁵⁸⁾ can here be directly carried over. Thus, if G is imprimitive, the substitutions of G transforming the system of imprimitivity of which a_1 is a member into itself constitute a subgroup K of G containing G_1 , and larger than G_1 . Conversely, if K is a subgroup of the transitive G containing G_1 , and larger than G_1 , expand G in right cosets as regards K . Each coset will consist of the same number $\alpha > 1$ of right cosets of G as regards G_1 , and hence will carry a_1 into α letters, distinct for each coset, and will

⁽⁵⁸⁾ Rather what the proof of the more general theorem of *Finite Groups*, page 39, would become if given directly for the more special result. We have interchanged the order of the two results.

consist of all the substitutions of G carrying a_1 into one of those letters. Finally, any substitution s of G will transform these mutually exclusive sets of letters as units. For if K_0s_1 is the coset carrying a_1 into one of these sets, that set will be transformed by s as a_1 is by K_0s_1s . But K_0s_1s is the same as $s_0K_0s_2$, with s_0 in G_1 , s_2 some element in G , and hence transforms a_1 into that one of the above sets into which the coset K_0s_2 carries a_1 .

We shall prove in §24 that if an element or subgroup of an m -group G is transformed by the elements of G , the resulting set of distinct transforms constitutes a "complete set of conjugates" under G , and is transformed by the elements of G according to a transitive m -adic group of ordinary substitutions having a $(1, N)$ isomorphism with G . We again then are concerned with the case $\mu = 2$; and either by applying the preceding result in conjunction with that isomorphism, or by directly extending the ordinary proof as was done above, we again obtain the complete analogue of the corresponding ordinary group theorem⁽⁵⁹⁾. *A necessary and sufficient condition that a complete set of conjugate elements or subgroups under an m -group G of an element or subgroup of G is transformed under G according to an imprimitive m -adic group of ordinary substitutions is that the largest subgroup of G which transforms into itself one of these elements or subgroups is contained in a larger subgroup of G .*

When G is a transitive (m, μ) group with $\mu > 2$ we no longer have an analogue of G_1 for G itself. We must therefore go outside of G for theorems on imprimitivity. G^* will still be transitive; and apart from the restriction $\alpha < n$, a set of systems of imprimitivity of either G or G^* will also be one of the other. Our description of the possible systems of imprimitivity of G therefore applies equally well to G^* , and we conclude that G will be imprimitive when and only when G^* admits a set of systems of imprimitivity for which $\alpha < n$. As G^* is an ordinary transitive substitution group, we easily supplement the standard result concerning its imprimitivity to obtain the following. *A transitive (m, μ) group G is imprimitive when and only when G^* has a subgroup containing G_{11}^* , larger than G_{11} , but not containing G_0 . G_{11}^* is of course the subgroup of G^* consisting of all of its substitutions omitting a_{11} . In proving this result we observe that as a consequence of the transitivity of G the substitutions of G_0 will carry any letter into every letter in its Γ . If then G , and hence G^* , is imprimitive, the subgroup K of G^* , composed of all the substitutions of G^* which transform the system of imprimitivity of which a_{11} is a member into itself, satisfies the conditions of the theorem. For K is known to be a subgroup of G^* containing G_{11}^* , and larger than G_{11} . And as it can carry a_{11} into only $\alpha < n$ letters of Γ_1 , it cannot contain G_0 . Conversely, if K is a subgroup of G^* satisfying the conditions of the theorem, the letters into which the substitutions of K carry a_{11} are known to form one of a set of systems of imprimitivity of G^* . As K will then contain all the substitutions of G^* which carry a_{11} into any letter that*

⁽⁵⁹⁾ At least as stated on page 39, *Finite Groups*.

one substitution of K carries a_{11} into, could it carry a_{11} into all the letters of Γ_1 it would contain G_0 , contrary to hypothesis. Hence, the α of the resulting systems of imprimitivity of G is less than n , whence G too is imprimitive.

A criterion for the imprimitivity of G in terms of G_0 would be preferable to one in terms of G^* . Our example of a primitive $(m, 2)$ group whose associated group was imprimitive precludes such a criterion for an arbitrary (m, μ) group. However when $\mu=m$ we do have the following partial criterion in terms of, better than G_0 , the associated constituent groups of G . *A transitive m -adic substitution group G admits systems of imprimitivity with $\alpha > 1$ when and only when the associated constituent group G'_0 is imprimitive.* In fact, systems of imprimitivity of G must be permuted as units under G_0 . Hence the portions of these systems in Γ_1 are permuted as units under G'_0 . As $\alpha > 1$ by hypothesis, and $\alpha < n$ by definition, we thus have a set of systems of imprimitivity of G'_0 . Conversely, given a set of systems of imprimitivity of G'_0 , any s of G will transform $G'_0 \rightarrow G''_0 \rightarrow \dots \rightarrow G^{(m-1)}_0 \rightarrow G'_0$, and hence will successively transform the systems of imprimitivity of G'_0 into systems of imprimitivity of $G''_0, \dots, G^{(m-1)}_0$. The result of transforming these systems of imprimitivity of $G^{(m-1)}_0$ by s is the same as that of transforming the given systems of imprimitivity of G'_0 by s^{m-1} . As s^{m-1} is in G_0 , it transforms the systems of imprimitivity of G'_0 as units. Hence s transforms the totality of systems of imprimitivity of $G'_0, G''_0, \dots, G^{(m-1)}_0$ as units. As G_0 does the same, so will $G = G_0 s$ which is therefore imprimitive with $1 < \alpha (< n)$.

For the exceptional case $\alpha=1$ we have to return to G^* . The same considerations that gave us our general criterion for the imprimitivity of a transitive (m, μ) group yield the following result. *A transitive (m, μ) group G admits systems of imprimitivity with $\alpha=1$ when and only when G^* has a subgroup containing G_{11}^* , larger than G_{11}^* , but having no other substitutions than those of G_{11}^* that carry each Γ into itself.* The last condition is equivalent to the crosscut of the subgroup in question and G_0 being identical with $(G_0)_{11}$, the crosscut of G_{11}^* and G_0 , and hence, for an m -adic substitution group G , with G_{11}^* .

Though all of the development of the next section, and, with certain restrictions, of the one following, can be given for (m, μ) substitution groups, we restrict ourselves, for the sake of simplicity, to m -adic substitution groups, i.e., (m, m) groups.

19. Multiple transitivity; cyclically transitive m -adic substitution groups. Various extensions of the concept of multiple transitivity suggest themselves. According to the simplest, an m -adic substitution group G would be said to be r -fold transitive if any r letters belonging to any one Γ can be transformed into any r letters of the succeeding Γ by the substitutions of the group. It is readily proved that a necessary and sufficient condition that G be thus r -fold transitive is that the associated constituent groups $G'_0, G''_0, \dots, G^{(m-1)}_0$ (or any one of them) be r -fold transitive in the ordinary sense. Since the order of G'_0 is a divisor of the order of G_0 , and hence of G , it follows from the corre-

sponding ordinary group result that the order of an r -fold transitive m -adic substitution group of degree n is a multiple of $n(n-1) \cdots (n-r+1)$.

Of special interest in polyadic theory is the type of multiple transitivity we term cyclic transitivity. An m -adic substitution group G will be said to be *cyclically transitive* if, given any two selections from the classes of letters $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$, some substitution of G will carry the letters of one selection into the letters of the other. Actually then, if one selection is $a_{1j_1}, a_{2j_2}, \dots, a_{(m-1)j_{m-1}}$, the other $a_{1k_1}, a_{2k_2}, \dots, a_{(m-1)k_{m-1}}$, the substitution will carry $a_{1j_1} \rightarrow a_{2k_2}, a_{2j_2} \rightarrow a_{3k_3}, \dots, a_{(m-1)j_{m-1}} \rightarrow a_{1k_1}$. Every cyclically transitive m -adic substitution group is then transitive, and, indeed, for $m=2$ cyclic transitivity reduces to transitivity. The symmetric and alternating m -adic groups of degree n , previously observed to be transitive—in the latter cases at least for $n > 2$ —are now seen to be cyclically transitive.

The $m-1$ Γ 's, of n letters each, give rise to n^{m-1} selections which we shall call *cycles*. Any m -adic substitution s on the Γ 's will merely permute these cycles, and hence will determine an ordinary substitution on these n^{m-1} cycles as new "permutants." Since this relationship is preserved under multiplication, the members of an m -adic substitution group G of degree n will thus give rise to substitutions on the cycles forming an m -adic group G' of ordinary substitutions, of degree n^{m-1} , isomorphic with G . Clearly, different m -adic substitutions yield different substitutions of the cycles. Hence G' is indeed simply isomorphic with G . In particular, then, G' and G are of the same order. Finally, if G is cyclically transitive, G' will be transitive, and, indeed, conversely. Since the order of an m -adic transitive group of ordinary substitutions is a multiple of its degree, we have, as our first result, *the order of a cyclically transitive m -adic substitution group of degree n is a multiple of n^{m-1}* .

We have seen that transitive m -adic groups of ordinary substitutions have the complete analogue of the G_1 of ordinary transitive groups. Actually, all of the corresponding theory goes over. In our simple isomorphism between G and G' , the subgroup of G' consisting of all of its substitutions "omitting" a given cycle C will correspond to the subgroup G_C of G consisting of all of its substitutions which carry C into itself. We shall call G_C the *cycle subgroup of G* , corresponding to C . Actually, if C is the selection $a_{1j_1}, a_{2j_2}, \dots, a_{(m-1)j_{m-1}}$, it will be transformed into itself according to the cyclic substitution $(a_{1j_1} a_{2j_2} \cdots a_{(m-1)j_{m-1}})$ by all the substitutions of G_C . Hence, if the m -adic substitutions of G be written as ordinary substitutions on $(m-1)n$ letters in cycle form, G_C will consist of those substitutions of G which have this cyclic substitution as component. It will be convenient to speak of these substitutions as having the cycle C . Clearly, then, *an m -adic substitution of degree n cannot have more than n cycles*.

Each of the n^{m-1} cycles yields thus a corresponding cycle subgroup of the cyclically transitive G . The simple isomorphism between G and G' then immediately transforms the corresponding properties of G' to yield, among

others, the following results on G . The order of each cycle subgroup of G is equal to the order of G divided by n^{m-1} . The cycle subgroups of G are conjugate, forming a complete set of conjugates under the substitutions of G . If all the substitutions of a given cycle subgroup have exactly α cycles in common, and they have one cycle in common by definition, then the n^{m-1} cycles can be separated into mutually exclusive sets of α cycles each such that different cycles yield the same cycle subgroup when and only when they belong to the same set.

There are thus n^{m-1}/α distinct cycle subgroups of G . The only information that G' yields concerning α is that it is a divisor of n^{m-1} . However, our observation that an m -adic substitution of degree n cannot have more than n cycles shows that $\alpha \leq n$. Hence, a cyclically transitive m -adic substitution group of degree n has a number N of cycle subgroups with $N \geq n^{m-2}$ and a divisor of n^{m-1} . Whether N is actually a multiple of n^{m-2} , i.e., α a divisor of n , is another of our unsolved problems.

20. Class of an m -adic substitution group. The class of an ordinary substitution group is the smallest number of letters appearing in any of its substitutions, other than the identity, when those substitutions are written in cycle form. Since the substitutions of an m -adic substitution group G never carry a letter into itself when $m > 2$, we are led to define the class of G as the class of its associated group G_0 . This also is the class of its containing group G^* ⁽⁶⁰⁾.

With this definition most of the elementary theory of class goes over to m -adic substitution groups. We have almost immediately that the m -adic symmetric group of degree n , $n > 1$, is a primitive group of class 2, while the m -adic alternating groups of degree n , $n > 2$, are primitive groups of class 3. That these m -adic groups are primitive follows from the fact that their constituent associated groups are either the symmetric, or alternating, ordinary groups of degree n , and hence primitive, so that the m -adic groups do not admit systems of imprimitivity with $\alpha > 1$; and, being cyclically transitive, they cannot admit systems of imprimitivity with $\alpha = 1$. As for their class, they are clearly at most of the class indicated. And could an alternating group actually be of class 2, the ϵ -subgroup of its corresponding δ -subgroup would have an ϵ -sequence with one -1; but then the δ -subgroup would be the complete δ -group, and hence the given group not an alternating group, but the symmetric group.

We now prove that, as in the standard theory, the converses of these results also hold. First then let G be a primitive m -adic substitution group of degree n and of class 2. On the one hand, its associated constituent group G'_0 will be primitive; on the other hand, its associated group G_0 will have some substitution whose component in each $G_0^{(0)}$ but one is the identity, and in

⁽⁶⁰⁾ Not necessarily so, however, for (m, μ) -groups with $\mu < m$.

that one a transposition. By the invariance of G_0 under G , we see that G_0 has a substitution t_0 of the form $t'_0 \cdot 1 \cdots 1$, with t'_0 in G'_0 and a transposition. Now let \bar{G}'_0 be that subgroup⁽⁶¹⁾ of G'_0 composed of all the substitutions t' of G'_0 for which $t' \cdot 1 \cdots 1$ is in G_0 . The subgroup \bar{G}'_0 is clearly an invariant subgroup of G_0 , and hence of G'_0 , and it has the transposition t'_0 . Now the standard proof of the fact that a primitive (ordinary) group of class 2 is the corresponding symmetric group also yields the following more general statement. An invariant subgroup⁽⁶²⁾ of class 2 of a primitive group is the corresponding symmetric group. Hence \bar{G}'_0 is the symmetric group of degree n . G_0 therefore has among its elements every substitution of the form $t' \cdot 1 \cdots 1$. Since G_0 is invariant under G , it also has every substitution of the form $1 \cdots t^{(i)} \cdots 1$, for each i , and hence every substitution of the form $t't'' \cdots t^{(m-1)}$. G_0 is therefore the associated group of the m -adic symmetric group of degree n , and hence G the symmetric group itself. Hence, *every primitive m -adic substitution group of class 2 and degree n is the corresponding symmetric group, and conversely for $n > 1$.*

If G is a primitive m -adic substitution group of degree n and class 3, we have as before that G'_0 is primitive, while G_0 has a substitution of the form $t'_0 \cdot 1 \cdots 1$ with t'_0 of the form abc , the last since the substitution of class 3 in G_0 must consist of a single cycle of three letters which, in turn, must then belong to a single Γ . Defining \bar{G}'_0 as before, we see that \bar{G}'_0 is of class 3, and hence, by the corresponding extension of the standard result, is the alternating group of degree n . We therefore conclude that G_0 has, perhaps among others, every substitution of the form $t't'' \cdots t^{(m-1)}$ with the $t^{(i)}$'s positive substitutions. G_0 therefore has every possible substitution corresponding to the ϵ -sequence $(+1, +1, \dots, +1)$, and hence every possible substitution for each of the ϵ -sequences of its substitutions. It is therefore the associated group of an alternating group, i.e., G is an alternating group. Hence, *every primitive m -adic substitution group of degree n and of class 3 is an alternating group of degree n , and conversely for $n > 2$.*

Actually two cases arise as far as G'_0 is concerned. When the above found substitutions of G_0 are its only substitutions, $(+1, +1, \dots, +1)$ is its only ϵ -sequence, G is an alternating group whose δ -subgroup is of the first order, while G'_0 is identical with \bar{G}'_0 , and hence is itself the alternating group. Otherwise, G'_0 will be larger than \bar{G}'_0 , while containing it, and hence will be the symmetric group, while G will be an alternating group whose δ -subgroup is of order greater than one. Note also that in both of the above results the hypothesis of the primitivity of G was used only in deducing the primitivity of G'_0 . We therefore conclude that there does not exist an m -adic substitution group G of class 2 or 3 for which G is imprimitive, G'_0 primitive.

⁽⁶¹⁾ If not G'_0 itself.

⁽⁶²⁾ Actually improper, therefore.

Let now G be a primitive m -adic group of degree n and of class p , p a prime greater than 3. As before, the substitution of class p in G_0 must consist of a single cycle of letters which therefore belong to a single Γ . Hence $n \geq p$. Furthermore, \bar{G}'_0 will have a substitution of class p for element, and hence be of class p . Finally G'_0 is primitive, with \bar{G}'_0 as invariant subgroup. With the corresponding ordinary proof generalized as in the preceding cases, we then find that \bar{G}'_0 is $(n-p+1)$ -fold transitive. The remainder of the standard proof is then directly applicable to \bar{G}'_0 and shows that n cannot be greater than $p+2$. Hence, if a primitive m -adic substitution group is of class p , p being a prime number greater than 3, its degree can only be p , $p+1$ or $p+2$. Note that actually \bar{G}'_0 is then itself primitive—immediately so for $n=p$, and as a consequence of its being more than simply transitive for $n=p+1$ or $p+2$. Hence, in each of these cases, \bar{G}'_0 is the unique primitive ordinary group of class p and degree n .

We consider in detail only the case $n=p$. \bar{G}'_0 is then the group of order p , as is also each $\bar{G}_0^{(i)}$, defined in analogous fashion. Each $\bar{G}_0^{(i)}$ is therefore a cyclic group, and is, in fact, generated by a single cycle of the p letters of Γ_i . By relettering the members of the Γ 's we may therefore assume that $\bar{G}_0^{(i)}$ is generated by the substitution $t_0^{(i)} = (a_{1i} a_{2i} \cdots a_{pi})$. Now any substitution t of G_0 will transform each $\bar{G}_0^{(i)}$ into itself, and hence will transform $t_0^{(i)}$, the generator of $\bar{G}_0^{(i)}$, into some power $\nu^{(i)}$ of itself, with $\nu^{(i)} = 1, 2, \dots, p-1$. Hence, with each t in G_0 we can thus associate a ν -sequence $(\nu', \nu'', \dots, \nu^{(m-1)})$. Likewise, if s is any substitution in G , s will transform each $\bar{G}_0^{(i)}$ into $\bar{G}_0^{(i+1)}$. It will therefore transform each $t_0^{(i)}$ into some power $\mu^{(i)}$ of $t_0^{(i+1)}$, $\mu^{(i)} = 1, 2, \dots, p-1$. Hence, with each s in G we can thus associate a μ -sequence $[\mu', \mu'', \dots, \mu^{(m-1)}]$. Since G_0 has the substitution $1 \cdots t^{(i)} \cdots 1$ whenever $\bar{G}_0^{(i)}$ has the substitution $t^{(i)}$, we see that G_0 has the invariant subgroup

$$\bar{G}_0 = \bar{G}'_0 \bar{G}_0'' \cdots \bar{G}_0^{(m-1)},$$

the direct product of the $\bar{G}_0^{(i)}$'s, when it is not \bar{G}_0 itself. Now $\bar{G}_0^{(i)}$ consists of all the substitutions on the letters of Γ_i that transform $t_0^{(i)}$ into itself. It follows that \bar{G}_0 consists of all the $(m-1)$ -ads, consequently p^{m-1} in number, which transform each $t_0^{(i)}$ into itself. G_0 , therefore, has among its elements each of the p^{m-1} $(m-1)$ -ads with which we can associate the ν -sequence $(1, 1, \dots, 1)$. By expanding G_0 in cosets as regards \bar{G}_0 , we then easily verify that G_0 likewise has each of the $(m-1)$ -ads with which we can associate the ν -sequence of any one of its members, there being exactly p^{m-1} $(m-1)$ -ads for each ν -sequence. Likewise, by expanding G in cosets as regards \bar{G}_0 , which is invariant under G , we find that G has every m -adic substitution on the Γ 's with which we can associate the μ -sequence of any one of its members, there being exactly p^{m-1} such substitutions for each μ -sequence.

G is therefore determined by the set of μ -sequences of its members. Actually, if s_i in G has the μ -sequence $[\mu_i', \mu_i'', \dots, \mu_i^{(m-1)}]$, then $s = s_1 s_2 \cdots s_m$,

also in G , will have the μ -sequence $[\mu', \mu'', \dots, \mu^{(m-1)}]$ given by the equations

$$\begin{aligned}\mu' &= \mu_1' \mu_2' \cdots \mu_m', \\ \mu'' &= \mu_1'' \mu_2'' \cdots \mu_m'', \\ &\vdots \\ \mu^{(m-1)} &= \mu_1^{(m-1)} \mu_2^{(m-1)} \cdots \mu_m^{(m-1)},\end{aligned}$$

the products being reduced modulo p to one of the numbers $1, 2, \dots, p-1$. It follows that the μ -sequences of the members of G constitute an m -adic group under this m -adic operation isomorphic with G , and, indeed, simply isomorphic with the quotient group G/\bar{G}_0 . G is therefore determined by its corresponding "μ-subgroup."

This suggests a development analogous to that of the symmetric and alternating groups, and their relationship to the complete δ -group. Let P , the "symmetric power group of degree p ," be the m -adic substitution group of degree p consisting of all m -adic substitutions on the Γ 's that transform each $t_0^{(i)} = (a_{i1}a_{i2} \cdots a_{ip})$ into a power of $t_0^{(i+1)}$. Its associated group P_0 will then consist of all $(m-1)$ -ads on the Γ 's transforming each $t_0^{(i)}$ into a power of itself. Note that actually $\bar{P}_0^{(i)}$ is identical with the preceding $\bar{G}^{(i)}$, \bar{P}_0 with \bar{G}_0 , a fact that merely emphasizes the fact that for the class of groups under consideration these groups are independent of G . With each t in P_0 we can associate a ν -sequence, with each s in P , a μ -sequence. Clearly, for each of the $(p-1)^{m-1}$ possible ν -sequences there is at least one t in P_0 having that ν -sequence, and hence, as for G_0 , p^{m-1} such t 's. P_0 is therefore of order $[p(p-1)]^{m-1}$. Hence the symmetric power group P is also of order $[p(p-1)]^{m-1}$, having p^{m-1} substitutions for each of the $(p-1)^{m-1}$ possible μ -sequences. Actually, if s_0 is the m -adic substitution carrying each a_{ij} into $a_{(i+1)j}$, it will be in P with the μ -sequence $[1, 1, \dots, 1]$, so that we can write $P = P_0 s_0$; and if in $s = ts_0$ we let t run through all substitutions in P_0 with a given ν -sequence, s will run through all substitutions in P having that ν -sequence for μ -sequence.

Under the above m -adic operation the $(p-1)^{m-1}$ possible μ -sequences constitute an m -adic group which we may call the complete μ -group. As in the case of the complete δ -group, the associated group of the complete μ -group may be considered to have the $(p-1)^{m-1}$ ν -sequences as elements under the dyadic operation $(\nu'_1, \nu''_1, \dots, \nu^{(m-1)}_1)(\nu'_2, \nu''_2, \dots, \nu^{(m-1)}_2) = (\nu'_1 \nu'_2, \nu''_1 \nu''_2, \dots, \nu^{(m-1)}_1 \nu^{(m-1)}_2)$, the ν -sequence corresponding to an $(m-1)$ -ad of μ -sequences being given by the above operation on μ -sequences with the $\mu_m^{(i)}$'s omitted. The complete μ -group is therefore semi-abelian; and by the use of the μ -sequence $[1, 1, \dots, 1]$ the unique ν -sequence into which $(\nu', \nu'', \dots, \nu^{(m-1)})$ is transformed by every μ -sequence is again found to be $(\nu^{(m-1)}, \nu', \dots, \nu^{(m-2)})$.

Corresponding to each subgroup of the complete μ -group there will be an

"alternating power group," the m -adic substitution group of degree p consisting of all the m -adic substitutions on the Γ 's with μ -sequences in the μ -subgroup under consideration. Each of our G 's is therefore an alternating power group⁽⁶³⁾. To complete our investigation within its present scope we need merely find which of the alternating power groups are primitive groups of class p . Actually they are all primitive. For their \bar{G}'_0 is the primitive \bar{P}'_0 , so that their G'_0 is primitive. And their \bar{G}_0 is always \bar{P}_0 , which can carry any selection of letters chosen from the Γ 's into any other selection, so that in fact, they are cyclically transitive. As for their class, it is immediately seen to be at most p . Now actually a substitution on the letters of Γ , carrying $t_0^{(0)} = (a_{i1}a_{i2} \dots a_{ip})$ into a power of itself other than the first must be of class $p-1$. It follows that an alternating power group of degree p is of class less than p , in fact $p-1$, when and only when the associated group of its μ -subgroup has a ν -sequence with one and only one number not unity. This is easily transformed into a condition on the μ -subgroup itself to yield the following result. *The primitive groups of class p and degree p , p being a prime greater than 3, are the alternating power groups of degree p whose μ -subgroups do not have a pair of μ -sequences differing in one and only one component*⁽⁶⁴⁾.

B. FINITE ABSTRACT POLYADIC GROUPS

21. **Cyclic polyadic groups; ordinary theory**⁽⁶⁵⁾. Given the m -adic operation c , we define the m -adic powers of an element s under c inductively as follows. s itself will be rewritten $s^{[0]}$; and having $s^{[n]}$, we define $s^{[n+1]}$ as $c(s \dots ss^{[n]})$. If then $s^{[n]}$ be written out in full, n is the number of c 's occurring in the resulting extended operation, the number of s 's being $n(m-1)+1$. By the associative law it follows that any extended operation involving n c 's and but the single element s repeated can be rewritten in the form $s^{[n]}$. We thus easily obtain the following m -adic power laws:

$$c(s^{[n_1]}s^{[n_2]} \dots s^{[n_m]}) = s^{[n_1+n_2+\dots+n_m+1]}, \quad (s^{[n_1]})^{[n_2]} = s^{[(m-1)n_1 n_2 + n_1 + n_2]}.$$

Note that for $m=2$ our n th power is the ordinary $(n+1)$ -st power⁽⁶⁶⁾.

⁽⁶³⁾ Unless it were P itself. But P is readily seen to be of class $p-1$.

⁽⁶⁴⁾ The actual problem of determining the subgroups of the complete μ -group remains unsolved. Gill has pointed out to the writer that while the problem of determining the associated groups of these μ -subgroups can superficially be expressed as a problem in V.A.G.'s, actually the theory is now inapplicable, since the coefficients of the polynomials no longer form a field.

⁽⁶⁵⁾ For the special case $m=3$, the results of the present section reduce to those given by Lehmer. Likewise those of the next section involving mere reducibility, now of necessity to a 2-group.

⁽⁶⁶⁾ By contrast, Dörnte writes a^z in usual notation with, however, z subject to the restriction $z \equiv 1 \pmod{m-1}$. While our laws of powers are, as a result, more complicated than Dörnte's, we find great comfort in the fact that our $s^{[n]}$ is an " m -adic element" for every positive integral, or zero, n . Our lack of negative m -adic powers could easily be supplied.

If s is an element of an m -adic group K , each of its m -adic powers will represent elements of K . With K a finite group we therefore must have for some n_0 and n_0+n , $n > 0$, $s^{[n_0]} = s^{[n_0+n]}$. Since $s^{[n_0+n]}$ can be rewritten $c(s^{[n_0]}s \dots ss^{[n-1]})$, it follows that $\{s, \dots, s, s^{[n-1]}\}$ is an identity, whence we have

$$s^{[n]} = s.$$

The smallest positive integral value of n for which this equation holds will be called the (m -adic) *order* of s . If then s is of order g , the sequence of its m -adic powers $s^{[0]}, s^{[1]}, s^{[2]}, \dots$ starts with g distinct elements which are then repeated in order. It follows on the one hand that $s^{[n]} = s$ when and only when n is a multiple of g ; and, more generally, that $s^{[n_1]} = s^{[n_2]}$ when and only when $n_1 - n_2$ is a multiple of g . On the other hand, since but a finite number of elements are involved, our first law of m -adic powers shows that the g distinct elements constitute an abelian m -adic group G of order g which may then be called the *cyclic m -adic group generated by s* . The order of s is therefore equal to the order of the cyclic group it generates. Again by the first law of m -adic powers it is immediately seen that two cyclic m -groups of the same order are simply isomorphic. Furthermore, the same law shows that apart from an assumed m -group K , g distinct elements s_0, s_1, \dots, s_{g-1} , subject to the m -adic operation obtained by writing $s_n = s^{[n]}$, with $s^{[0]} = s$, constitute an m -group which is then the cyclic m -group of order g generated by $s = s_0$. Hence, as in ordinary group theory, we may say there is one and only one cyclic m -group whose order is an arbitrary natural number⁽⁶⁷⁾.

Let then G be the cyclic m -group of order g , s a generator of G . We first ask for the order of any power $s^{[n]}$ of s . This will be the least value of N for which $(s^{[n]})^{[N]} = s^{[n]}$, hence the least value of N for which $(m-1)nN + N + n - n = [(m-1)n+1]N$ is a multiple of g . It follows that *the order of $s^{[n]}$ is equal to the order of s divided by the highest common factor of $(m-1)n+1$ and the order of s* . In particular, the order of $s^{[n]}$ will be the same as the order of s when and only when $(m-1)n+1$ is prime to the order of s . Hence *an element s is generated by those and only those of its m -adic powers $s^{[n]}$ for which $(m-1)n+1$ is prime to the order of s* .

We can now determine what orders the elements of G can have. γ will be the order of an element of G if $\gamma = g/d$, $d = \text{H.C.F. } [(m-1)n+1, g]$ for some n . It is necessary then that d be a divisor of g , and prime to $m-1$. We now show that this is also sufficient. We have to find, then, an n and k such that $(m-1)n+1 = kd$, $g = \gamma d$ with k relatively prime to γ . Since $m-1$ is prime to d by hypothesis, for some $n = n_0$, $k = k_0$, we will have $(m-1)n_0+1 = k_0d$.

(67) The following discussion tacitly assumes that a symbol representing the order of an element or group is restricted to positive integral values, one representing an m -adic power to non-negative integral values. On the other hand, symbols entering into a diophantine equation may at first be allowed to assume arbitrary integral values which are then restricted in the above manner as the need arises.

The general solution of $(m-1)n+1=kd$ is then given by $n=n_0+\lambda d$, $k=k_0+\lambda(m-1)$ with arbitrary λ . Now the particular solution shows k_0 to be prime to $m-1$. Hence the arithmetic progression $k_0+\lambda(m-1)$ has, indeed, an infinite number of primes, and hence certainly a number prime to γ as was to be proved. We thus have the following result. *A cyclic m-group of order g has at least one element of every order γ such that γ is a divisor of g, and g/γ is prime to $m-1$, and no element of any other orders.* In particular, *a cyclic m-group of order g has a first order element when and only when g is prime to m-1.*

We can now generalize the ordinary cyclic group argument to prove the following. *A cyclic m-group of order g has one and only one subgroup whose order is any given divisor γ of g such that g/γ is prime to $m-1$, and no others.* The one subgroup is immediately yielded by the cyclic subgroup generated by an element of order γ , whose existence is insured by the preceding result. For the converse, consider any subgroup of the given cyclic group, and let its order be γ . By Lagrange's theorem extended, γ is a divisor of g . By the same theorem, each element s of the subgroup has an order which is a divisor of γ , and hence must satisfy the equation $s^{[\gamma]}=s$. Now consider all the elements of the given cyclic group, generated, say, by s_0 , that satisfy this equation. If $s=s_0^{[k]}$, we have, as in a preceding argument, that $\gamma[(m-1)n+1]=kg$ for some k , and hence, with $g/\gamma=d$, that $(m-1)n+1=kd$ —and conversely. We first see that d is prime to $m-1$, and hence that γ is the order of a cyclic subgroup of the given group. Furthermore, since $\gamma d=g$, our general solution $n=n_0+\lambda d$, $k=k_0+\lambda(m-1)$, of the equation $(m-1)n+1=kd$ shows that exactly γ such n 's are to be found with values in the set $0, 1, 2, \dots, g-1$. Hence the elements s satisfying the equation $s^{[\gamma]}=s$ are exactly γ in number, and consequently must be the γ elements of the above cyclic subgroup of order γ . Our assumed subgroup of order γ must therefore be that cyclic subgroup, whence our result.

From this proof flow a number of corollaries. We have immediately that *every subgroup of a cyclic polyadic group is cyclic.* Furthermore, our proof shows that an element of given order of a cyclic group is contained in those and only those subgroups of the cyclic group whose orders are multiples of the order of the element. It follows, on the one hand, that the necessary and sufficient condition that one element of a cyclic group generate a second is that the order of the first be a multiple of the order of the second. On the other hand, we see that two subgroups of a cyclic polyadic group intersect in the subgroup whose order is the highest common factor of the orders of the given subgroups, and generate the subgroup whose order is the least common multiple of those orders.

Apart from the possible orders of elements and subgroups of a cyclic polyadic group the above results are the same as for ordinary cyclic groups. Our condition on those possible orders γ can be transformed into the following more usable form. *Let g_0 be the largest divisor of g prime to $m-1$, and let*

$\gamma_0 = g/g_0$. Then the cyclic m -group of order g has at least one element, and exactly one subgroup, of those and only those orders γ for which $\gamma = \delta\gamma_0$, δ a divisor of g_0 . In fact, if γ is a divisor of g with g/γ prime to $m-1$ as per our original condition, g/γ , being a divisor of g prime to $m-1$, must be a divisor of g_0 . Hence $g_0 = \delta(g/\gamma)$ with δ a divisor of g_0 , whence $\gamma = \delta\gamma_0$. Conversely, if $\gamma = \delta\gamma_0$ with δ a divisor of g_0 , $g/\gamma = g_0/\delta$, so that γ is a divisor of g with g/γ prime to $m-1$.

We thus see that γ_0 is the least order of a subgroup of our cyclic group, with all of the subgroups of the cyclic group containing the unique subgroup of order γ_0 . At one extreme, when $\gamma_0 = 1$, which is equivalent to g prime to $m-1$, the cyclic group has a subgroup of first order. This corresponds to the element of first order previously noted, which is now seen to be unique. Every subgroup then contains this first order element, and their orders are the same as the orders of the subgroups of an ordinary cyclic group of order g . At the other extreme $\gamma_0 = g$, which is equivalent to every distinct prime factor of g being a factor of $m-1$. The cyclic group then has no (proper) subgroup, each of its elements being of order g , and thus generating the entire cyclic group⁽⁶⁸⁾. In particular, if g is a prime p , the corresponding cyclic group is always of one of these two special types. Note that unlike an ordinary group, a polyadic group whose order is a prime p need not be cyclic. By the extended Lagrange theorem its elements must be of order 1 or p . If it has an element of order p , it must be the cyclic group of order p . However all of its elements may be of order one, in which case it is noncyclic.

In the general case the orders of the subgroups are multiples of γ_0 , the multipliers being the orders of the subgroups of an ordinary cyclic group of order g_0 . Hence, if $g = p_1^{\alpha_1}p_2^{\alpha_2} \cdots p_r^{\alpha_r}q_1^{\beta_1}q_2^{\beta_2} \cdots q_s^{\beta_s}$, with p_1, p_2, \dots, p_r distinct primes not factors of $m-1$, q_1, q_2, \dots, q_s factors of $m-1$, then the number of subgroups of the cyclic m -group of order g is $(\alpha_1+1)(\alpha_2+1) \cdots (\alpha_r+1) - 1$.

We can now also find an expression for the number of elements of given order in a cyclic m -group. With that order one for which there is at least one element, the total number of elements of that order will be the same as the number of generators of a cyclic m -group of that order. We proceed therefore to find the number of generators of a cyclic m -group of order g generated, say, by s_0 . We first show that if $m-1$ is prime to g the number of generators is $\phi(g)$ as for ordinary cyclic groups. In fact, if we recall our formula for the order of $s_0^{[n]}$ we see that the number of generators in question is the number of numbers $(m-1)n+1$, $n=0, 1, \dots, g-1$, prime to g . But with $m-1$ prime to g this is the same as the number of numbers $0, 1, \dots, g-1$ prime to g , that is, $\phi(g)$. Now, in the general case, expand the given cyclic m -group of order g in cosets as regards its subgroup of order γ_0 . The resulting quotient group is then an m -group of order g_0 prime to $m-1$. Now let s be any element of the given group, σ the corresponding element of the quotient group. Then s

(68) Whence our correction of a statement of Miller.

is a generator of the given group when and only when σ is a generator of the quotient group. That σ generates the quotient group if s generates the given group is immediate. As for the converse, s will then generate a group having a complete set of multipliers for our coset expansion. But it must also generate all the elements of the subgroup of order γ_0 , and hence all the elements of the group. It follows, on the one hand, that the quotient group is itself cyclic, and hence has $\phi(g_0)$ generators, and hence, finally, that *the number of generators of a cyclic m-group of order g is $\gamma_0\phi(g_0)$* .

Among the few extensions of topics of the ordinary theory of cyclic groups omitted in the above development is that of the k th powers of elements of a cyclic group. We state the result for m -groups without further proof. *The distinct kth powers of a cyclic m-group of order g constitute a subgroup of order g/h where h is the highest common factor of g and $(m-1)k+1$; furthermore, each element of this subgroup is the kth power of exactly h elements of the given group.*

22. Cyclic polyadic groups; polyadic theory. We have observed that a cyclic m -group of order g has a first order element when and only when $m-1$ is prime to g . As this element, when it exists, is invariant under the group, it follows that *a cyclic m-group of order g is reducible to a 2-group when and only when g is prime to m-1*. We turn now to the general discussion of reducibility for cyclic polyadic groups. Our first result is immediate. *Every group to which a cyclic group is reducible is cyclic.* For if s is a generator of the given cyclic group, c its operation, c' the operation of the reduced group, every element of the given group is given by an extended c operation involving s 's only, hence also by an extended c' operation involving s 's only. s is therefore a generator of the reduced group, which is thus cyclic.

In applying our general criterion of reducibility to cyclic groups, questions of commutativity are automatically disposed of, since every cyclic group is abelian. A cyclic m -group will then be reducible to a μ -group, $m=k(\mu-1)+1$, if for some $(\mu-1)$ -ad, which may be written $\{s^{[n_1]}, s^{[n_2]}, \dots, s^{[n_{\mu-1}]}\}$, s being a generator of the cyclic group, the $(m-1)$ -ad $\{s^{[n_1]}, s^{[n_2]}, \dots, s^{[n_{\mu-1}]}, \dots, s^{[n_1]}, s^{[n_2]}, \dots, s^{[n_{\mu-1}]}\}$ is an identity of the cyclic group. Hence also if $s^{[k(n_1+n_2+\dots+n_{\mu-1})+1]}=s$, i.e., if $kn+1$ is a multiple of g , where g is the order of the group, $n=n_1+n_2+\dots+n_{\mu-1}$. It follows first that the cyclic m -group is reducible to a μ -group when and only when $k=(m-1)/(\mu-1)$ is prime to g . Furthermore, with k prime to g , if $kn'+1$ and $kn''+1$ are both multiples of g , then $k(n'-n'')$, and hence $n'-n''$, must be a multiple of g . It follows from our first law of m -adic powers that the $(\mu-1)$ -ads corresponding to n' and n'' are equivalent. Recalling our general theory of reducibility, we thus see that a cyclic m -group is reducible to but one μ -group for each admissible μ .

The least value of μ for which $(m-1)/(\mu-1)$ is prime to g corresponds to a k which is the largest divisor of $m-1$ prime to g . We thus obtain the following result. *The real dimension of a cyclic m-group of order g is equal to $(m-1)/k_0+1$ where k_0 is the largest divisor of $m-1$ prime to g.* In particular

a cyclic m -group of order g is irreducible when and only when each prime factor of $m-1$ is also a prime factor of g . Our previous uniqueness result easily enables us to complete the picture as far as mere reducibility is concerned. We thus see that the real dimension of a cyclic m -group of order g is its only irreducible dimension; and the groups to which the cyclic m -group is reducible, all cyclic, consist of a single group of dimension equal to the real dimension of the given group, and those of its extensions whose dimensions are of the form $k(\mu_0 - 1) + 1$, where μ_0 is the real dimension in question, k any proper divisor of $k_0 = (m-1)/(\mu_0 - 1)$.

Since the subgroups of a cyclic group are themselves cyclic, we can find their real dimensions by applying the above formula. γ will be the order of a subgroup of a cyclic m -group of order g if it is a divisor of g with g/γ prime to $m-1$. Writing $g = d\gamma$, with d prime to $m-1$, we see that the largest divisor of $m-1$ prime to γ is also the largest divisor of $m-1$ prime to g . Hence, all the subgroups of a cyclic polyadic group have the same real dimension, namely the real dimension of the group itself. It follows that a subgroup of a cyclic m -group is reducible to a μ -group when and only when the given group is reducible to a μ -group. In particular, all the subgroups of an irreducible cyclic group are irreducible. We now readily verify that the following simple situation holds. If a cyclic m -group be reduced to a μ -group, the subgroups of the m -group are thereby reduced to the subgroups of the μ -group⁽⁶⁹⁾. In fact, half of this situation obtains for arbitrary polyadic groups. For from the very definition of reducibility, if a polyadic group G is reduced to a polyadic group G' , the subgroups of G' are also subgroups of G , or more exactly, reductions of subgroups of G . Moreover, if G is abelian, every reduction of a subgroup of G can be effected by thus reducing G to some G' . For the satisfaction of our general criterion of reducibility by the subgroup then holds equally well for G , and the same operation that serves to reduce the subgroup is shown by the proof of that criterion to reduce G as well. If now G is cyclic and reducible to a μ -group, every subgroup of G is reducible to a μ -group; and since G can be reduced to but a single μ -group G' , that reduction must reduce all the subgroups of G to subgroups of G' , and hence by the first part of the proof to the subgroups of G' . Our proof incidentally shows that by varying μ every possible reduction of a subgroup of G will thus be obtainable. We furthermore have the following corollary which, indeed, can easily be proved directly, and itself used to give a different turn to our proofs. The polyadic orders of the elements of a cyclic polyadic group remain unchanged under every reduction of the group.

While cyclic groups form a closed set with respect to the two operations "subgroup of" and "reduction of," they do not form a closed set under the operation of extension, of which reduction is the inverse, and hence under the more general operation of derivation. We proceed to prove that a cyclic

⁽⁶⁹⁾ We prove this result independently of the discussion of complexes which concluded §5, since that discussion was extremely sketchy.

m-group of order g remains cyclic when extended to a μ -group, $\mu = k(m-1)+1$, when and only when k is prime to g . Let the polyadic n th powers of an element s in the two groups be written more explicitly $s^{[n]m}$ and $s^{[n]\mu}$. By counting c 's we then have immediately

$$s^{[n]\mu} = s^{[kn]m}.$$

Let s be a generator of the extended group, assuming that group to be cyclic. The elements of that group, and hence of the given group, will then be given by the μ -adic powers of s , and hence by those m -adic powers of s of the form $s^{[kn]m}$. For each N , therefore, there must be an n such that $s^{[N]m} = s^{[kn]m}$, and hence an n and v such that $N = kn + gv$. This will be so when and only when k is prime to g .

A group may therefore be reducible to a cyclic group without itself being cyclic. It will be convenient to have the phrase "reducible to a cyclic group" cover even the irreducible cyclic groups. The class of groups reducible to cyclic groups is therefore a wider class than the class of cyclic groups. While it is obviously closed under the operation "extension of," the situation has become obscured so far as the operations "reduction of" and "subgroup of" are concerned. It turns out that the following discussion of the corresponding associated groups clears up the entire situation.

We first reinterpret m -adic power and m -adic order in terms of the coset theorem. More generally, let s be an element of an arbitrary m -group K , K^* an arbitrary containing group of K . Then, in the notation of K^* , the m -adic n th power $s^{[n]}$ of s is the ordinary power of s , $s^{(m-1)n+1}$. The m -adic order of s is therefore the least positive integral value of n for which $s^{(m-1)n+1} = s$, i.e., for which $s^{(m-1)n} = 1$. It follows that the m -adic order g of s is identical with the ordinary order of s^{m-1} . As for the ordinary order of s as element of K^* , we can offhand merely say that it is a divisor of $(m-1)g$. If, however, K^* is of index $m-1$, in particular if it be the abstract containing group K^* of K , then $s^n = 1$ is possible only if N is a multiple of $m-1$, and hence the ordinary order of s will be exactly $(m-1)g$.

These observations are immediately applicable to our discussion of cyclic polyadic groups, and are in turn illuminated thereby. We first observe that *every containing group G^* of a cyclic m -group G is cyclic*. For if s is a generator of G , the elements of G being the m -adic powers of s are also ordinary powers of s in G^* . Hence the elements of G^* , being products of elements of G , are also ordinary powers of s , and G^* is an ordinary cyclic group generated by s . Note that if G^* is of index $m-1$, as is always the case when s is an element of K with K^* of index $m-1$, then the order of G^* is $m-1$ times the order of G and thus again the ordinary order of s , $m-1$ times its m -adic order.

Since the abstract containing group G^* of a cyclic m -group G is cyclic, it follows that its subgroup G_0 , the associated ordinary group of G , is cyclic. Indeed, our earlier result to the effect that the m -adic order of s is equal to the

ordinary order of s^{m-1} shows that an element s of an m -group G generates G when and only when s^{m-1} , then an element of G_0 , generates G_0 .

This result can be immediately generalized to the following. *The associated ordinary group of a group reducible to a cyclic polyadic group is cyclic.* For the abstract containing group of the cyclic polyadic group is a containing group of the given group. The abstract associated group of the cyclic polyadic group, cyclic by the preceding result, is therefore the associated group of the given group corresponding to the above containing group. But we have shown in §6 that all containing groups of a given polyadic group yield simply isomorphic associated groups.

We have seen that every containing group of a cyclic polyadic group is cyclic. While the last argument shows that some containing group of a group reducible to a cyclic polyadic group is cyclic, it is not true that every containing group of such a group is cyclic. In fact it is readily proved that if the abstract containing group of a polyadic group is cyclic, the polyadic group itself must be cyclic. Hence, while cyclic polyadic groups are characterized by the fact that their abstract containing groups are cyclic, we must seek elsewhere for a similarly definite characterization of groups reducible to cyclic polyadic groups.

This characterization cannot consist merely of the associated ordinary group of a polyadic group being cyclic; for the abelianism of cyclic polyadic groups makes every group reducible to a cyclic polyadic group abelian, while non-abelian polyadic groups exist whose associated ordinary groups are cyclic. The added hypothesis of abelianism is however sufficient. We proceed to prove the following result which will enable us to close the entire polyadic development of cyclic groups. *Every abelian polyadic group with cyclic associated ordinary group is reducible to a cyclic polyadic group.* Since the commutativity of two elements can be tested by any extended operation, it follows that an abelian group can be reducible only to an abelian group. Coupled with the previous observations on containing and associated groups, it follows that if an abelian group with cyclic associated group is reducible to a second group, the latter is also an abelian group with cyclic associated group. Our result will therefore have been proved if we show that every irreducible polyadic group of this type is in fact cyclic.

Let then G be an irreducible abelian m -adic group of order g with cyclic associated group G_0 . With s_0 a fixed element of G , t a generator of G_0 , the g elements of G may be written $s_0 t^n$, $n = 0, 1, 2, \dots, g-1$, in accordance with the coset theorem. The $(m-1)$ -ad s_0^{m-1} will itself be in G_0 . Let then $s_0^{m-1} = t^k$. Since G is abelian, its reducibility to a μ -group, with $m-1 = k(\mu-1)$, would be equivalent to the existence of a $(\mu-1)$ -ad $\{s_0 t^{i_1}, s_0 t^{i_2}, \dots, s_0 t^{i_{\mu-1}}\}$ such that $(s_0 t^{i_1} s_0 t^{i_2} \dots s_0 t^{i_{\mu-1}})^k = 1$, i.e., such that

$$k(i_1 + i_2 + \dots + i_{\mu-1}) + \kappa \equiv 0 \pmod{g} \quad (70),$$

(70) G being abelian, s_0 and t are commutative.

and hence to the H.C.F.(k, g)'s being a divisor of κ . It follows that the irreducibility of G is equivalent to the combined condition, each prime divisor of $m-1$ is a divisor of g , κ is prime to $m-1$. On the other hand, we have seen that an element s of G generates G when and only when s^{m-1} generates G_0 . With $s = s_0 t^\nu$, $s^{m-1} = t^{\kappa+(m-1)\nu}$. Since, for our irreducible G , κ is prime to $m-1$, the arithmetic progression

$$\kappa + (m-1)\nu, \quad \nu = 0, 1, 2, \dots,$$

will certainly include a value which is prime to g . With ν thus chosen, $t^{\kappa+(m-1)\nu}$, i.e., s^{m-1} , is a generator of the cyclic G_0 , and hence s of the consequently cyclic G .

We thus see that the class of polyadic groups reducible to cyclic polyadic groups is identical with the class of abelian polyadic groups with cyclic associated groups. The first formulation immediately showed this class of groups to be closed under the operation "extension of." The second formulation was already used to prove it closed under the operation "reduction of." It also easily shows the class to be closed under the operation "subgroup of." For such a subgroup must be abelian; while its associated group, being a subgroup of the associated group of the parent group, must be cyclic. Hence the class of polyadic groups reducible to cyclic polyadic groups is closed under the three operations "reduction of," "extension of," and "subgroup of."

In particular, the net of groups derivable from a cyclic polyadic group, or for that matter from a group reducible to a cyclic polyadic group, consists wholly of groups reducible to cyclic polyadic groups. The irreducible groups of the net are therefore all cyclic. Since we are dealing with abelian groups of finite order, the outer real dimension of these groups is 2. Hence the net of groups is in fact also derivable from an ordinary cyclic group. We proceed then to study the net of groups derivable from a cyclic 2-group of order g . Our general theory shows that for each $m \geq 2$ there will be g m -groups in the net, one for each class of equivalent $(m-1)$ -ads of the 2-group, said class serving as the class of identities of the m -group. In terms of the given 2-group, equivalent polyads are equivalent to a unique element of the group. Hence the groups of the net are determined in 1-1 fashion by letting m run through the values 2, 3, 4, ..., and s , the element of the 2-group equivalent to their identities, run through the g elements of that 2-group. By utilizing the expression for the operation of a polyadic group in terms of the operation of a group it is reducible to, and the fact that for any two groups of a net there is a third reducible to each, we find the following expression, in terms of the operation of the 2-group, for the operation c of an m -group of the net with identities equivalent to s :

$$c(s_1 s_2 \cdots s_m) = s_1 s_2 \cdots s_m s^{-1}.$$

By means of this formula we easily find which groups of the net are cyclic,

and hence also which are the irreducible groups of the net. While it also enables us to study in detail the relation of reducibility for the groups of the net, the resulting picture is quite complicated, and will not be entered into here.

Let s_0 be a generator of the cyclic 2-group, and let $s = s_0^\lambda$. If then s_0^ν be any element of the m -group of the net with identities equivalent to s , the above operation yields the following expression for the corresponding m -adic n th power of s_0^ν :

$$(s_0^\nu)^{[n]} = s_0^{n(m-1)\nu + \nu - n\lambda}.$$

We then easily find the condition under which, for some ν , s_0^ν is a generator of the m -group, and thus obtain the following result. *If s_0 is a generator of an ordinary cyclic group of order g , the cyclic groups of the net of groups derivable from the given group are those m -groups whose identities are equivalent to an s_0^λ for which $H.C.F.(m-1, \lambda, g) = 1$.* If γ is the order of s_0^λ in the 2-group, this condition is equivalent to g/γ prime to $m-1$. Thus all of the g m -groups of the net for given m are cyclic when and only when $m-1$ is prime to g . Since the irreducible groups of the net are the irreducible cyclic groups of the net, we see that *the irreducible groups of the net are those for which the prime divisors of $m-1$ are all divisors of g while λ is prime to $m-1$.* Hence, for $g \geq 2$, a cyclic polyadic group of order g has an infinite number of outer irreducible dimensions.

The full force of our closedness results for groups reducible to cyclic polyadic groups is brought out by the complexes obtained from such groups. We have then that the complex of groups obtainable from a cyclic polyadic group, or, in general, from a group reducible to a cyclic polyadic group, consists wholly of groups reducible to cyclic polyadic groups. We recall that the groups of any complex separate into mutually exclusive nets, there being a 1-1 correspondence between these nets and the different classes of elements the groups of the complex can have. In the present instance each net is of the type discussed above, being derivable from a group reducible to a cyclic polyadic group. Furthermore, these "group-bearing" classes now admit of very simple description. As most of the resulting picture holds good for arbitrary finite abelian polyadic groups we so present our development.

Observe first that our simplification of the operations yielding an arbitrary complex shows that its group-bearing classes, apart from that of the initial group, can all be obtained from the subgroups of the extensions of the initial group. Since a finite abelian polyadic group is always derivable from a 2-group, we may then assume that initial group to be a 2-group. That 2-group is then its own associated and containing group, and can be identified with the associated and containing group of each of its extensions. The relationship between the subgroups of a polyadic group and of its associated group, actually valid for an arbitrary containing group, then yields the following result. *The group-bearing classes of a complex obtained from a finite*

abelian polyadic group are, apart from the class of elements of the given group, the classes of elements of the subgroups of any 2-group derivable from the given group and the cosets of those subgroups.

We recall that the problem of the intersection of two subcomplexes of a complex was reduced to that of the intersection of their corresponding group-bearing classes. The above result then shows that, for finite abelian groups, either two group-bearing classes have no elements in common, or their common elements constitute an augmented coset of the crosscut of the subgroups of the 2-group of which they are augmented cosets. Note actually that the 2-groups derivable from the given finite abelian group are in 1-1 correspondence with the elements of the group, the element corresponding to a 2-group being the identity of the 2-group. If then s be any element of the given group, the group-bearing classes containing s constitute the subgroups of the 2-group having s as identity. It follows that *if two group-bearing classes of the complex obtained from a finite abelian group have a common element, they are the classes of elements of two subgroups of one and the same 2-group derivable from the given group, and intersect accordingly.*

In particular, then, for a cyclic polyadic group of order g the group-bearing classes of its complex are g/γ in number, of γ elements each, for every divisor γ of g . And two group-bearing classes, with γ_1 and γ_2 elements respectively, either have no elements in common, or exactly $H.C.F.(\gamma_1, \gamma_2)$ elements in common.

The above development can be given a somewhat different turn. For any finite polyadic group a finite number of extensions of the group, and subgroups of those extensions, suffice to yield all group-bearing classes, as these are now finite in number. From the corresponding situation for a pair of groups of a net it follows that for any finite number of groups of a net there is a group of the net itself reducible to each of the given groups. Hence the above extensions can themselves be extended to one and the same group. In this process the subgroups of these groups are extended to subgroups of the resulting group. Hence, *the group-bearing classes of the complex obtained from a finite polyadic group are the classes of elements of a single suitable extension of the group, and of the subgroups of that extension.* For any finite polyadic group, therefore, the intersection of two group-bearing classes can be pictured as the intersection of two subgroups of one and the same extension of that group. And now for the earlier picture. Clearly any element of finite order in a polyadic group is of first order in some extension of that group, and hence is the sole member of a group-bearing class of the complex obtained from that group. It follows that the elements of the above "suitable extension" of a finite polyadic group are all of first order. Hence that extension will itself be reducible to each of the 2-groups derivable from the given group. If, furthermore, the given group is abelian, each of its elements s will be the identity of a 2-group to which that extension is reducible, and the subgroups of that ex-

tension containing s will thereby be reduced to the subgroups of the 2-group—hence that first picture.

In conclusion, then, while the theory of cyclic groups requires for its completion the introduction of groups reducible to cyclic polyadic groups, the theory of these groups is entirely self-contained. While it would therefore be desirable to complete this theory by developing the properties of these groups, and we have at hand the instruments that would yield this development, we have perhaps already spent too much time on such very special developments, and so pass on to the more general topics of the theory.

23. Abstract polyadic groups of the first three orders. The concepts of the last two sections give a certain basis for distinguishing between polyadic groups. As in ordinary theory, in counting abstract polyadic groups no distinction will be made between groups that are simply isomorphic. By contrast, in the theory of reducibility such a distinction is imperative, for two groups on the same class of elements, but with different multiplication tables, must there be considered different even if simply isomorphic. Our present interest lies not only in the results to be obtained but in the illustrations of method thus afforded.

For each $m \geq 2$ there is of course the single abstract m -group of order one. Its sole element is of the first order, and hence the group is cyclic, and reducible to the cyclic 2-group whose sole element is the identity.

The abstract m -groups of order two can be determined directly from their possible multiplication tables⁽⁷¹⁾. If they are written on the abstract elements α and β , and c represents the m -adic operation, the value of $c(\alpha\alpha \cdots \alpha)$, that is, of $\alpha^{[1]}$, determines the table; for each change in the value of an argument must change the value of the result. Hence there are at most two abstract m -groups of order two. It further follows that $\alpha^{[1]}$ is, or is not, equal to $\beta^{[1]}$ according as m is even or odd. If m is even, then if $\alpha^{[1]} = \alpha$, $\beta^{[1]} = \alpha$, while if $\alpha^{[1]} = \beta$, $\beta^{[1]} = \beta$, and the two possible groups are changed into each other on interchanging α and β . On the other hand, if m is odd, if $\alpha^{[1]} = \alpha$, $\beta^{[1]} = \beta$, and if $\alpha^{[1]} = \beta$, $\beta^{[1]} = \alpha$, and the two groups cannot be simply isomorphic. These groups are then readily identified to yield the following result. When m is even, there is but one abstract m -group of order two, namely, the cyclic m -group of order two. It then consists of one first order element and one second order element, and is reducible to the ordinary cyclic group of order two, if it be not that group. When m is odd there are exactly two abstract m -groups of order two; one group consisting of two first order elements, and being the non-cyclic second order m -group reducible to the ordinary cyclic group of order two, the other group being the cyclic m -group of order two, consisting of two second order elements, and hence not reducible to a 2-group.

⁽⁷¹⁾ Dörnte used this method to determine the number of m -groups on two symbols as elements, but did not consider the question of those m -groups being abstractly the same.

To obtain the abstract m -groups G of order three, we employ the general coset theorem method of §8. The associated ordinary group G_0 must be cyclic, and hence its elements may be written $1, t, t^2$. If s_0 be a fixed element of G with $s_0^{m-1} = t_0, t_0$ in G_0 , we may assume that either (1) $t_0 = 1$, (2) $t_0 = t$; for were $t_0 = t^2$, groups simply isomorphic with those of case (2) would result. G_0 , furthermore, admits of but two automorphisms, i.e., (a) the identical automorphism, (b) the automorphism interchanging t and t^2 while, of course, leaving 1 invariant. With either of these automorphisms as the automorphism of G_0 under s_0 , and either of the two choices of t_0 , an m -group will be correspondingly determined provided (A) the automorphism carries t_0 into itself, (B) the $(m-1)$ -st power of the automorphism is the automorphism of G_0 under t_0 . Of the four cases thus to be considered (1) (a) and (2) (a) satisfy both (A) and (B) for all m 's, and hence always determine a corresponding m -group. (1) (b) satisfies (A) for all m 's, but (B) only for m odd; for if m be even, the $(m-1)$ -st power of the automorphism interchanges t and t^2 whereas $t_0 = 1$ leaves them unchanged. Hence (1) (b) determines an m -group when and only when m is odd. Finally, there is no polyadic group of order three corresponding to (2) (b), as (A) is then never satisfied.

We now identify and distinguish between the groups thus determined. The group (1) (a) is abelian since s and t are then commutative. Since G_0 is cyclic, G is therefore cyclic, or reducible to a cyclic group. Direct calculation then shows that if $m-1$ is a multiple of 3, each element is of first order, and hence the group is noncyclic, but reducible to the ordinary cyclic group of order three. On the other hand, when $m-1$ is not a multiple of 3 we find that while s_0 is of first order, ts_0 , and in fact t^2s_0 , are not, and hence must be of the third order. The group is therefore cyclic, but reducible to the ordinary cyclic group.

In the case of the group (2) (a), s_0 , not being of the first order, must be of the third order. The group is therefore cyclic. When $m-1$ is not a multiple of 3 it is therefore simply isomorphic with the group (1) (a). On the other hand, when $m-1$ is a multiple of 3, and hence not prime to $g=3$, the group contains no first order element. It is therefore not reducible to an ordinary group, and consists of three third order elements.

Finally group (1) (b), m odd, is non-abelian, since s_0 does not leave t invariant. Being therefore noncyclic, each of its elements is of the first order. We have already given this group with $m=3$ as an example of one with no invariant element. This property holds for each admissible m . In fact, since any two of the three elements must generate the whole non-abelian group, each element is invariant under no other element than itself. It follows that each element transforms a second element into the third, a property which by itself can be shown to determine the multiplication table of that third order m -group for odd m . It is needless to add that this group is not reducible to an ordinary group.

The third order abstract polyadic groups may then be tabulated as follows, the numbers in the parentheses being the orders of the elements.

$$\mu = 0, 1, 2, \dots$$

$m-1 =$	$6\mu+1$	$6\mu+2$	$6\mu+3$	$6\mu+4$	$6\mu+5$	$6\mu+6$
cyclic (3, 3, 1)	1	1		1	1	1
cyclic (3, 3, 3)			1			1
abelian (1, 1, 1)			1			1
non-abelian (1, 1, 1)		1		1		1
total	1	2	2	2	1	3

In particular, the one ordinary third order group comes under the case $m-1 = 6\mu+1$ with $\mu=0$. We further see that the smallest value of m for which there are three abstract third order groups is 7.⁽⁷²⁾.

24. **Properties of transforms.** The coset theorem enabled us to write the transform of an element s by a polyad r in the ordinary form $r^{-1}sr$. A fundamental m -group is of course tacitly presupposed. Since the m -adic n th power of an element can likewise be written as an ordinary $(m-1)n+1$ power, it follows that

$$(r^{-1}sr)^{[n]} = r^{-1}s^{[n]}r.$$

Hence, also, the m -adic order of an element is unchanged under transformation.

Suppose now that $r^{-1}sr = s^{[\alpha]}$. By raising both sides of this equation to the m -adic β th power we then have

$$r^{-1}s^{[\beta]}r = (s^{[\beta]})^{[\alpha]},$$

for our m -adic formula for the power of a power shows that $(s^{[\alpha]})^{[\beta]} = (s^{[\beta]})^{[\alpha]}$. Hence we have the following generalization of the corresponding ordinary theorem. *If a polyad transforms a generator of a cyclic m -group into its α th power, it transforms every element of this cyclic group into its α th power.*

Commutativity is related to transform through invariance. Given two noncommutative elements s_0 and s , we consider what m -adic powers, if any, of s are commutative with s_0 . If s is of m -adic order k , its ordinary order in the fundamental abstract containing group is $(m-1)k$. Let γ_0 be the least positive value of γ for which the ordinary power s^γ is commutative with s_0 . γ_0 is then a divisor of $(m-1)k$ and the distinct ordinary powers of s commutative with s_0 are $s^{n\gamma_0}$, $n=1, 2, \dots, (m-1)k/\gamma_0$. The m -adic powers $s^{[\beta]}$ commuta-

(72) The two third order m -groups falling under the case $m-1 = 6\mu+2$ have been given by Miller for $\mu=0$.

tive with s_0 are those for which $(m-1)\beta+1$ is a multiple of γ_0 . It follows first that there will be an m -adic power of s commutative with s_0 when and only when γ_0 is prime to $m-1$. γ_0 is then a divisor of k ; and if β_0 is the least value of β for which $s^{[\beta]}$ is commutative with s_0 , the m -adic powers of s commutative with s_0 are $s^{[\beta_0+n\gamma_0]}$, $n=0, 1, \dots, (k/\gamma_0-1)$. Actually these k/γ_0 m -adic powers of s commutative with s_0 must constitute a subgroup, necessarily cyclic, of the cyclic m -group generated by s . They are therefore the m -adic powers of some one of their number, not, however, necessarily of $s^{[\beta_0]}$ ⁽⁷³⁾.

If we form the successive transforms

$$s^{-1}s_0s = s_1, s^{-1}s_1s = s_2, \dots, s^{-1}s_{n-1}s = s_n, \dots,$$

the resulting elements are the transforms of s_0 under the various ordinary powers of s . In general, therefore, they will not all be gotten by transforming s_0 by the elements of the cyclic m -group generated by s , but by the elements of the abstract containing group of that cyclic m -group, or, what is the same thing, by the various polyads of the cyclic m -group.

This suggests that given any m -group G and element s_0 we consider the transforms of s_0 under the various polyads of G . These will then constitute a complete set of conjugates of s_0 under the abstract containing group G^* of G . The following discussion applies equally well to an m -group K taking the place of the element s_0 .

With s an element in G , G_0 the associated ordinary group of G , we have the expansion $G^* = G_0s + G_0s^2 + \dots + G_0s^{m-2} + G_0$, with $G_0s = G$. Since G_0 is an ordinary group, the number of transforms of s under the elements of G_0 is some divisor ν of g , the common order of G and G_0 . Each coset G_0s^i therefore transforms s_0 into ν distinct elements. If two cosets yield a common transform of s_0 , by writing those cosets in the form r_1G_0, r_2G_0 , r_1 and r_2 being elements of the cosets yielding that common transform, we see that the set of transforms yielded by one coset is identical with the set yielded by the other. The transforms of s_0 under G^* thus fall into a certain number κ of mutually exclusive classes of ν elements each. By a method entirely analogous to that used in the analysis of an arbitrary containing group, we easily find that κ is a divisor of $m-1$, and that the first κ cosets all yield distinct sets of ν transforms each, these being repeated in order by each succeeding set of κ cosets. We thus have the following theorem. *The number of transforms of an element under the polyads of an m -group of order g is of the form $\kappa\nu$, where ν is a divisor of g , κ a divisor of $m-1$. For each i the i -ads of the group yield ν distinct transforms. The $\kappa\nu$ transforms can be obtained from the i -ads with $i=1, 2, \dots, \kappa$; and these κ mutually exclusive sets of ν transforms each are cyclically repeated for i -ads with $i > \kappa$.*

We can now connect the theory of transforms with that of groups of sub-

(73) As may be shown by an example.

stitutions. For convenience set $\kappa = \mu - 1$, and let $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$ be the mutually exclusive sets of ν transforms each corresponding to $i = 1, 2, \dots, \mu - 1$ respectively. If s_i is any element of G , and s' is the transform of s_0 by an i -ad of G , $s_i^{-1}s's_i$ will be the transform of s_0 by an $(i+1)$ -ad of G . It follows that s_i transforms the members of each Γ_i in 1-1 fashion into the members of Γ_{i+1} . Thus each element of G determines a μ -adic substitution on $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$. Clearly, the product of m elements of G yields a μ -adic substitution which is the product of the μ -adic substitutions yielded by those elements. Certainly then, for our finite G the class of all μ -adic substitutions corresponding to elements of G constitutes an m -adic group of μ -adic substitutions isomorphic with G . It is readily seen that if N elements of G correspond to one μ -adic substitution, exactly N elements of G correspond to each μ -adic substitution, and the isomorphism is $(1, N)$. Finally, this (m, μ) substitution group is transitive. For if element s' of Γ_i is the transform of s_0 by the i -ad $\{s_{i_1}, s_{i_2}, \dots, s_{i_\mu}\}$ of G , the transforms of s' by the elements s_j of G are the transforms of s_0 by the $(i+1)$ -ads $\{s_{j_1}, s_{j_2}, \dots, s_{j_\mu}, s_j\}$ of G , hence by all $(i+1)$ -ads of G , and so constitute the whole class Γ_{i+1} .

When $\kappa = 1$, the (m, μ) substitution group becomes a transitive m -group of ordinary substitutions. The transforms of s_0 under the elements of G , now identical with the transforms of s_0 under the polyads of G , then include s_0 , and are such that each is transformed into the entire set by the elements of G . On the other hand, when $\kappa > 1$, the transforms of s_0 under the elements of G are transformed by the elements of G into an entirely different set. Nor can they then include s_0 ; for s_0 , being transformed into itself by that $(m-1)$ -ad of G which is the identity of G_0 , appears for the first time in the set of transforms for which $i = \kappa$. We thus see that the transforms of s_0 under the elements of G must be said to constitute a complete set of conjugates of s_0 under G when and only when $\kappa = 1$. And the fact that then and only then is s_0 included in that set of transforms needs only restatement to become the following useful criterion. *The necessary and sufficient condition that the transforms of an element s_0 by the elements of an m -group G constitute a complete set of conjugates under G is that s_0 is commutative with some element of G .* As in the case of ordinary groups, the elements of G thus leaving s_0 invariant constitute a subgroup H of G . If G is expanded in right cosets as regards H , each coset consists of all the elements of G transforming s_0 into some one element. Hence, here too the number of conjugates of s_0 under G is the order of G divided by the order of the largest subgroup of G leaving s_0 invariant.

If s_0 is actually an element of G , the above condition is automatically satisfied with s_0 itself as element commutative with s_0 . We thus have the significant fact that the transforms of an element of a polyadic group under the elements of the group always constitute a complete set of conjugates under the group⁽⁷⁴⁾. Hence, as for ordinary groups, *all the elements of an m -group G*

⁽⁷⁴⁾ Essentially a result of Miller's when stated for "perfect cosets."

can be separated into distinct complete sets of conjugates as regards G , and this separation can be performed in only one manner.

In the case of an i -ad of G with $i > 1$ the transforms of the i -ad by the elements of G need no longer constitute a complete set of conjugates. Thus in the non-abelian 3-group of order three a dyad not the identity has but one transform under the elements of the group, two under the polyads of the group. However, our general theorem holds in this case; and since the i -ad is invariant under itself, it readily follows that κ is a divisor of $H.C.F.(i, m - 1)$.

25. Generation of polyadic groups by two groups, one invariant under the elements of the other. We shall consider two distinct cases. In the first, a 2-group H_0 is invariant under each element of an m -group K , in the second, an m -group H is invariant under each element of an m -group K . The discussion of the m -group G generated by the two given abstract groups can also be carried through from two different points of view, the first, that of the investigation of properties of groups assumed given, the second, that of the construction of groups hitherto unknown.

We have already illustrated the constructional point of view in §8. Our present interests being largely theoretical, we shall not further pursue the complexities introduced by that point of view in the field of abstract group theory, but merely obtain the results given by the first point of view⁽⁷⁵⁾.

(75) This is the point of view really followed by Miller in §25, *Finite Groups*, despite the section heading "Construction of Groups with Invariant Subgroups." He thus obtains the theorem: "If all the elements of a group H transform G into itself, then H and G generate a group whose order is the order of G multiplied by the index under H of the crosscut of G and H ." The constructional point of view, while using his treatment for purposes of analysis, would necessitate the following complications. H and G would be given by group-satisfying multiplication tables on specified symbols as elements. These tables must then satisfy the consistency condition that H and G have at least one element in common, and that the product of two elements common to H and G is the same in H as in G . With each element of H there would be given a corresponding automorphism of G which is to be the automorphism of G induced by transforming it by that element of H . These automorphisms must then satisfy the consistency conditions that the product of the automorphisms corresponding to two elements of H is the automorphism corresponding to the product of those elements, while the automorphism corresponding to any element of H common to H and G is the automorphism of G induced by that element as element of G . That posited, our guess is that H and G , assumed finite, will generate a unique group in the sense that there exists a group K which, with respect to itself as fundamental group, is the group generated by H and G , while all such groups are simply isomorphic; a simple isomorphism being in fact determined by letting each element of H and G correspond to itself.

The above criticism assumes that we are dealing with abstract groups, the title of the chapter in which the above section appears. If the generating groups be given as substitution groups, for example, the divergence between the two points of view disappears, as there is always the symmetric group on all the letters involved to act as fundamental group. A similar situation obtains for m -adic groups of ordinary substitutions as is shown in the last footnote of our present section.

It should be pointed out that what we have termed the constructional point of view is followed in the related theory of group extensions. (See the first footnote to §8.) That it is but

This point of view assumes a given m -group F . In the first of our two cases, H_0 is a subgroup of the associated ordinary group F_0 of F which is invariant under each element of a subgroup K of F . It is convenient here to consider F a subgroup of itself. The crosscut of all subgroups of F which are such that H_0 is a subgroup of their associated groups, K of themselves, is itself one of these subgroups, and will be said to be the m -group G generated by H_0 and K . We may then also say that the m -group G generated by H_0 and K is the smallest subgroup G of F such that H_0 is a subgroup of G_0 , K of G . Similarly, if H and K are two subgroups of F with H invariant under each element of K , the m -group G generated by H and K is the smallest subgroup G of F such that H and K are subgroups of G . The existence and uniqueness of G is thus assured, but is entirely relative to the given m -group F .

We shall first consider the subcase of the general H_0 , K case where K is the cyclic m -group generated by an element s of F . The m -group generated by H_0 and K may then also be said to be generated by H_0 and s . The invariance condition now reduces to H_0 being transformed into itself by s . Consider the cosets H_0s , $H_0s^{[1]}$, $H_0s^{[2]}$, \dots . If γ is the m -adic order of s , $H_0s = H_0s^{[\gamma]}$. Let then κ be the smallest positive integer for which $H_0s = H_0s^{[\kappa]}$. It then easily follows that the cosets H_0s , $H_0s^{[1]}$, \dots , $H_0s^{[\kappa-1]}$ are mutually exclusive, while succeeding cosets are cyclic reproductions of these. Hence, also, κ is a divisor of γ .

We now readily show that the m -group G generated by H_0 and s is given by

$$G = H_0s + H_0s^{[1]} + \dots + H_0s^{[\kappa-1]}.$$

Since H_0 is a subgroup of F_0 , s an element of F , the set G thus defined is contained in F . Furthermore, the invariance of H_0 under s coupled with the above coset analysis shows that the product of m elements of G is in G . Hence G is, indeed, a subgroup of F . G has s for element, in fact, as a member of H_0s . Hence $G_0 = Gs^{-1}$ has H_0 as subgroup. Finally, as with F , every subgroup of F whose associated group has H_0 as subgroup, while it has s as element, contains G . Hence G is the m -group generated by H_0 and s . We thus have the theorem: *If s is an element of an m -group, H_0 a subgroup of the associated group of that m -group invariant under s , then if $s^{[\kappa]}$ is the smallest positive m -adic power of s which is in the coset H_0s , H_0 and s generate an m -group whose order is κ times the order of H_0 .*

In the general H_0 , K case let L_0 be the crosscut of H_0 and K_0 . Since H_0 and K_0 are invariant under each element of K , the same is true of L_0 . Expand K in cosets as regards L_0 , and let $s_1, s_2, \dots, s_\kappa$ be a corresponding set of multi-

a related theory may be seen from the definition of an extension K of G by H as a group having G as invariant subgroup, with the quotient group K/G simply isomorphic to H . The above complication arising from the common elements of H and G goes not then arise.

pliers. We then show that the m -group G generated by H_0 and K has the expansion

$$G = H_0s_1 + H_0s_2 + \cdots + H_0s_k$$

with all indicated elements distinct. As a consequence of the invariance of H_0 under each s_i we can reduce the product of m elements of the set G thus defined to the form ts , with t in H_0 , s in K . As s can further be written $t's_i$ with t' in L_0 , and hence in H_0 , s_i one of the above k multipliers, we see that the product in question is in G . It then follows as in the special case that G is the m -group generated by H_0 and K . Moreover, suppose that with t_1 and t_2 in H_0 we have $t_1s_{i_1} = t_2s_{i_2}$. Then $t_2^{-1}t_1 = s_{i_2}s_{i_1}^{-1}$. Since the left side of this equation represents an element of H_0 , the right of K_0 , this one element τ is in L_0 . But then $s_{i_2} = \tau s_{i_1}$, contradicting the assumption that s_1, s_2, \dots, s_k were the set of multipliers in question. The indicated elements of G are thus distinct, and we have the theorem: *If K is a subgroup of an m -group, H_0 a subgroup of the associated group of the m -group invariant under each element of K , then H_0 and K generate an m -group whose order is the order of H_0 multiplied by the index under K of the crosscut of H_0 and the associated ordinary group K_0 of K .*

We turn now to the more interesting case of the m -group G generated by two m -groups H and K , with H invariant under each element of K . Note that H_0 , the associated ordinary group of H , is then also invariant under K . It is readily seen that while the m -group generated by H_0 and K is contained in the m -group generated by H and K , it will be identical with that m -group when and only when it contains an element of H . This means that for some t in H_0 , s in K , s' in H , $ts = s'$. But this is equivalent to $s = t^{-1}s'$, i.e., to s 's also being in H . Hence *the m -group generated by H and K is identical with the m -group generated by H_0 and K when and only when H and K have a common element.*

In particular, if H and K have but one common element, while each element of H is commutative with each element of K , we shall say that the m -group G generated by H and K is their *direct product*. G , then, is also the m -group generated by H_0 and K , and by K_0 and H , and correspondingly has expansions which may be briefly written $G = H_0 \times K = K_0 \times H$. For L_0 now reduces to the identity, so that in the first case, for example, the multipliers s_i are all the elements of K . More symmetrically then, $G = (H_0 \times K_0)s$, with s , say, the unique common element of H and K . It follows that $G_0 = H_0 \times K_0$ is the ordinary direct product of H_0 and K_0 . Clearly s must be of first order, since all of its m -adic powers must be common elements of H and K ; and being invariant under H and K , it must also be invariant under G . All three groups are therefore reducible to ordinary groups, and simultaneously so. These considerations immediately extend to the "direct product" of any number of m -groups provided the unique common element is also the only element common to each group and the group generated by the remaining groups.

Special as this concept of direct product thus turns out to be, it is very useful in the theory of abstract polyadic groups. By contrast, the direct product method as applied to m -adic substitution groups, while involving no restriction on the groups per se, did not yield the m -group generated by the given m -groups, and hence is restricted in its usefulness to the construction of desired m -groups.

In the most general H, K case consider the abstract containing groups of the m -groups involved. Since H, K , and G are subgroups of the fundamental m -group F , their abstract containing groups H^* , K^* , and G^* may be considered subgroups of the abstract containing group F^* of F . As the elements of an m -group generate the corresponding containing group, it easily follows that G^* is the ordinary group generated by H^* and K^* . Since H^* will be invariant under each element of K^* , the standard theorem tells us that the order of G^* is equal to the order of H^* multiplied by the index under K^* of the crosscut of H^* and K^* . But the order of the abstract containing group of an m -group is $m-1$ times the order of the m -group. Hence the order of G is equal to the order of H multiplied by that index.

It is easy indeed to write the actual expansion of G . According to the standard theory, if we expand K^* in cosets as regards the crosscut of H^* and K^* , and let r_1, r_2, \dots, r_n be the corresponding set of multipliers, then $G^* = H^*r_1 + H^*r_2 + \dots + H^*r_n$ with all indicated elements distinct. If then r_j is an i_j -ad of K , the elements of H^*r_j in G will be the $(m-i_j)$ -ads of H multiplied by r_j . We may therefore write, in notation thus suggested,

$$G = (H)_{m-i_1}r_1 + (H)_{m-i_2}r_2 + \dots + (H)_{m-i_n}r_n.$$

Returning to the order of G , we seek a useful expression for the index in question. The crosscut \bar{L} of H^* and K^* will consist of the common i -ads of H and K for $i=1, 2, \dots, m-1$. For $i=m-1$, these common i -ads constitute the crosscut L_0 of H_0 and K_0 . Let l be the order of L_0 , κ the smallest value of i for which H and K have a common i -ad. Then, by methods already made familiar, we find that κ is a divisor of $m-1$, while \bar{L} consists of l i -ads for each of the $(m-1)/\kappa i$'s, $i=\kappa, 2\kappa, \dots, m-1$. The order of \bar{L} is thus $l(m-1)/\kappa$. If now k is the order of K_0 , the order of K^* is $(m-1)k$. Hence the index under K^* of \bar{L} is $\kappa k/l$. But k/l is the index under K_0 of L_0 . Hence the index of \bar{L} under K^* is κ times the index of L_0 under K_0 . We therefore have the following theorem. *If H and K are two subgroups of an m -group such that all the elements of K transform H into itself, then H and K generate an m -group whose order is the order of H multiplied by the index under K_0 of the crosscut of H_0 and K_0 multiplied by a divisor κ of $m-1$, where κ is the smallest value of i for which H and K have a common i -ad.*

Of special interest is the case where K is the cyclic m -group generated by an element s which transforms H into itself. K^* is then an ordinary cyclic group also generated by s . Hence if s^λ is the smallest positive ordinary power

of s in H^* , λ will be the index under K^* of the crosscut of H^* and K^* . We thus have as our first result: *If an element s of an m -group transforms a subgroup H of that m -group into itself, and if s^λ is the smallest positive ordinary power of s in the containing group H^* of H , then s and H generate an m -group G whose order is λ times the order of H .* Indeed, the expansion of G is now readily seen to be

$$G = (H)_{m-1}s + (H)_{m-2}s^2 + \cdots + (H)_{m-\lambda}s^\lambda.$$

Note that if k is the m -adic order of s , and hence $k(m-1)$ its ordinary order, λ is a divisor of $k(m-1)$. Since the order of G must exceed the order of H whenever s is not in H , we obtain the following useful corollary further generalized below. *If s is of m -adic order one, and not in H , then the order of G is equal to the order of H multiplied by a divisor, not unity, of $m-1$.*

More refined results are yielded by our earlier analysis of the above mentioned index. The associated group K_0 of the cyclic m -group K generated by s will be the cyclic ordinary group generated by s^{m-1} . Hence, if $s^{\nu(m-1)}$ is the smallest positive power of s^{m-1} in H_0 , ν will be the index under K_0 of the crosscut of H_0 and K_0 . Consequently, *the order of G is also equal to the order of H times ν times κ , where $s^{\nu(m-1)}$ is the smallest positive power of s^{m-1} in H_0 , κ the smallest value of i for which H and the cyclic m -group generated by s have a common i -ad.*

We may note certain relationships between the constants thus involved. The connecting link between our two expressions for the order of G is the equation $\lambda = \nu\kappa$. λ is thus determined by ν and κ . Conversely ν and κ are determined by λ and m . For the common elements of H^* and K^* are $s^\lambda, s^{2\lambda}, \dots, s^{k(m-1)}$. It therefore easily follows that $\kappa = \text{H.C.F.}(m-1, \lambda)$, and hence $\nu = \lambda/\text{H.C.F.}(m-1, \lambda)$. By means of m -adic groups of ordinary substitutions it is readily shown that λ and m may assume arbitrary values. In the case of κ , ν , and m , we have already observed that κ is a divisor of $m-1$. Our expressions for κ and ν in terms of λ and m further show that $(m-1)/\kappa$ is prime to ν . Now it is readily verified that if κ , ν , and m are arbitrarily chosen subject to these two conditions, then $\lambda = \nu\kappa$ redetermines the same κ and ν by means of the above formulas. It follows that κ , ν , and m may assume any values subject to these conditions. If we now further introduce the m -adic orders h and k of H and s , we obtain the further conditions ν a divisor of k , $h\nu$ a multiple of k ; the first from the index interpretation of ν , the second from the order requirement imposed by $s^{\nu(m-1)}$'s being in H_0 . We have not carried the investigation far enough, however, to see whether the resulting four necessary conditions on h , k , κ , ν and m suffice to insure a corresponding H and s ⁽⁷⁶⁾.

(76) When $h=k$, the fourth condition is automatically satisfied. In this case the writer has verified by an example that $h=k$, κ , ν , and m may have arbitrary values subject to the first three conditions.

In constructing such examples by means of m -adic groups of ordinary substitutions, we

H is clearly an invariant subgroup of the generated group G . If then σ is the element of the m -adic quotient group G/H corresponding to s , ν is seen to be the m -adic order of σ . For the least positive ν with $s^{\nu(m-1)}$ in H_0 is the least positive ν with $s^{[\nu]}$ in H_0s , and hence the least positive ν with $\sigma^{[\nu]} = \sigma$. By a simple result of our later §29 the m -adic order of σ is a divisor of the m -adic order of s . Our previous corollary thus generalizes to the following.

The order of G is equal to the order of H multiplied by a multiple of a divisor other than unity of the m -adic order of s whenever the m -adic order of the element of G/H corresponding to s is not unity; when the latter order is unity, and yet s

are naturally led to the ordinary groups they generate as containing groups. On the other hand, our theory concerns their abstract containing groups only. In the λ, m example referred to above, it was possible to avoid this difficulty by so choosing H, K , and the fundamental F that their concrete containing groups were all of index $m-1$, and so simply isomorphic with their abstract containing groups. On the other hand, especially in the case of F , it is desirable to dispense with this requirement. For we could then fully make use of the fact that as for ordinary substitution groups, so for m -adic substitution groups, a fundamental F is always at hand, namely, the extension to an m -group of the ordinary symmetric group on all the letters involved; and clearly all fundamental F 's which are m -adic groups of ordinary substitutions yield the same G .

Actually, we can easily obtain the desired information concerning the abstract containing groups, and so the order of G , from any containing groups. We shall consider our general H, K case. Let F be a corresponding fundamental m -group, $F^{*''}$ any containing group of F . The subgroups of $F^{*''}$ generated by the elements of H and K respectively will then be containing groups $H^{*''}$ and $K^{*''}$ of H and K . In the above case of m -adic groups of ordinary substitutions, $F^{*''}$ may be the ordinary substitution group generated by the substitutions of F , in which case $H^{*''}$ and $K^{*''}$ will be the ordinary substitution groups generated by the substitutions of H and K , and so obtainable without the explicit use of $F^{*''}$. Let $H^{*''}, K^{*''}, F^{*''}$ be of indices ρ_1, ρ_2, ρ . All three indices will then be divisors of $m-1$. Furthermore, it is readily seen that ρ_1 and ρ_2 will be multiples of ρ . Now if the cosets into which these containing groups are broken up are cyclically repeated until there are $m-1$ of each, the i th cosets of $H^{*''}$ and $K^{*''}$ will be contained in the i th coset of $F^{*''}$ for $i=1, 2, \dots, m-1$. In particular, the $(m-1)$ -st cosets will be the associated ordinary groups H_0', K_0', F_0' . And in the simple isomorphism between F_0' and F_0 , the abstract associated ordinary group of F , the subgroups H_0' and K_0' will correspond to H_0 and K_0 . Hence, the index under K_0 of the crosscut of H_0 and K_0 is also the index under K_0' of the crosscut of H_0' and K_0' , where the latter may now be considered the ρ_1 th and ρ_2 th cosets in $H^{*''}$ and $K^{*''}$. As for κ , note that two products of i elements each taken from an m -group will be identical in a containing group of an m -group when and only when those two i -ads of elements are equivalent. Hence, the smallest value of i for which H^* and K^* have a common i -ad is also the smallest value of i for which $H^{*''}$ and $K^{*''}$, repeated as above, have a common i -ad. Actually, the pairs of i th cosets of $H^{*''}$ and $K^{*''}$ start repeating after $i=L.C.M.(\rho_1, \rho_2)$. Hence, κ may be found from $H^{*''}$ and $K^{*''}$ if their cosets be cyclically repeated to a total number equal to L.C.M. (ρ_1, ρ_2) each. Clearly κ is a divisor of L.C.M. (ρ_1, ρ_2) . If desired, it is not difficult to give a number theoretic expression for κ in terms of the distribution of (i, j) 's for which an i -ad of $H^{*''}$ and a j -ad of $K^{*''}$ in their unrepeated form are identical.

The order of G is thus determinable from $H^{*''}$ and $K^{*''}$. Explicitly $F^{*''}$ does not enter. Hence, in the case of H and K m -adic groups of ordinary substitutions, no further reference need be made to F . In particular, then, if $H^{*''}$ and $K^{*''}$ are each of index $m-1$, the order of G is found exactly as if they were the abstract containing groups of H and K .

is not in H , then the order of G is equal to the order of H multiplied by a divisor, not unity, of $m-1$.

26. *m -adic groups of order g prime to $m-1$.* Let G be any m -group whose order g is prime to $m-1$. The order of any element s of G , being a divisor of g , will then also be prime to $m-1$. The cyclic m -group generated by s therefore has one and only one first order element s_0 , i.e., s generates one and only one first order element s_0 . G therefore has at least one first order element; and if it has exactly λ first order elements, all of its elements can be separated into λ corresponding mutually exclusive classes of elements, each class consisting of all the elements of G which separately generate the corresponding first order element. Now no first order element of G can transform another first order element of G into itself. For otherwise, by the first of the two corollaries of the last section, the two would generate a subgroup of G whose order would be a divisor, not unity, of $m-1$. But, as in the case of an element of G , the order of any subgroup of G must be prime to $m-1$. It follows that if element s of G generates the first order element s_0 , and hence transforms s_0 into itself, it can transform no other first order element of G into itself; for otherwise s_0 , a power of s , would transform that other first order element into itself. The class of elements of G each generating s_0 therefore consists of all the elements of G which transform s_0 into itself, and hence constitute a subgroup of G . As this subgroup has s_0 for invariant first order element, it is reducible to an ordinary group. We have thus proved that *if G is an m -group whose order is prime to $m-1$, the elements of G can be separated into a number λ of mutually exclusive subgroups of G , all reducible to ordinary groups, where λ is the number of first order elements of G , and each subgroup contains one and only one first order element of G , and, indeed, consists of all the elements of G that transform that first order element into itself.*

Other immediate consequences of the above proof are the following. *G is reducible to a 2-group when and only when it has but a single first order element. If G has more than one first order element, it has no invariant element, and hence is not derivable from a 2-group.* In particular, *every abelian m -group whose order is prime to $m-1$ has one and only one first order element, and hence is reducible to a 2-group⁽⁷⁷⁾.*

We may note in passing the marked simplicity, from the standpoint of polyadic theory, of those m -groups of order prime to $m-1$ which are reducible to 2-groups. As seen below, the one first order element of such an m -group is also the one and only first order element of each of its subgroups. These subgroups are therefore also reducible to 2-groups. Furthermore, both the group and its subgroups are reducible to 2-groups in one and only one way. It easily follows by a slight modification of our cyclic m -group argument that when the above m -group is reduced to the 2-group, its subgroups are reduced to the subgroups of that 2-group.

⁽⁷⁷⁾ This generalizes a theorem of Lehmer on abelian 3-groups.

Returning to our arbitrary m -group G of order g prime to $m - 1$, we proceed to show that the λ first order elements of G , as well as the corresponding λ subgroups into which G was decomposed, constitute a complete set of conjugates under G . It will then follow that these λ subgroups are all of the same order, and hence that the number of first order elements of G is a divisor of the order of G ⁽⁷⁸⁾. Let $s'_0, s''_0, \dots, s^{(\lambda)}_0$ be the first order elements of G , $k_1, k_2, \dots, k_\lambda$ the orders of the corresponding λ subgroups of G . Since exactly k_i elements of G transform $s^{(i)}_0$ into itself, $s^{(i)}_0$ is transformed into g/k_i different elements by all the elements of G . As the transform of a first order element is also of the first order, $g/k_i \leq \lambda$, i.e., $k_i \geq g/\lambda$. Since $g = k_1 + k_2 + \dots + k_\lambda$, it follows that the equality sign must hold for each i . Each $s^{(i)}_0$ therefore has the λ first order elements of G for its different transforms under the elements of G , whence the first half of our theorem. Now if s_1 generates the first order element $s^{(i)}_0$, say $s_1^{[n]} = s^{(i)}_0$, then $(s^{-1}s_1s)^{[n]} = s^{-1}s_1^{[n]}s = s^{-1}s^{(i)}_0s$; that is, the transform of s_1 under s generates that first order element which is the transform of $s^{(i)}_0$ under s . Hence, if element s of G transforms $s^{(i)}_0$ into $s^{(j)}_0$, it transforms the subgroup corresponding to $s^{(i)}_0$ into the subgroup corresponding to $s^{(j)}_0$, whence the rest of our result.

It follows from the above that the λ first order elements of G also constitute a complete set of conjugates under the m -group they generate. For that m -group will have an order prime to $m - 1$, while its first order elements will be the λ first order elements of G . Since the m -group generated by a given set of elements chosen from a finite m -group will actually consist of all extended products of elements chosen from the set, it follows that the λ first order elements of G constitute a "generalized" complete set of conjugates under themselves, that is, each can be obtained from any other by a succession of transforms by first order elements only. Actually, this statement is weaker than the one immediately preceding, since it amounts to saying that the λ first order elements of G constitute a complete set of conjugates under any containing group of the m -group they generate. In any case, the question whether they constitute a complete set of conjugates under themselves, in the sense that any one can be transformed into any other by a third, is left open⁽⁷⁹⁾.

We have already observed that λ is a divisor of g . While it is therefore prime to $m - 1$, we now find an additional restriction imposed upon it by $m - 1$. The first order element s'_0 is of course invariant under itself. On the

(78) For, if k is the common order of these λ subgroups, $g = k\lambda$. That the number of first order elements of an arbitrary m -group need not be a divisor of its order is illustrated by the ordinary symmetric group of degree three extended to a 3-group. This 3-group of order six has four first order elements.

(79) Note that the statement of Miller, page 30 of *Finite Groups*, to the effect that the Sylow subgroups of order p^3 of a group constitute a complete set of conjugates under themselves must also be interpreted in the above sense of a generalized complete set of conjugates. At least, that is all the proof there given allows us to infer.

other hand, since any other first order element $s_0^{(t)}$ is not invariant under s_0' , it will be transformed by the polyads $\{s_0'\}$, $\{s_0', s_0'\}$, $\{s_0', s_0', s_0'\}$, ... into a number, not unity, of first order elements which either directly, or by our general theorem on transforms, is seen to be a divisor of $m-1$. Since the sets of transforms of different $s_0^{(t)}$'s by the above polyads are mutually exclusive when not identical, a separation of the λ first order elements into mutually exclusive classes is thus effected, one class consisting of but one element, every other class of a number of elements which is a divisor, not unity, of $m-1$. Hence, if p_1, p_2, \dots, p_v are the distinct prime divisors of $m-1$, λ is of the form $\lambda = 1 + k_1 p_1 + k_2 p_2 + \dots + k_v p_v$, $k_i \geq 0$. In particular, if $m-1$ is a power of a single prime p , the number of first order elements of G is of the form $\lambda = 1 + kp$. While for $v > 1$ the expression for λ gives information concerning small λ 's only, every sufficiently large number being so representable, when $m-1$ is a power of a single prime p the condition includes the condition λ prime to $m-1$, and for $p > 2$, is stronger than that condition.

A peculiar property of the sets of transforms arising in the preceding proof is that each set, clearly invariant under s_0' , in turn generates s_0' . More generally, any set of first order elements of G which is transformed into itself by a first order element s_0 of G in turn generates s_0 . This result is itself an immediate consequence of the following. A first order element of G which transforms a subgroup of G into itself must be contained in that subgroup. The proof of the last result consists in noting that, otherwise, that first order element and the subgroup would generate a subgroup of G whose order was the order of the given subgroup multiplied by a divisor, not unity, of $m-1$. As for the result preceding, the subgroup of G generated by the given set, being consequently invariant under s_0 , must contain s_0 .

Since the order of a subgroup of G must also be prime to $m-1$, there will be associated with every subgroup of G an existent subset of the λ first order elements of G , namely, the set of first order elements of the subgroup. These "group-bearing" subsets of the λ first order elements of G can be independently characterized as those existent subsets of the λ first order elements which generate no other first order elements. By the reasoning of the preceding paragraph, a first order element which transforms a group-bearing subset of first order elements into itself must be contained in that subset. As the converse must also be true, it follows that a first order element, and hence indeed any element, of G either leaves both a subgroup of G and the set of first order elements of that subgroup invariant, or else transforms neither into itself. Clearly, two subgroups of G have a common element when and only when their sets of first order elements have a common element. We finally note the following. If $s_0^{(t_1)}, s_0^{(t_2)}, \dots, s_0^{(t_\lambda)}$ are the first order elements of some subgroup of G , then of all subgroups of G with exactly those first order elements there is one contained in, and one containing each. The smallest subgroup is of course the crosscut of all the subgroups in question, and will indeed be

the subgroup H generated by those first order elements⁽⁸⁰⁾. Now let K be the subgroup of G consisting of all the elements of G which transform H into itself. K will then contain all of the above subgroups. And since each of the first order elements of K transforms H into itself, they will all be in H , and hence will be the given first order elements. K is therefore that largest subgroup of our theorem.

The above theory is significant only if there exist m -groups of order prime to $m-1$ with more than one first order element, and, preferably, not consisting wholly of first order elements. For odd $m-1 \neq 1$ such an m -group is furnished by the complete m -adic δ -group which is of order 2^{m-1} and has 2^{m-2} first order elements. The 2^{m-2} second order subgroups are then the corresponding mutually exclusive subgroups into which the elements of the group are separated. For $m-1$ even, and λ prime to $m-1$, the λ second order elements of the ordinary dihedral group of order 2λ constitute such an m -group under the product of m elements as operation. In this m -group all λ elements are of m -adic order one. However, by the direct product method, we can obtain from this m -group, and a cyclic m -group of order g/λ , an m -group of arbitrary order g prime to the even $m-1$, and with an arbitrary divisor λ of g as the number of its first order elements. Most of the theory can be illustrated by means of these examples.

27. **Sylow subgroups of order p^α with g/p^α prime to $m-1$.** That Sylow's theorem is not universally valid for polyadic groups is shown by cyclic polyadic groups. We recall that a cyclic m -group of order g has a subgroup of order γ , γ a divisor of g , when and only when g/γ is prime to $m-1$. Hence, if p is a prime divisor of g , and p^α is the largest power of p which divides g , a cyclic m -group of order g will have a "Sylow subgroup" of order p^α when and only when g/p^α is prime to $m-1$. This example shows that our extension of Sylow's theorem to polyadic groups as given below is the most general that can be given in terms of a condition involving only the order and dimension of the group⁽⁸¹⁾. Note also that our cyclic group will have a Sylow subgroup for each of two distinct prime divisors of g when and only when g itself is prime to $m-1$, in which case it will have a Sylow subgroup for every distinct

⁽⁸⁰⁾ In this connection a theorem of Dörnte's is of interest. To wit, if an m -group is semi-abelian, and has at least one first order element, then its first order elements themselves constitute a subgroup of the m -group.

⁽⁸¹⁾ Other theorems however are possible. Thus, if G is an m -group of order g whose associated ordinary group G_0 has but one Sylow subgroup corresponding to a prime divisor p of g , in particular if G is semi-abelian, then the necessary and sufficient condition that G have a Sylow subgroup corresponding to p is that G have at least one element whose order is a power, possibly the zeroth, of p . Necessary, immediately; and sufficient. For if H_0 is that sole Sylow subgroup of G_0 of order a power of p , s the element of G , then s can transform H_0 only into itself, while s^{m-1} , being of ordinary order a power of p , must be in H_0 . Hence $H=H_0s$ is an m -group, and thus a subgroup of G of the requisite order. However, the Sylow subgroups of G corresponding to the prime p need not then constitute a complete set of conjugates under G . Thus, if G' is

prime divisor of g . The same situation holds for the *applicability* of our extension of Sylow's theorem to polyadic groups.

We proceed then to prove the following. *If the order g of an m -group G is divisible by p^α but not by $p^{\alpha+1}$, p a prime divisor of g , then if g/p^α is prime to $m-1$, G will have at least one subgroup of order p^α .* Our proof consists in expressing G in accordance with our basic coset theorem, and applying the Sylow theorem for ordinary groups to the associated ordinary group G_0 of G . By that coset theorem, and in the notation of the abstract containing group G^* of G , we may write $G = s'G_0$, where s' is any element of G . Since G_0 is also of order g , it will have at least one Sylow subgroup H_0 of order p^α . As G_0 is invariant under s' , H_0 will be transformed by s' into a Sylow subgroup H'_0 of G_0 of order p^α . But the Sylow subgroups of G_0 of order p^α constitute a complete set of conjugates under G_0 . Hence some element t of G_0 will transform H'_0 into H_0 . It follows that the element $s'' = s't$ of G transforms H_0 into itself.

Now s'' as element of G will be of some m -adic order γ which is a divisor of g . If then p^β is the largest power of p which divides γ , γ/p^β will be prime to $m-1$. It follows from our theory of cyclic groups that s'' will generate an element s , also in G , of m -adic order p^β . That is, s as element of G^* will be of ordinary order $p^\beta(m-1)$, and hence s^{m-1} of ordinary order p^β . But H_0 , being invariant under s'' , must also be invariant under s , and hence under s^{m-1} . Since s^{m-1} of order p^β is in G_0 , and transforms Sylow subgroup H_0 of G_0 of order p^α into itself, s^{m-1} must be in H_0 . It follows from the converse of the coset theorem that $H = H_0s$ is an m -group, hence a subgroup of G , and of order p^α .

Our proof actually shows then that for each Sylow subgroup of order p^α of G_0 there is at least one "Sylow subgroup" of order p^α of G whose associated ordinary group is that Sylow subgroup of G_0 . Conversely, the associated ordinary group of any subgroup of order p^α of G will be a subgroup of order p^α of G_0 , and hence a Sylow subgroup of order p^α of G_0 . Since one and only one subgroup of G_0 can be the associated ordinary group of a given subgroup of G , we thus see that there is a one-many correspondence thus set up between the Sylow subgroups of order p^α of G_0 , and those of G .

Of the three results which together constitute Sylow's theorem for ordi-

an ordinary abelian group, some extension of it G , also abelian, will consist wholly of first order elements. There will then be g/p^α Sylow subgroups of G of order p^α , yet each is invariant under G .

Again, in attempting to generalize the standard substitution group proof of the existence of Sylow subgroups by means of m -adic substitution groups, the writer succeeded in constructing a Sylow subgroup corresponding to the prime p for any symmetric m -adic substitution group of degree a power of p . It may be of interest to note that the rest of that standard proof goes over except for the last step. This one point of failure, and failure there must be for an arbitrary m -group, lay in our being able to establish that the number of elements in a double coset H_1sH_2 was the order of a subgroup of H_1 only for the case when H_2 and the transform of H_1 under s have a common element.

nary groups we have therefore proved that the first, pertaining to the existence of Sylow subgroups, go over for polyadic groups under the given order condition. We now show that under the same condition the third result also goes over. That is, *under the condition of the preceding theorem the Sylow subgroups of order p^α of the m -group G constitute a complete set of conjugates under G .* We have to show then that each subgroup of order p^α of G can be transformed into any other by an element of G . Let H' and H be any two such Sylow subgroups of G , H'_0 and H_0 the corresponding Sylow subgroups of G_0 . Some element t of G_0 will transform H'_0 into H_0 . That same t will then transform H' into a Sylow subgroup H'' of G also corresponding to H_0 , i.e., having H_0 for associated ordinary group. If then we can show that some element s' of G will transform H'' into H , it will follow that element $s=ts'$ of G must transform H' into H as required by our theorem.

Our problem therefore reduces to showing that of all Sylow subgroups $H^{(i)}$ of G corresponding to one and the same Sylow subgroup H_0 of G , each can be transformed into any other by an element of G . Since H_0 is the associated ordinary group of each $H^{(i)}$, it will be transformed into itself by the elements of each $H^{(i)}$. If then \bar{G} is the subgroup of G consisting of all the elements of G which transform H_0 into itself, each $H^{(i)}$ will be a subgroup of \bar{G} . On the one hand, therefore, Lagrange's theorem for polyadic groups shows that if \bar{g} is the order of \bar{G} , then \bar{g} will be divisible by p^α , but not by $p^{\alpha+1}$, while \bar{g}/p^α will be prime to $m-1$. On the other hand, since H_0 is invariant under each element of \bar{G} , it will be an invariant subgroup of \bar{G}_0 , the associated ordinary group of \bar{G} . First then, H_0 , whose order proclaims it to be a Sylow subgroup of \bar{G}_0 , is the only Sylow subgroup of \bar{G}_0 of order p^α . And since \bar{G} satisfies the order condition of our first theorem, it follows from the proof of that theorem that the subgroups $H^{(i)}$, which constitute all the Sylow subgroups of order p^α of G , and hence of \bar{G} , corresponding to H_0 , actually are the only subgroups of order p^α of \bar{G} .

If we expand \bar{G} in cosets as regards H_0 , each subgroup $H^{(i)}$, having H_0 for associated group, will appear as one of these cosets. Since H_0 is invariant under each element of \bar{G} , these cosets are the elements of the m -adic quotient group $\Gamma=G/H_0$. H_0 then appears as the identity of Γ_0 , the associated ordinary group of Γ , each $H^{(i)}$ as an element $\sigma^{(i)}$ of Γ . If s is an element of $H^{(i)}$, s^{m-1} is in H_0 . Hence for each $\sigma^{(i)}$, $[\sigma^{(i)}]^{m-1}=1$. That is, each $\sigma^{(i)}$ is a first order element of the m -group Γ . Conversely, if σ be any first order element of Γ , the corresponding coset of \bar{G} constitutes a subgroup of \bar{G} with H_0 for associated group, and hence is an $H^{(i)}$. The elements $\sigma^{(i)}$ are therefore the only first order elements of Γ . But the order of Γ is \bar{g}/p^α which is prime to $m-1$. The preceding section therefore tells us that the elements $\sigma^{(i)}$ constitute a complete set of conjugates under the elements of Γ . It follows that each of the subgroups $H^{(i)}$ of \bar{G} can be transformed into any other by an element of \bar{G} , and hence of G . Our proof is thus completed.

Clearly, the Sylow subgroups of order p^α of G are also the Sylow subgroups of order p^α of the subgroup of G generated by those Sylow subgroups. As that generated subgroup must satisfy the order condition of our theorem, it follows that the Sylow subgroups of order p^α also constitute a complete set of conjugates under the elements of the m -group they generate. As in the case of the preceding section, a weaker form of this result is that the Sylow subgroups of order p^α of G constitute a generalized complete set of conjugates under their own elements, that is, each can be obtained from another by a succession of transforms by their own elements.

Under the condition g/p^α prime to $m - 1$, two of the three parts of Sylow's theorem have thus been shown to hold verbatim for polyadic groups. Not so for the remaining part concerning the number of Sylow subgroups of order p^α . Let us return to the one-many correspondence between the Sylow subgroups of order p^α of G_0 and of G . As stated in different guise in the preceding proof, an element t of G_0 which transforms one Sylow subgroup of G_0 into a second will transform the Sylow subgroups of G corresponding to that first Sylow subgroup of G_0 into those corresponding to the second. Each Sylow subgroup of order p^α of G_0 therefore has the same number λ of corresponding Sylow subgroups of G . As seen above, λ is actually the number of first order elements of an m -group of order \tilde{g}/p^α prime to $m - 1$. Hence our result of the preceding section, coupled with the corresponding part of the Sylow theorem for ordinary groups, yields the following as the remaining part of our Sylow theorem for polyadic groups. *Under the condition of the preceding theorems the number of Sylow subgroups of order p^α of the m -group G of order g is of the form $(1+kp)\lambda$ where λ is a divisor of g/p^α and hence prime to $m - 1$ and p .*

In contrast with the above, we are able to extend the ordinary result that every element and subgroup of order a power of p is contained in a Sylow subgroup of order p^α , only for several still narrower classes of polyadic groups. It will be convenient to refer to this as the *inclusion property*. We do have immediately that *under the conditions of the preceding theorems if element s of order p^β of G , $\beta \geq 0$, transforms a Sylow subgroup H of order p^α of G into itself, then s is in H .* For otherwise, by our generalized corollary of §25, s and H would generate a subgroup of G whose order would be either p^α times a multiple of p , or p^α times a divisor, not unity, of $m - 1$, neither of which possibility is consistent with the given conditions. Hence also, if each element of a subgroup K of order p^β of G transforms H into itself, then K is contained in H . It follows that if G has but one Sylow subgroup of order p^α , in particular then if G is abelian, the inclusion property holds. Again, as in the proof of the first part of our extension of Sylow's theorem, we see that if element s of order p^β of G transforms a Sylow subgroup H_0 of order p^α of the associated ordinary group G_0 into itself, then s must be in a Sylow subgroup of order p^α of G , namely, H_0s ; likewise then for a subgroup K of order p^β of G that transforms H_0 into itself. For K_0 will then be contained in H_0 ; and with s in K , Sylow

subgroup H_0s of G will contain $K = K_0s$. Hence, if G_0 has but one Sylow subgroup, in particular if G_0 is abelian, i.e., G semi-abelian, the inclusion property is satisfied.

If we attempt to generalize the standard proof of the inclusion property for ordinary groups, we see that while the number of Sylow subgroups of order p^α of the m -group G is shown by our formula to be again prime to p , our work on transforms merely shows the number of transforms of a Sylow subgroup under the polyads formed from s or K to be a divisor of $p^\beta(m-1)$. We are thus led to the inclusion property only when $m-1$ itself is a power of the prime p . More generally, however, let G be reducible to a μ -group G' , with $\mu-1$ a power of p , say p^γ . The abstract containing group G'^* of G' , of order $p^\gamma g$, will then be a containing group of G . The corresponding containing group of the cyclic m -group generated by s , or of K , will be a subgroup of G'^* . It follows that the above number of transforms will also be a divisor of $p^\gamma g$, and hence actually be a power of p . The standard proof therefore again generalizes. Hence, *under the condition of the preceding theorems the inclusion property holds whenever G is reducible to a μ -group with $\mu-1$ a power of p* ; in particular, then, whenever G is reducible to an ordinary group.

An interesting consequence of this result is that the inclusion property for G holds under the condition of this section *whenever G has an invariant element*. For let s be an invariant element of G . Since its m -adic order is a divisor of g , the condition g/p^α prime to $m-1$, coupled with our formula for the real dimension of a cyclic m -group, shows that the cyclic m -group generated by s is reducible to a μ -group with $\mu-1$ a power of p . If then we apply our general criterion of reducibility to a μ -group to this cyclic μ -group, we obtain a condition which, with the invariance of s under G , becomes the condition that G be reducible to a μ -group. Note that in this case, which is that of a G derivable from a 2-group, for each Sylow subgroup of order p^α of G_0 there is but one corresponding Sylow subgroup of G . For the invariant element s will generate some invariant element of order a power of p , which, consequently, must be in every Sylow subgroup of order p^α of G . On the other hand two Sylow subgroups of G corresponding to the same Sylow subgroup of G_0 can have no common element.

All of the above concerned the Sylow subgroups of G corresponding to the single prime p . As stated early in this section, if the condition g/p^α prime to $m-1$ is to be satisfied for two distinct prime factors of g , then g itself must be prime to $m-1$, in which case the condition is satisfied for every prime factor of g . Hence, when g is prime to $m-1$, our extension of Sylow's theorem is universally valid. In particular, if G is abelian with g prime to $m-1$, then G has one and only one Sylow subgroup for each distinct prime divisor of g . By the preceding section, G then has one and only one first order element, which must then be in each of the Sylow subgroups of G , and, indeed, be the only element common to one such subgroup and the subgroup generated by

the others. G , therefore, is then the direct product of its Sylow subgroups; and when it is reduced to a 2-group, in the one manner allowed by its unique first order element, its Sylow subgroups are reduced to the Sylow subgroups of that 2-group.

Actually, this last result is but a special instance of a general result. We have earlier observed that when an m -group G is reduced to a μ -group G' , each subgroup of G' is the reduction of a subgroup of G , but a subgroup of G may not reduce to a subgroup of G' . On the other hand, let G satisfy our general condition g/p^α prime to $m-1$. Then G' satisfies the corresponding condition g/p^α prime to $\mu-1$. Our extension of Sylow's theorem is therefore applicable to both groups. Since transforms of elements by elements are the same in G and in G' , our complete set of conjugates result, applied to a Sylow subgroup of order p^α of G' and that of G reducing to it, shows that *when G is reduced to G' the Sylow subgroups of order p^α of G are reduced to the Sylow subgroups of order p^α of G'* . Finally, if $m-1$ is prime to g , the Sylow subgroups of G , without qualification, are reduced to the Sylow subgroups of G' .

28. Representation of an arbitrary m -adic group as a regular m -adic substitution group. We shall prove our result without the use of the coset theorem. The proof will then, indeed, immediately lead to another proof of the coset theorem, actually, the writer's original proof⁽⁸²⁾.

Let G be an arbitrary m -group of order g . The classes $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ are then to have for members the g classes of equivalent i -ads for $i=1, 2, \dots, m-1$. It will be convenient to symbolize the g members of Γ_i by $a_{ij}, j=1, 2, \dots, g$. Let s be any element of G . Then, as proved in more general form in §3, if the i -ads $\{s'_1, s'_2, \dots, s'_i\}$ and $\{s''_1, s''_2, \dots, s''_i\}$ of G are equivalent, the $(i+1)$ -ads $\{s'_1, s'_2, \dots, s'_i, s\}$ and $\{s''_1, s''_2, \dots, s''_i, s\}$ of G are equivalent, and conversely. s thus becomes an operator which carries the g classes of equivalent i -ads in 1-1 fashion into the g classes of equivalent $(i+1)$ -ads. Furthermore, if c represents the m -adic operation of G , then if the $(m-1)$ -ads $\{s'_1, s'_2, \dots, s'_{m-1}\}$ and $\{s''_1, s''_2, \dots, s''_{m-1}\}$ are equivalent, the elements $c(s'_1 s'_2 \dots s'_{m-1} s)$ and $c(s''_1 s''_2 \dots s''_{m-1} s)$ are identical, and conversely. It follows that s thus carries in 1-1 fashion the letters of $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$, that is, determines an m -adic substitution on the Γ 's.

Now given any i -ad $\{s_1, s_2, \dots, s_i\}$, and any $(i+1)$ -ad $\{s'_1, s'_2, \dots, s'_i, s'_{i+1}\}$, there is one and only one element s of G for which the $(i+1)$ -ads $\{s_1, s_2, \dots, s_i, s\}$ and $\{s'_1, s'_2, \dots, s'_i, s'_{i+1}\}$ are equivalent. It follows on the one hand that no two distinct elements of G can yield the same m -adic substitution on the Γ 's. The correspondence between the elements of G and the m -adic substitutions they determine is therefore 1-1. And since the m -adic substitution determined by $c(s_1 s_2 \dots s_m)$ is clearly the product of the m -adic substitutions determined by s_1, s_2, \dots, s_m , it follows that the m -adic substi-

⁽⁸²⁾ While the proof as given is for finite m -groups, it holds with little change for all m -groups. Hence the full generality of the consequent proof of the coset theorem.

tions determined by the elements of G constitute an m -adic substitution group simply isomorphic with G . Furthermore, the initial observation of this paragraph shows that given any two letters in successive Γ 's there is one and only one element s of G , and hence one and only one m -adic substitution of the simply isomorphic substitution group, that carries the letter in the first Γ into that of the second. This m -adic substitution group is therefore regular. We have consequently proved the following generalization of Cayley's theorem. *Every m -adic group can be represented as a regular m -adic substitution group.* In this connection, as seen in §16, the argument of §14 shows that *two regular m -adic substitution groups on the same letters which are simply isomorphic are conjugate.*

If we now wish to obtain the coset theorem from this result, we need merely observe that the ordinary group generated by the m -adic substitutions of the representation of G , as in the case of all m -adic substitution groups, is a containing group of the representation of G of index $m-1$, and hence by resymbolization of its elements can be made a containing group of G leading to the desired result. Since we have developed our theory of abstract polyadic groups abstractly, comparatively few applications of this generalization of Cayley's theorem are to be found in the present paper. Perhaps the most important of these is that it allows the concept of holomorph to apply to an arbitrary abstract polyadic group.

29. Invariant subgroups and quotient groups; the m -adic central quotient group. The present section may be considered a continuation of §4, our attention now being restricted to finite polyadic groups. We recall that if G is an m -group with ordinary associated group G_0 , then every subgroup H_0 of G_0 that is invariant under G leads to an m -adic quotient group $Q = G/H_0$ isomorphic with G . Clearly, if H_0 is of order h , the isomorphism between G and Q is $(h, 1)$. H_0 and Q may be called complementary groups as regards G . Since the elements of Q are the cosets of G as regards H_0 , the order of G is the product of the orders of H_0 and Q . Similarly for an actual subgroup H of G corresponding to H_0 .

Let σ be any element of Q , s any one of the elements of the corresponding coset. Then the m -adic order n of s must be divisible by the m -adic order v of σ . For, since $s^{[n]} = s$, $\sigma^{[v]} = \sigma$, and hence n is a multiple of v . That is, *the order of any element of an m -adic quotient group divides the orders of all the elements of the corresponding coset.* We recall that each coset corresponding to a first order element of Q constitutes a subgroup of G . These subgroups in fact are all the subgroups of G having H_0 for associated ordinary group, and hence also are semi-invariant subgroups of G . In particular, if H_0 is of order prime to $m-1$, each coset thus corresponding to a first order element of Q has at least one first order element.

Unlike the corresponding situation for ordinary groups, an element σ of Q may be of order a power of a prime p without any element of the correspond-

ing coset being of order a power of that prime. Thus, let G be a cyclic m -group of order $p^\alpha k$ where k , prime to p , is not prime to $m-1$. Then no element of G can have an order a power of p . But with H_0 the subgroup of G_0 of order k , $Q=G/H_0$ is cyclic, and of order p^α . Some element σ of Q will then indeed be of order p^α , while the corresponding coset has no element of order a power of p .

However, let σ be of order p^β , H_0 of order $p^\alpha k$, k prime to p , and suppose that k is prime to $m-1$. The elements of the cosets corresponding to the m -adic powers of σ will then together constitute a subgroup G' of G of order $p^{\alpha+\beta}k$. Since k is prime to $m-1$, G' will have a Sylow subgroup K of order $p^{\alpha+\beta}$ ⁽⁸³⁾. As the crosscut of K_0 and H_0 must be of order a power of p , it follows that K must have exactly p^α elements in each of the p^β cosets of G' as regards H_0 . The coset corresponding to σ therefore has at least one element of order p^γ with, of course, $\gamma \geq \beta$. That is, if the order of an element of an m -adic quotient group is a power of a prime number p , while the largest divisor prime to p of the order of the complementary group is prime to $m-1$, then the corresponding coset involves an element whose order is a power of p .

We recall the ordinary group result that every invariant subgroup of index 2 under any group includes all the elements of odd order contained in this group. In the case of an m -adic quotient group of order two, we recall our results of §23, and note that for m odd no such result can be expected. In fact, when the quotient group consists of two first order elements, each of the corresponding cosets, both then invariant subgroups of the given group as a consequence of the abelianism of the quotient group, may have an element of odd order; while when the quotient group consists of two second order elements both cosets consist of even order elements only. On the other hand, for m even the quotient group must consist of one first and one second order element. The coset corresponding to the first order element of the quotient group will then be an invariant subgroup of the given group, and any elements of odd order in the given group must be included in that invariant subgroup.

If H_0 is a subgroup of G_0 , the index of H_0 under G may be defined as the order of G divided by the order of H_0 , and, of course, gives the number of cosets in the expansion of G in either right or left cosets as regards H_0 —likewise for an H actually a subgroup of G . In the case of ordinary groups, we know that the index of the crosscut of two subgroups of a group under one of those subgroups is less than or equal to the index of the other subgroup under the group; while if the two subgroups are conjugate under the group, the inequality always prevails. If now H is a subgroup of an m -group G , K_0 a subgroup of G_0 , let L_0 be the crosscut of the associated ordinary group H_0 of H , and K_0 . Then, by writing G in the form $G_0 s$, with s in H , we see that the expansion of H_0 in right cosets as regards L_0 , and the expansion of G_0 in right

(83) Unless $\alpha=\beta=0$. But that case has already been treated. Actually, the first order elements of G may then conveniently be considered its Sylow subgroups of order p^0 .

cosets as regards K_0 , become the expansions of H and G in right cosets as regards L_0 and K_0 respectively. It then follows immediately that the index of L_0 under H is less than or equal to the index of K_0 under G . Now let K_0 be the associated ordinary group of a subgroup K of G conjugate to H under G . Since H and K are subgroups of G , we see from the discussion in §24 that H can also be transformed into K by some element t of G_0 . Since t then transforms H_0 into K_0 , the 2-group result for conjugate subgroups is applicable and thus yields the following. *If H and K are conjugate subgroups of an m -group G , the index of the crosscut of H_0 and K_0 under one of the subgroups is always less than the index of these subgroups under G .*

In this formulation we use "subgroup" in the strict sense, and thereby avoid the need of specifying that H_0 and K_0 , or H and K , are distinct. Now, as in the corresponding 2-group illustration, let H be of index 2 under G . With K conjugate to H , the above result shows H_0 and K_0 to be identical. H is then at least a semi-invariant subgroup of G . But since the resulting quotient group G/H , being of order two, is abelian, it follows that H is actually invariant under G . Hence, as for ordinary groups, *a subgroup of index 2 under any polyadic group is invariant*.

If an m -group G has at least one invariant element, these invariant elements clearly constitute an invariant subgroup of G which may be called the *central* of G . Note that a necessary and sufficient condition that our finite m -group G have a central is that it be derivable from an ordinary group. The central C of G , when it exists, is of course abelian, and coincides with G when and only when G is abelian. The quotient group G/C may be called the central quotient group of G , and, as with ordinary groups, is easily proved noncyclic whenever G is non-abelian.

It is readily seen that, when the central C of G exists, the associated ordinary group C_0 of C consists of all the elements of G_0 which are invariant under G . In general then, let us define the *associated central* C_0 of G as the subgroup of G_0 consisting of all the elements of G_0 invariant under G . C_0 then always exists, and being a subgroup of G_0 invariant under G , always leads to a quotient group G/C_0 . Since $G/C = G/C_0$ whenever C exists, we may call G/C_0 the *central quotient group* of G irrespective of the existence of C . Since each element of C_0 is also invariant under G_0 , C_0 is a subgroup of the central of G_0 when it does not coincide with the central of G_0 . It is readily seen, in fact, that the central of G_0 is invariant under G , each element of G yielding the same automorphism of that central. It follows that C_0 consists of those elements of the central of G_0 which are left invariant under any one element of G . In particular, when C exists, C_0 will coincide with the central of G_0 . In any case, C_0 is abelian, and coincides with G_0 when and only when G is abelian. It is then again easily proved that *the central quotient group of an m -group G is noncyclic whenever G is non-abelian*.

Any subgroup of G having C_0 for associated group leads to the central

quotient group G/C_0 and may be called a *relative central* of G . The relative centrals of G are then those cosets, if any, of the expansion of G as regards C_0 which correspond to first order elements of the central quotient group. They are of course semi-invariant subgroups of G , and are easily seen to be abelian. They can be independently characterized as the maximal subgroups of G having the property that, on being transformed by an element of G , each element of the subgroup is multiplied by one and the same element t of G_0 . Together, the elements of the relative centrals of G constitute all elements s of G with s^{m-1} in C_0 . The relative centrals corresponding to invariant first order elements of the central quotient group are characterized by the above multiplier t 's always being in C_0 , in which case, indeed, $t^{m-1}=1$. The unique central C , when it exists, is then the only one for which t is always 1.

30. Commutator, semi-commutator, and quasi-commutator subgroups. A direct extension to polyadic groups of the concepts of commutator, and commutator subgroup, is immediately obtainable. Given an m -group G , and in the notation of the abstract containing group of G , if s_1 and s_2 are any two elements of G , we may, as in ordinary theory, define the commutator of s_1 and s_2 to be $t = s_1^{-1}s_2^{-1}s_1s_2$. We shall also refer to s_1 and s_2 as the elements of the commutator. The commutator of s_1 and s_2 is then not an element of G , but of G_0 , the associated ordinary group of G , and is indeed that element of G_0 by which s_1 has to be multiplied on the right to yield the transform of s_1 under s_2 . The different commutators thus formed from elements of G therefore generate a subgroup of G_0 , if not G_0 itself, which may then be called the *commutator subgroup for G* .

As in ordinary group theory, the theory of commutator subgroups for polyadic groups is intimately bound up with the property of abelianism. But now our general formulation of semi-abelianism given in §7 suggests the need of a corresponding formulation of semi-commutator subgroup. The relative complexity of the resulting formulation then suggests a still further generalization of both concepts to what we term quasi-abelianism, and quasi-commutator subgroup. This wider generalization is also significant for ordinary groups. But while thus intimately related to certain recent work, in particular of Hall and Neumann⁽⁸⁴⁾, its direction seems to be new.

The immediate connection between abelianism and commutator subgroup is more clearly in evidence if we rewrite the usual $s_1s_2 = s_2s_1$ for the former in the equivalent form $s_1^{-1}s_2^{-1}s_1s_2 = 1$. Now the expression $s_1^{-1}s_2^{-1}s_1s_2$ that thus enters into both concepts is but a special instance of a word in the sense of Hall, or a rational expression in the sense of Baer. In general, a word W will be any expression of the form $s_{i_1}^{v_1}s_{i_2}^{v_2} \cdots s_{i_N}^{v_N}$, where the exponents are arbitrarily +1 or -1, the subscripts arbitrarily equal or unequal. If such an expression is to assume the value 1 for any choice of s 's in an m -group G , the notation

⁽⁸⁴⁾ B. H. Neumann, *Identical relations in groups I*, Mathematische Annalen, vol. 114 (1937), pp. 506-525. References will here be found to the work of Hall.

being that of the abstract containing group of G , the exponents must satisfy the condition $\nu_1 + \nu_2 + \dots + \nu_N \equiv 0 \pmod{m-1}$. Given m , consider then any specific class of words W_i whose exponents satisfy this condition. An m -group G will then be said to be *quasi-abelian* of corresponding formal type if the equations $W_i = 1$ are satisfied for every assignment of elements in G as values of the s 's, i.e., form a set of identical relations for G in the sense of Neumann. Now given an arbitrary m -group G , as a result of the exponent condition on the given class of words W_i each word assumes an element of G_0 as value when its letters are assigned elements of G as values. We shall call these words formal quasi-commutators, their values quasi-commutators, of the given formal type. The subgroup of G_0 generated by all of the quasi-commutators thus obtainable from elements of G will then be called the *quasi-commutator subgroup for G* of corresponding formal type.

In particular, any formulation of semi-abelianism as given in §7 can be rewritten in the above form. We correspondingly have formal semi-commutators, semi-commutators, and *semi-commutator subgroup* for an m -group G . While a certain degree of arbitrariness enters into the manner in which the equations of §7 are thus rewritten, it will be seen that this is irrelevant in the formation of the corresponding semi-commutator subgroup for G . In fact, our central theorem will be to the effect that the correspondence between type of quasi-abelianism and type of quasi-commutator subgroup, at present purely formal, is in fact intrinsic⁽⁸⁶⁾.

Our initial development, paralleling that of ordinary theory up to its main conclusion, will be given for quasi-commutator subgroups, the results then also holding for the successive specialization to semi-commutator and commutator subgroups. Consider then any one formulation of quasi-commutator subgroup for m -groups. From its very definition we then have that the *quasi-commutator subgroup for an m -group G reduces to the identity when and only when G is quasi-abelian of corresponding formal type*. Clearly the transform W_i by s is the same expression with each letter in W_i replaced by its transform

(86) Note that while we are interested in all, in the present instance finite, m -groups satisfying a given set of identical relations, Neumann considered instead the class of all identical relations satisfied by a given, of course ordinary, group. But it is the former concept that generalizes abelianism. Again, Hall, in the first paper cited by Neumann, builds up higher commutator forms merely out of ordinary commutators. His later concept of word-subgroup is identical, for ordinary groups, with our quasi-commutator subgroup. But again the emphasis is on all word-subgroups of a given group, rather than word-subgroup of given type for all groups—say of cardinal number less than, or less than or equal to, a given cardinal. And so our particular contribution of the relation between type of word-subgroup and type of identical relations is again unnoticed. We hasten to add that the researches of these authors in the directions they do pursue are profound. We also note that on reading Neumann's paper we changed our original formulation involving a finite number of identical relations to an arbitrary set of identical relations. In the case of our formulation of semi-abelianism, the finite can stand; for our theorem of §7 shows that an infinite set would always be equivalent to a finite subset thereof.

under s . That is, the transform of each quasi-commutator by an element of G is also a quasi-commutator. Hence, the *quasi-commutator subgroup for G of the given formal type* is a subgroup of G_0 invariant under G , when not G_0 itself. We may therefore form the m -adic quotient group of G relative to this quasi-commutator subgroup, i.e., the corresponding *quasi-commutator quotient group* of G . We then readily see that as in the ordinary theory, the *quasi-commutator quotient group of G of given formal type is quasi-abelian of the corresponding formal type*. For the isomorphism between G and the quotient group shows that a quasi-commutator formed from any elements of the quotient group corresponds to the quasi-commutator formed in the same way from corresponding elements of G , and hence is always the identity. Conversely, consider any quotient group of G which is quasi-abelian according to the given formulation. Again quasi-commutators of G correspond to quasi-commutators of this quotient group. Since the latter quasi-commutators can only be the identity, the former must be in the subgroup of G_0 complementary to this quotient group. That is, *every subgroup of G_0 which is invariant under G , and whose complementary quotient group is quasi-abelian of given formal type, contains the quasi-commutator subgroup for G of corresponding formal type*.

We are now able to prove the following fundamental theorem. *If two formulations of quasi-abelianism for m -adic groups are such that every m -group satisfying either satisfies the other, then the corresponding quasi-commutator subgroups for an m -group are always identical.* For let A' and A'' symbolize the two formulations of quasi-abelianism. If then, for a given m -group G , C'_0 and C''_0 are the quasi-commutator subgroups corresponding to A' and A'' respectively, the quasi-commutator quotient group G/C'_0 satisfies A' , G/C''_0 satisfies A'' . By our hypothesis, therefore, the m -group G/C'_0 also satisfies A'' , G/C''_0 also satisfies A' . Hence, by our last theorem, C'_0 contains C''_0 and C''_0 contains C'_0 , that is, C'_0 and C''_0 are identical.

The converse of this theorem is immediate; for if two formulations of quasi-commutator subgroup lead to identical subgroups for each m -group, then, if either of these subgroups is the identity, the other also is the identity. If then we say that two formulations of quasi-abelianism for m -adic groups define the same *type of quasi-abelianism* if every m -group satisfying either satisfies the other, while two formulations of quasi-commutator subgroup for m -adic groups define the same *type of quasi-commutator subgroup* if they yield identical subgroups for each m -group, we can conclude that *there is a 1-1 correspondence between types of quasi-abelianism for m -adic groups and types of quasi-commutator subgroup*. The correspondence between quasi-abelianism and quasi-commutator subgroup, originally depending on a particular formulation, has thus been shown to be intrinsic.

A useful partial consequence of our earlier proof is the following. *If two formulations of quasi-abelianism for m -adic groups are such that every m -group satisfying the first satisfies the second, then the quasi-commutator subgroup for an*

m-group corresponding to the first formulation always contains the one corresponding to the second. In this connection note that quasi-commutator subgroups of different types may be identical for a particular *m*-group. We therefore pause to prove the following. Given any finite set of distinct types of quasi-abelianism, there exists an *m*-group for which the corresponding quasi-commutator subgroups are all distinct. In fact, for each pair of these types there must exist an *m*-group quasi-abelian according to one type, but not according to the other. Represent these *m*-groups say as *m*-adic substitution groups on different letters, and form the *m*-group *G* therefrom by the direct product method. *G* then has the desired property. For it is readily proved from commutativity considerations that each quasi-commutator of *G* is the product of quasi-commutators of the same form, one for each of the above constituent groups of *G*, and conversely. Hence the quasi-commutator subgroups for *G* corresponding to any two of the given types of quasi-abelianism have, on the letters of the corresponding constituent group of *G*, a constituent group which is the identity in one case, not the identity in the other, and hence are themselves distinct.

Our basic "equivalence theorem" immediately translates our determination of the distinct types of semi-abelianism effected in §7 into a determination of the distinct types of semi-commutator subgroup. Since the proof of distinctness for the former was carried through by means of finite groups, we can therefore state that *there are as many distinct types of semi-commutator subgroups for m-adic groups as there are distinct divisors of m - 1*. For a divisor ρ of $m - 1$, the semi-commutator subgroup corresponding to ρ -semi-abelianism may be called the ρ -semi-commutator subgroup. From the above more general result it follows that *there exists an m-group for which the semi-commutator subgroups of all the distinct types are distinct*. In this case a simpler example of such a group is obtained merely by taking the direct product of groups, one for each divisor $\rho - 1$ of $m - 1$, which, as in §7, are *m*-groups ρ -semi-abelian, but not ρ' -semi-abelian for any divisor $\rho' - 1$ of $\rho - 1$ other than $\rho - 1$. Whether the semi-commutator subgroups of a given *m*-group are distinct or not, we may note the following relations between them. Since ρ_1 -semi-abelianism implies ρ_2 -semi-abelianism whenever $\rho_1 - 1$ is a divisor of $\rho_2 - 1$, it follows that in this case the ρ_1 -semi-commutator subgroup contains the ρ_2 -semi-commutator subgroup. More generally then, the crosscut of the ρ_1 and ρ_2 -semi-commutator subgroups contains the ρ_3 -semi-commutator subgroup, where $\rho_3 - 1 = \text{L.C.M.}(\rho_1 - 1, \rho_2 - 1)$, while the subgroup generated by the ρ_1 and ρ_2 -semi-commutator subgroups is contained in the ρ -semi-commutator subgroup, where $\rho - 1 = \text{H.C.F.}(\rho_1 - 1, \rho_2 - 1)$. In the second case, however, we can prove that *the subgroup generated by the ρ_1 and ρ_2 -semi-commutator subgroups is the ρ -semi-commutator subgroup with $\rho - 1 = \text{H.C.F.}(\rho_1 - 1, \rho_2 - 1)$* . For by the general theorem of §7, the semi-abelianism defined by the combination of ρ_1 -semi-abelianism and ρ_2 -semi-abelianism is equivalent to ρ -semi-

abelianism with the above ρ . The ρ -semi-commutator subgroup is therefore also the subgroup generated by all semi-commutators of the ρ_1 and ρ_2 formal types, and hence by the ρ_1 and ρ_2 -semi-commutator subgroups themselves.

In our march to the equivalence theorem we neglected certain developments related only to semi-commutators, or merely commutators, which might well have come first. In the limited generality of the first specialization we note that *each semi-commutator subgroup for an m-group G contains the commutator subgroup of the ordinary associated group G_0 of G*. In fact, if H_0 be such a semi-commutator subgroup, the quotient group G_0/H_0 can be identified as the associated ordinary group of the semi-commutator quotient group G/H_0 . Since G/H_0 is semi-abelian, G_0/H_0 , by a result of §7, is abelian, whence the above.

Clearly, two elements of a polyadic group are commutative when and only when their commutator is the identity. As in the corresponding situation for ordinary groups, it is readily proved that if the elements of a commutator respectively belong to two invariant subgroups of a polyadic group, the commutator is contained in the crosscut of the associated ordinary groups of those subgroups. It follows that *if two invariant subgroups of a polyadic group are such that their associated ordinary groups have only the identity in common, then every element of one of these subgroups is commutative with every element of the other*. Since two subgroups having at least one element in common have as many elements in common as have their associated ordinary groups, the above result is in this case equivalent to the following. *If two invariant subgroups of a polyadic group have one and only one element in common, then every element of one of these subgroups is commutative with every element of the other*. Actually, this special case is almost an immediate consequence of the corresponding ordinary theorem; for the one common element is then an invariant first order element of each of the subgroups, and hence of the polyadic group they generate⁽⁸⁶⁾, so that all three of these groups are reducible, and simultaneously so, to ordinary groups.

We have observed that the commutator of elements s_1 and s_2 of G is the element of G_0 which must be multiplied into s_1 to obtain the transform of s_1 under s_2 . Hence the complete set of conjugates of s_1 under G can be obtained by multiplying s_1 by commutators formed from elements of G . Since the commutator subgroup for G is invariant under G , it readily follows from this that all the transforms of an i -ad of G by polyads of G can be obtained by multiplying the i -ad by elements of the commutator subgroup for G . More specifically, it can be proved by way of the equivalence theorem that the transforms of an i -ad of an m -group G by the elements of G can be obtained by multiplying one such transform by elements of the ρ -semi-commutator subgroup for G , where $\rho - 1 = \text{H.C.F.}(i, m - 1)$; whence likewise for the transforms of the i -ad by the j -ads of G with fixed j . It follows from this result that if G is ρ -semi-

⁽⁸⁶⁾ Their direct product, therefore, as defined in §25.

abelian, all elements of G transform the i -ad into the same i -ad, as also do all j -ads with fixed j , a fact also easily shown directly.

We have defined an m -group G to be simple if G_0 has no subgroup other than the identity invariant under G . It follows then immediately that if a simple m -group G is not quasi-abelian of specified type, the corresponding quasi-commutator subgroup for G is identical with G_0 . If then, rather narrowly, we define G to be perfect if the commutator subgroup for G is identical with G_0 , it follows that every simple polyadic group of composite order is perfect. For otherwise G would be abelian, while G_0 would possess a subgroup other than the identity, yet invariant under G .

As in the case of ordinary groups, a subgroup of an m -group G may be called a characteristic subgroup of G if it corresponds to itself under every automorphism of G . Every automorphism of G determines an automorphism of G_0 . We may then define a subgroup of G_0 to be an associated characteristic subgroup of G if it corresponds to itself under every automorphism of G . In the case of invariance, a subgroup of G_0 invariant under G is always invariant under G_0 , but not conversely. Here the reverse situation holds. For clearly a characteristic subgroup of G_0 is also an associated characteristic subgroup of G , but not always conversely, as shown by the following example. The complete m -adic δ -group for $m=3$ is a triadic group of order four which has exactly two second order subgroups, one cyclic, the other non-cyclic. Each of the subgroups is therefore a characteristic subgroup of the group. Evidently the associated ordinary group of any characteristic subgroup of a polyadic group is an associated characteristic subgroup of the group. On the other hand, the associated ordinary group of this triadic δ -group is the ordinary axial group, and hence itself has no characteristic subgroup of order two.

It is readily proved that if G is non-abelian, then the central of G , if existent, is a characteristic subgroup of G , while the associated central of G is an associated characteristic subgroup of G . We now observe that every quasi-commutator subgroup for G , when not identical with G_0 , is an associated characteristic subgroup of G . In fact it is readily seen that under any automorphism of G a quasi-commutator involving certain elements of G will correspond to a quasi-commutator of the same form involving the corresponding elements of G . As the first set of elements take on all values in G , so do the second, so that actually the set of quasi-commutators of G of given formal type corresponds to itself under the automorphism.

Granting that the concept of quasi-abelianism and quasi-commutator subgroup has a certain degree of generality, ever further generalizations suggest themselves⁽⁸⁷⁾. Perhaps a guiding principle in such generalizations might

⁽⁸⁷⁾ Thus, if the above concepts be termed categorical, the following generalization, which we give only for ordinary groups, can be effected. With each of a given class of words W_i is associated a class of words W_{ik} involving only the letters of W_i . A group G will then be conditionally quasi-abelian of corresponding formal type if each $W_i=1$ is satisfied for every assign-

be the existence of an equivalence theorem. It may then be of interest to present our equivalence theorem in the following light. Each type of quasi-commutator subgroup for m -groups may be thought of as a function which assumes for each m -group G a subgroup of G_0 , if not G_0 , as value. Our equivalence theorem then asserts that this function is completely determined when it is known for what values of its argument it assumes the value 1.

31. The ϕ -subgroup of an m -adic group. The concept of a set of elements of a group being a set of independent generators of the group is equally applicable to a polyadic group. Whereas an ordinary group always has at least one element, namely the identity, which can never be one of a set of independent generators of the group⁽⁸⁸⁾, this need not be so in the case of a polyadic group. Thus a cyclic m -group of order g such that each prime divisor of g divides $m - 1$ can be generated by any one of its elements, and hence fails to possess an element of the type in question. If, however, an m -group G has at least one element which cannot be one of a set of independent generators of the group, then the set of all such elements constitutes a characteristic subgroup of G which may be called the ϕ -subgroup of G . It is a mark of the generality of the concept of the ϕ -subgroup that the self-same proofs which yield the corresponding results for ordinary groups apply verbatim to polyadic groups to give the following. *The ϕ -subgroup of an m -group G is the crosscut of all the maximal subgroups of G . If the ϕ -subgroup of an m -group G involves a non-invariant element or subgroup, the number of conjugates under G of this element or subgroup is greater than the number of the corresponding conjugates under the ϕ -subgroup.* As an application of the first of these results we may note that if a cyclic m -group G is of order $g = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} \gamma_0$, the p 's being the distinct prime divisors of g not divisors of $m - 1$, then the ϕ -subgroup of G exists if there be at least one such prime p , and is then the subgroup of order $p_1^{\alpha_1-1} p_2^{\alpha_2-1} \cdots p_r^{\alpha_r-1} \gamma_0$. Hence also, if we continue forming ϕ -subgroups starting with the cyclic m -group G , we finally arrive at the subgroup of order γ_0 which has no ϕ -subgroup. Since the ϕ -subgroup is always a "proper" subgroup, if we start with any finite m -group and successively form ϕ -subgroups,

ment of elements in G as values of its letters for which each $W_{ik} = 1$ is satisfied. Correspondingly, the conditional quasi-commutator subgroup of G is to be the smallest subgroup of G having the property that each W_i is in that subgroup for every assignment of elements in G as values of its letters for which each W_{ik} is in that subgroup. Our development up to, and including, the equivalence theorem then goes over. But now symbolic logic suggests that our conditions might involve more explicitly its apparent variables and other apparatus, and our horizon keeps receding. Thus, also, Neumann suggests the possibility of allowing constant elements of a group to enter into his identical relations, while Hall, in his higher commutator forms, from the start allows arbitrary subgroups of G individually to replace G as domain of a corresponding variable. It may be that a postulational procedure, perhaps centering around our actual development, or around the point of view about to be suggested, would bring order out of the chaos that thus threatens.

(88) Unless the group is the identity.

we arrive at a subgroup whose ϕ -subgroup is nonexistent. This weak statement, supported by the above example for $m > 2$, contrasts with the case $m = 2$ when the last existent ϕ -subgroup is always the identity.

In applying the second of the above two general results to the Sylow subgroups of the ϕ -subgroup of an arbitrary m -group G , we are hampered by the order condition of our extension of Sylow's theorem. Within the scope of that condition, we note first that if the ϕ -subgroup of G is of order g' , and if, with $p^{\alpha'}$ the largest power of the prime p dividing g' , $g'/p^{\alpha'}$ is prime to $m - 1$, then the ϕ -subgroup has a Sylow subgroup of order $p^{\alpha'}$ which then, as in the ordinary case, is unique. If then g' itself is prime to $m - 1$, the ϕ -subgroup will have one and only one Sylow subgroup for each distinct prime divisor of g' . Since, with g' prime to $m - 1$, the first order elements of the ϕ -subgroup constitute a complete set of conjugates under the ϕ -subgroup, it follows as for the Sylow subgroups that the ϕ -subgroup then has one and only one first order element. That is, when the order of the ϕ -subgroup of an m -group is prime to $m - 1$, the ϕ -subgroup is reducible to a 2-group. When so reduced its Sylow subgroups are reduced to the Sylow subgroups of the 2-group. As in the ordinary case, the ϕ -subgroup is then the direct product of its Sylow subgroups.

This result has an interesting consequence when the order of the given m -group is itself prime to $m - 1$. The ϕ -subgroup, if it exists, then has but one first order element. The invariance of the ϕ -subgroup therefore entails the invariance of this first order element under the given m -group. But this can only be the case if the m -group has no other first order element. Hence, *if an m -group of order prime to $m - 1$ has more than one first order element, its ϕ -subgroup is nonexistent*; that is, if an m -group of order prime to $m - 1$ is not reducible to a 2-group, each of its elements can be one of a set of independent generators of the group. On the other hand, if the m -group is reducible to a 2-group, its sole first order element can be generated by any other element, and hence is in the consequently existent ϕ -subgroup of the group.

We restrict our discussion of the ϕ -subgroups of primitive groups to primitive m -adic groups of ordinary substitutions. By the corresponding theorem of §18, the subgroups consisting of all substitutions omitting a given letter are maximal subgroups. Since these maximal subgroups can only have the identity in common, it follows that *the ϕ -subgroup of a primitive m -adic group of ordinary substitutions is either the identity, or else is nonexistent*. Certainly then when the primitive group in question does not possess the identity, and hence a fortiori when it is not reducible to a 2-group, its ϕ -subgroup is nonexistent. Strangely enough, the same may be true even when the identity is in the primitive group, then consequently reducible to a 2-group. Thus, the ordinary cyclic substitution group of order and degree a prime p remains primitive when extended to a $(p + 1)$ -group. Yet, while the identity and any other element together generate the $(p + 1)$ -group, each alone generates only itself.

32. Simply isomorphic m -adic groups; group of inner isomorphisms. We have defined simply isomorphic m -groups in §4, and have shown there that the transform of an m -group by an element or polyad is an m -group simply isomorphic with the given m -group. Restricting our attention to the case when the simple isomorphism is an automorphism, i.e., between an m -group and itself, we then have conversely, as in the case of ordinary groups, that any automorphism of an m -group can be effected by transforming it by an element. This really means that an m -group can be found of which the given m -group is a subgroup and which has an element so transforming the given m -group. This result may be proved as in the ordinary case by representing the given m -group as a regular m -adic substitution group in accordance with §28. Then, by §16, the principal holomorph of the m -group so represented certainly transforms it into each of its possible automorphisms.

Since the abstract containing group of an m -group is determined abstractly by the m -group, we see that a simple isomorphism between two m -groups determines a simple isomorphism between their abstract containing groups. Conversely, any simple isomorphism between the abstract containing groups of two m -groups which makes the classes of elements of the m -groups correspond determines a simple isomorphism between the m -groups. The simple isomorphism theorem of §8 may be considered a refinement of this obvious result. As that theorem is related to the determination theorem preceding it, so the following theorems are related to two of the generation theorems of §25. Their proofs, easily supplied, are therefore here omitted.

Two m -groups of the same order G' and G'' are simply isomorphic if their associated ordinary groups G'_0 and G''_0 contain two simply isomorphic subgroups H'_0 and H''_0 invariant under G' and G'' respectively, while G' and G'' are generated by H'_0 and H''_0 and two elements s_1 and s_2 such that if $s_1^{(m-1)}$ is the smallest positive power of s_1^{m-1} that occurs in H'_0 , then $s_2^{(m-1)}$ is the smallest positive power of s_2^{m-1} that occurs in H''_0 , and $s_1^{(m-1)}, s_2^{(m-1)}$ correspond in the given simple isomorphism of H'_0 and H''_0 . Moreover, it is assumed that s_1 and s_2 transform corresponding generators of H'_0, H''_0 into corresponding elements in the given simple isomorphism.

Two m -groups of the same order G_1 and G_2 are simply isomorphic if they contain two simply isomorphic invariant subgroups H_1 and H_2 respectively, and are generated by these subgroups and two elements s_1 and s_2 such that if s_1^λ is the smallest positive power of s_1 which occurs in the abstract containing group H_1^ of H_1 , then s_2^λ is the smallest positive power of s_2 which occurs in the abstract containing group H_2^* of H_2 , and s_1^λ and s_2^λ correspond as a consequence of the given simple isomorphism of H_1 and H_2 . Moreover, it is assumed that s_1, s_2 transform corresponding generators of H_1, H_2 into corresponding elements in the given simple isomorphism.*

We have observed that cyclic m -groups of the same order are simply isomorphic, and, obviously, no noncyclic m -group can be simply isomorphic

with a cyclic m -group. The following is a rather interesting application of the simple isomorphism theorem of §8. Let G' and G'' be two m -groups of order g reducible to cyclic polyadic groups, and let element s'_0 of G' be of the same m -adic order as element s''_0 of G'' . Then element s'^{m-1}_0 of G'_0 is of the same ordinary order as element s''^{m-1}_0 of G''_0 . Since G'_0 and G''_0 are ordinary cyclic groups of order g , a simple isomorphism can be set up between them which makes s'^{m-1}_0 correspond to s''^{m-1}_0 . The theorem in question then yields the following result. *If two m -groups reducible to cyclic polyadic groups are of the same order, and one m -group has an element of the same order as an element of the other, then the m -groups are simply isomorphic.*

Every automorphism of an m -group G permutes the elements of G according to a certain ordinary substitution. These substitutions clearly constitute an ordinary substitution group which may be called the group of isomorphisms of G . This terminology may be reconciled with that of §16 by noting that when G is represented as a regular substitution group, the corresponding $(K_0)_{11}$ of §16 is simply isomorphic with the group of isomorphisms of G .

On the other hand the substitutions which result merely from transforming G by its own elements need not form a 2-group. In fact, it is readily verified that they do form an ordinary substitution group when and only when G has an invariant element. However they clearly do form an m -adic group of ordinary substitutions which may then be called the *group of inner isomorphisms of G* . It is easily proved that as in the ordinary theory this m -group is *simply isomorphic with the central quotient group of G* . Hence it is simply isomorphic with G if and only if the associated central of G is the identity.

By using the fact that every automorphism of G can be obtained by transforming it by some element, it is readily proved that the group of inner isomorphisms of G is an invariant subgroup of the group of isomorphisms of G , if not identical with it, when the latter is extended to an m -group. On the other hand, the containing group of the group of inner isomorphisms is directly an invariant subgroup of the group of isomorphisms, when not identical with it. This containing group clearly consists of the substitutions according to which the elements of G are permuted when G is transformed by all of its polyads.

In extending the Sylow subgroup property of the group of inner isomorphisms of an ordinary group to m -groups, we have to restrict our m -adic G to be of order g with g/p^α prime to $m-1$, p^α being the largest power of the prime p dividing g . Since the order of I_{11} , the m -group of inner isomorphisms of G , divides g , I_{11} has the same order property. We can then show that I_{11} contains the same number of Sylow subgroups corresponding to p as G does, it being understood that if p does not divide the order of I_{11} , the corresponding Sylow subgroups of I_{11} are its subgroups of first order. While the proof differs little from the corresponding ordinary group proof, we cannot follow Miller

in dismissing it with a line, and instead present it at least in outline. The elements of I_{11} corresponding to the elements of a subgroup H of G constitute a subgroup H' of I_{11} which may be called H 's corresponding subgroup. Let H be a Sylow subgroup of G for the prime p in accordance with our hypothesis. Then, by considering I_{11} to be the central quotient group of G , and comparing the largest powers of p dividing the orders of H , I_{11} , and C_0 with those dividing the orders of H , H' , and the crosscut of H_0 and C_0 , we are enabled to conclude that H' is a Sylow subgroup of I_{11} for the prime p . Since corresponding elements of G and I_{11} transform corresponding subgroups into corresponding subgroups, the relation between the Sylow subgroups of G for the prime p and their corresponding subgroups of I_{11} is shown by the complete set of conjugates theorem to be a correspondence between all the Sylow subgroups of G , and all the Sylow subgroups of I_{11} , for the prime p . Finally, since any subgroup of G with given corresponding subgroup of I_{11} would be transformed into itself by any other subgroup of G with that corresponding subgroup of I_{11} , the above correspondence must be 1-1.

The fact that the central quotient group of a non-abelian group cannot be cyclic leads in ordinary group theory to the result that the order of the group of inner isomorphisms of a non-abelian group is at least four. In the case of a non-abelian m -group, the same theorem, used in conjunction with our determination of the m -groups of the first three orders, shows that the least order of the group of inner isomorphisms of m -groups is at least two when $m-1$ is even, three when $m-1$ is odd but divisible by 3, four when $m-1$ is neither divisible by 2 nor 3. The following examples show that these actually are the least orders of I_{11} for such m 's as well as the fact that the order of I_{11} may have any value from that least order up to and including the order four. First, by extending an ordinary group with I_1 of order four to an m -group, we see that for any m , I_{11} may be of order four. An I_{11} of order three is immediately furnished for $m-1$ even by the non-abelian m -group of order three itself. For $m-1$ odd, but divisible by 3, we have the following example with $m-1=3$, and hence by extension for any m with $m-1$ divisible by 3. Let G_0 be the ordinary cyclic group of order nine generated by the cyclic substitution $t = (a_1a_2a_3a_4a_5a_6a_7a_8a_9)$. Then $s = (a_2a_5a_8)(a_3a_9a_6)$ transforms t into t^4 while $s^3 = 1$. $G = G_0s$ is then a 4-group of order nine. Since G_0 is abelian, the associated central C_0 of G consists of the elements of G_0 invariant under s , i.e., of 1, t^3 , t^6 . The I_{11} of G is therefore also of order three. Finally an I_{11} of order two for $m-1=2$, and hence by extension for any even $m-1$, is exhibited by the following 3-group of order four. Let G_0 be the axial group 1, (ab) , (cd) , $(ab)(cd)$, s the substitution $(ac)(bd)$. Since s transforms G_0 into itself, while $s^2 = 1$, $G = G_0s$ is a 3-group of order four. As s transforms but 1 and $(ab)(cd)$ of G_0 into themselves, the C_0 of G , and hence also the I_{11} of G , is of order two.

When I_{11} is of order two it can abstractly be but the noncyclic m -group of

order two with its two first order elements. G is correspondingly separated into two abelian subgroups of half its order. It is readily proved that every abelian subgroup of G is contained in one of these subgroups. Conversely, if non-abelian G can be separated into two abelian subgroups, its I_{11} is of order two.

When I_{11} is of order three, it can be but the non-abelian group when $m-1$ is of the form $6\mu+2$ and $6\mu+4$, the abelian noncyclic group when $m-1$ is of the form $6\mu+3$, and either of these two when $m-1$ is of the form $6\mu+6$ as shown by extensions of the cases where $m-1=2$ and 3. In any event I_{11} consists of three first order elements, so that G is separated into three abelian subgroups of one-third its order. Again every abelian subgroup of G is contained in one of these three subgroups. We have not however been able to decide the question whether a non-abelian G which can be separated into three abelian subgroups of one-third its order must have I_{11} of order three.

We restrict our discussion of I_{11} of order four to m 's for which four is the least order of I_{11} , i.e., to $m-1$ not divisible by 2 or 3. Since $m-1$ is then prime to the order of I_{11} , while the smallest prime divisor of $m-1$ cannot be less than 5, our seemingly trivial form for the number of first order elements of an m -group with $m-1$ prime to g shows that I_{11} has exactly one first order element. I_{11} is therefore reducible to an ordinary group of order four, and indeed to the axial group. Furthermore, the subgroups of I_{11} reduce to the subgroups of the axial group when I_{11} is so reduced. It follows that G then has three abelian subgroups of half its order, while every abelian subgroup of G is contained in one of these subgroups. Conversely, if a non-abelian m -group with $m-1$ not divisible by 2 or 3 has more than one abelian subgroup of half its order, its I_{11} is reducible to the axial group.

33. Extension of Frobenius's theorem to m -adic groups. Thanks to recent work of Hall⁽⁸⁹⁾ on a wide generalization of Frobenius's theorem, the extension of the original theorem of Frobenius to polyadic groups is immediate. A very special case of Theorem III of Hall's paper may be stated as follows. If a subgroup H is transformed into itself by an element P , then the number of solutions of $X^N=1$ which lie in the coset HP is congruent to 0 modulo H.C.F. (N, h), where h is the order of H . Given, then, an arbitrary m -group G of order g , express G in the form $G=G_0s_0$ in accordance with our coset theorem. With n a divisor of g , the elements s of G whose m -adic orders divide n are those for which $s^{[n]}=s$, i.e., $s^{(m-1)n}=1$. Since G_0 is transformed into itself under s_0 , the above special case of Hall's theorem is immediately applicable to yield the following result. *The number of elements of an m -group G of order g whose (m -adic) orders divide an arbitrary divisor n of g is, if not 0, not only a multiple of n , but of n H.C.F. ($g/n, m-1$).*

That the number in question may be 0 is shown by a cyclic m -group of

⁽⁸⁹⁾ P. Hall, *On a theorem of Frobenius*, Proceedings of the London Mathematical Society, (2), vol. 40 (1935–1936), pp. 468–501.

order g with g/n not prime to $m-1$. If γ is any divisor of n , g/γ will also fail to be prime to $m-1$, and the cyclic group has no elements of orders dividing n . Note that when g is prime to $m-1$ this can never occur, for our otherwise arbitrary G must then have at least one first order element. Actually, by applying the above result, restated for n not a divisor of g , to the conjugate subgroups of G of §26—and for these subgroups, indeed, the result is easily obtainable with but the help of the ordinary Frobenius theorem—we obtain the following stronger result. *If an m -group G is of order g prime to $m-1$, and n is any divisor of g , then the number of elements of G whose orders divide n is a multiple not only of n , but of n H.C.F.($g/n, \lambda$), λ being the number of first order elements of G .*

34. Representation of an abstract m -adic group as a transitive (m, μ) substitution group. We shall consider the general question of representing an abstract m -group G of order g by a transitive m -adic group of μ -adic substitutions of degree n . (See §17.) The result can then immediately be specialized to the two cases of chief interest, $\mu=m$ and $\mu=2$, as well as to the case $n=g$, i.e., when the representing group is regular.

In the general case it is necessary to introduce polyadic groups intermediate between G and its associated ordinary group G_0 , groups whose introduction simultaneously with that of G_0 could have been used to generalize the theory at a number of points⁽⁹⁰⁾. Clearly each coset in the expansion of the abstract containing group G^* of G as regards G_0 is a polyadic group of order g under suitable extensions of the dyadic operation of G^* . In particular, if i is a divisor of $m-1$, the coset consisting of the i -ads of G , regarded as members of G^* , will thus constitute a group of dimension $(m-1)/i+1$. It will suffice to refer to this group as the polyadic group G_i of the i -ads of G . In particular $G_1=G$, $G_{m-1}=G_0$. As in the case of the subgroups of G , we may identify $(G_i)^*$ with the subgroup of G^* generated by the elements of G_i . $(G_i)_0$ is then simply G_0 . Finally, since the isomorphism between G^* and any other containing group of G established in §6 involves but a 1-1 correspondence between the elements of two corresponding cosets, G_i may similarly be set up by means of any containing group of G .

Suppose then that G can be represented by a transitive (m, μ) group G' of degree n , with, of course, $\mu-1$ a divisor of $m-1$. Corresponding to the polyadic group $G_{\mu-1}$ of the $(\mu-1)$ -ads of G there will then be the polyadic group $G'_{\mu-1}$ of the $(\mu-1)$ -ads of G' , conveniently set up by means of the containing group of G' generated by the substitutions of G' . $G'_{\mu-1}$ then consists of substitutions carrying each of the $\mu-1$ Γ 's on which G' is written into themselves⁽⁹¹⁾. Since G' is transitive, at least one substitution of $G'_{\mu-1}$ carries all

⁽⁹⁰⁾ E.g., see the end of the last footnote to §7. Likewise the concept of semi-invariant subgroups could correspondingly be generalized.

⁽⁹¹⁾ Note that these will also be the substitutions forming G'_0 when and only when the containing group generated by G' is of index $\mu-1$.

into itself. The set of all such substitutions in $G'_{\mu-1}$ then constitutes a subgroup $H'_{\mu-1}$ of $G'_{\mu-1}$ of order g/n . The associated ordinary group H'_0 of $H'_{\mu-1}$ is a subgroup of the associated ordinary group G'_0 of G' , and, in fact, consists of the substitutions of G'_0 carrying a_{11} into itself. It then follows from the transitivity of G' that neither H'_0 , if it be not the identity, nor any subgroup of H'_0 other than the identity is invariant under G' .

It therefore follows that for G to be representable by a transitive (m, μ) group of degree n , $\mu-1$ a divisor of $m-1$, it is necessary that $G_{\mu-1}$ have a subgroup $H_{\mu-1}$ of order g/n such that neither H_0 , that is, $(H_{\mu-1})_0$, if it be not the identity, nor any subgroup of H_0 other than the identity is invariant under G . We now prove this condition also sufficient. Each right coset of G^* as regards H_0 consists of g/n i -ads with fixed i . $H_{\mu-1}^*$ consists of $(m-1)/(\mu-1)$ of these cosets, one for each i a multiple of $\mu-1$. Each right coset of G^* as regards $H_{\mu-1}^*$ therefore also consists of $(m-1)/(\mu-1)$ of the right cosets of G^* as regards H_0 , one for each i differing from a fixed $i = i_0$ by a multiple of $\mu-1$. We may then choose i_0 so that $1 \leq i_0 \leq \mu-1$. And for each such i_0 there will be exactly n right cosets of G^* as regards $H_{\mu-1}^*$ which together exhaust all i -ads with $i - i_0$ a multiple of $\mu-1$. Now symbolize the n right cosets of G^* as regards $H_{\mu-1}^*$ with $i_0 = 1$ by the letters $a_{11}, a_{12}, \dots, a_{1n}$. These together will form the Γ_1 of the basis of our representation. Similarly for $\Gamma_2, \dots, \Gamma_{\mu-1}$, with i_0 correspondingly $2, \dots, \mu-1$. If now we multiply the elements of G^* on the right by an element s of G , the effect on the right cosets of G^* as regards $H_{\mu-1}^*$ is merely to permute them as units, the i_0 of such a coset becoming i_0+1 , reduced modulo $\mu-1$ if need be. In terms of the a 's therefore, the letters of Γ_1 go over in 1-1 fashion into those of Γ_2 , of Γ_2 into those of Γ_3, \dots , of $\Gamma_{\mu-1}$ into those of Γ_1 . Corresponding to s there is thus determined a μ -adic substitution s' of degree n on the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$. The set of all such μ -adic substitutions corresponding to elements of G clearly constitute an m -group G' , under the product of m substitutions as operation, isomorphic with G . This isomorphism is also simple. For if s_1 and s_2 are any two elements of G corresponding to the same substitution s' of G' , $t = s_1 s_2^{-1}$ must be both in G_0 and $H_{\mu-1}^*$, and hence in H_0 . The set of such t 's must then be a group contained in H_0 , and invariant under G , and hence consists of the identity only. That is, $s_1 = s_2$.

We have thus proved the following theorem. *A necessary and sufficient condition that an abstract m -group G of order g can be represented as a transitive m -adic group of μ -adic substitutions of degree n , $\mu-1$ a divisor of $m-1$, is that the polyadic group of $(\mu-1)$ -ads of G contains a subgroup of order g/n whose associated ordinary group, if not the identity, is not invariant under G , and contains no subgroup besides the identity invariant under G .* For the representation of G by a transitive m -adic substitution group of degree n this condition reduces to the condition that *the associated ordinary group of G contains a subgroup of order g/n which, if not the identity, is not invariant under G , and*

contains no subgroup besides the identity invariant under G , while for the representation of G by a transitive m -group of ordinary substitutions the condition becomes G contains a subgroup of order g/n whose associated ordinary group has the above property.

When $g=n$ the non-invariantive property is vacuously satisfied. Hence a necessary and sufficient condition that an abstract m -group G can be represented by a regular m -adic group of μ -adic substitutions is that the polyadic group of $(\mu-1)$ -ads of G possesses a first order element. When $\mu=m$ this leads again, through the identity of G_0 , to the universal representability of abstract m -groups as regular m -adic substitution groups. On the other hand, for the representation of G as a regular m -adic group of ordinary substitutions, it is necessary and sufficient that G possess a first order element. In particular, every abstract m -group of order prime to $m-1$ can be so represented.

C. FINITE m -ADIC LINEAR GROUPS

35. m -adic linear transformations. An ordinary transformation in n variables may be thought of as transforming an m -dimensional space Σ into itself. By analogy with m -adic substitutions, an m -adic transformation in n variables will then transform $m-1$ spaces Σ' , Σ'' , ..., $\Sigma^{(m-1)}$, of n dimensions each, cyclically into each other, i.e., $\Sigma' \rightarrow \Sigma''$, $\Sigma'' \rightarrow \Sigma'''$, ..., $\Sigma^{(m-1)} \rightarrow \Sigma'$. In particular, if $x_{i1}, x_{i2}, \dots, x_{in}$ are the old coordinates in $\Sigma^{(i)}$, and $x'_{i1}, x'_{i2}, \dots, x'_{in}$ the new, an m -adic linear transformation of Σ' , Σ'' , ..., $\Sigma^{(m-1)}$ will consist of $m-1$ sets of linear homogeneous equations of the form

$$\begin{aligned} A: \quad & \dots, x_{i1} = a_{11}^{(i)} x'_{(i+1)1} + a_{12}^{(i)} x'_{(i+1)2} + \dots + a_{1n}^{(i)} x'_{(i+1)n}, \dots, \\ & \dots, x_{i2} = a_{21}^{(i)} x'_{(i+1)1} + a_{22}^{(i)} x'_{(i+1)2} + \dots + a_{2n}^{(i)} x'_{(i+1)n}, \dots, \\ & \dots, \dots \dots \dots \dots \dots \dots \dots, \dots, \\ & \dots, x_{in} = a_{n1}^{(i)} x'_{(i+1)1} + a_{n2}^{(i)} x'_{(i+1)2} + \dots + a_{nn}^{(i)} x'_{(i+1)n}, \dots, \end{aligned}$$

where $i=1, 2, \dots, m-1$, the $i+1$ in the last case being replaced by 1, and where for each i the determinant of the n^2 coefficients is not zero.

As in the case of m -adic substitutions, we shall assume for simplicity that the spaces Σ' , Σ'' , ..., $\Sigma^{(m-1)}$ are mutually exclusive. The m -adic linear transformation A may then be considered to be an ordinary linear transformation of the $(m-1)n$ variables $x_{11}, \dots, x_{(m-1)n}$, but of the above special form⁽⁹²⁾. The product of m such linear transformations will again be a linear transformation of the same form, and hence serves to define an m -adic operation on m -adic linear transformations of Σ' , Σ'' , ..., $\Sigma^{(m-1)}$. It then readily follows that the class of all m -adic linear transformations of

⁽⁹²⁾ The above requirement that each of the $m-1$ separate determinants be different from zero is equivalent to this ordinary linear transformation's being nonsingular. See the end of the present section.

$\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ with complex coefficients form an m -group under this operation. For the associative law follows immediately from this reinterpretation. Furthermore, if in the equation $A_1 A_2 \cdots A_m = A_{m+1}$ all but A_i are specified m -adic linear transformations, A_i will be determined as an ordinary linear transformation and be given by the equation $A_i = A_{i-1}^{-1} \cdots A_1^{-1} A_{m+1} A_m^{-1} \cdots A_{i+1}^{-1}$. Now each A^{-1} carries Σ_j into Σ_{j-1} . Hence A_i carries Σ_j into Σ_k where $k \equiv j - (m-1) + 1 \pmod{m-1}$, i.e., $k \equiv j + 1 \pmod{m-1}$, and A_i is also an m -adic linear transformation.

We shall call any set of m -adic linear transformations of $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ which constitute an m -group under the above operation an *m -adic linear group in n variables*. Any such m -group will then be a subgroup of the above "complete" m -adic linear group in n variables. It follows that the necessary and sufficient condition that a finite set of m -adic linear transformations of $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ with complex coefficients form an m -adic linear group is that the product of any m members of the set is in the set. Unless otherwise indicated, m -adic linear group will mean finite m -adic linear group in the present paper. However, the infinite complete m -adic linear group is useful in serving as fundamental m -group for operations on arbitrary m -adic linear transformations. Its members, as ordinary linear transformations in $(m-1)n$ variables, will generate a containing group of index $m-1$ which may therefore be used in place of its abstract containing group. Its ordinary associated group, consisting of the products of $m-1$ m -adic linear transformations of $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$, will therefore consist of transformations which carry each $\Sigma^{(i)}$ into itself, and indeed of all linear transformations with complex coefficients which carry each $\Sigma^{(i)}$ into itself. We may therefore refer to such transformations as $(m-1)$ -ads of m -adic linear transformations, or briefly $(m-1)$ -ads.

While it will continue to be useful every so often to consider m -adic linear transformations as special forms of ordinary linear transformations, it is as generalization of ordinary linear transformation that they lend themselves to a corresponding generalization of the ordinary theory. For this purpose we return to our arbitrary m -adic linear transformation A , and as in ordinary theory represent it by the *m -adic matrix*

$$A = [A', A'', \dots, A^{(m-1)}],$$

where the component $A^{(i)}$ is the ordinary matrix

$$A^{(i)} = \begin{pmatrix} a_{11}^{(i)} & a_{12}^{(i)} & \cdots & a_{1n}^{(i)} \\ a_{21}^{(i)} & a_{22}^{(i)} & \cdots & a_{2n}^{(i)} \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1}^{(i)} & a_{n2}^{(i)} & \cdots & a_{nn}^{(i)} \end{pmatrix}$$

formed from the coefficients in the equations expressing each x_{ij} , with fixed i , in terms of the $x_{(i+1)k}$'s. The product of m m -adic linear transformations A_1, A_2, \dots, A_m is the m -adic linear transformation A obtained by performing these m transformations in succession. If then the corresponding m -adic matrices are $A_1 = [A'_1, A''_1, \dots, A^{(m-1)}_1], \dots, A_m = [A'_m, A''_m, \dots, A^{(m-1)}_m]$, then by following through the m transformations, we find that the m -adic matrix $A = [A', A'', \dots, A^{(m-1)}]$ is given by the following ordinary matrix equations

These equations then completely determine an m -adic operation on m -adic matrices. We shall call A the product of A_1, A_2, \dots, A_m , and write simply $A = A_1 A_2 \cdots A_m$. From the correspondence between m -adic matrices and m -adic linear transformations we then have immediately that the set of all m -adic matrices with complex a 's and fixed n , forms an m -group under this m -adic operation. Hence also we have the group property criterion for a finite m -adic group of m -adic matrices. As in ordinary theory, we therefore reinterpret m -adic linear group as an m -adic group of m -adic matrices.

We could correspondingly reinterpret the concrete containing group of the complete m -adic linear group. It suffices for our purpose merely to do so for the $(m-1)$ -ads of the containing group. If $\alpha^{(i)}$ is the matrix of the coefficients in the transformation thus expressing the x_{ij} 's in terms of the x_k 's, we shall represent the $(m-1)$ -ad α by the sequence of matrices $\alpha = (\alpha', \alpha'', \dots, \alpha^{(m-1)})$. The dyadic operation on matrix $(m-1)$ -ads is then seen from the corresponding transformations to be

$$(\alpha'_1, \alpha''_1, \dots, \alpha^{(m-1)}_1)(\alpha'_2, \alpha''_2, \dots, \alpha^{(m-1)}_2) = (\alpha'_1\alpha'_2, \alpha''_1\alpha''_2, \dots, \alpha^{(m-1)}_1\alpha^{(m-1)}_2),$$

while the product of an $(m-1)$ -ad and a monad, i.e., a single m -adic linear transformation, will be given by

$$(\alpha', \alpha'', \dots, \alpha^{(m-1)}) [A', A'', \dots, A^{(m-1)}] = [\alpha' A', \alpha'' A'', \dots, \alpha^{(m-1)} A^{(m-1)}],$$

of a monad by an $(m-1)$ -ad by

$$[A', A'', \dots, A^{(m-1)}](\alpha', \alpha'', \dots, \alpha^{(m-1)}) = [A'\alpha'', A''\alpha''', \dots, A^{(m-1)}\alpha'].$$

Clearly the identity among $(m-1)$ -ads is (E, E, \dots, E) , where E is the ordinary matrix identity, while the inverse of $(\alpha', \alpha'', \dots, \alpha^{(m-1)})$ is $((\alpha')^{-1}, (\alpha'')^{-1}, \dots, (\alpha^{(m-1)})^{-1})$.

We consider now the important question of change of variable. Let S be an m -adic linear transformation carrying the x_{ij} 's into the $x_{(i+1)k}$'s, T an m -adic linear transformation expressing the x_{ii} 's in terms of $X_{(i+1)k}$'s, and likewise the x'_i 's in terms of $X'_{(i+1)k}$'s. As a result, the X_{ii} 's are carried into the $X'_{(i+1)k}$'s according to an m -adic linear transformation R . We shall say that R is the result of m -adically changing variables in S according to T . Now with R , S , and T considered to be ordinary linear transformations on $(m-1)n$ variables, R is the result of an ordinary change of variables in S according to T , and hence is the transform of S with respect to T . If then in the equation $R = T^{-1}ST$ we follow through the successive linear transformations, we obtain the following results on the corresponding m -adic matrices. If

$$S = [S', S'', \dots, S^{(m-1)}], \quad T = [T', T'', \dots, T^{(m-1)}],$$

then the transform

$$R = [R', R'', \dots, R^{(m-1)}]$$

of S with respect to T , which is the result of m -adically changing the variables of S according to T , is given by the equations

$$R^{(i)} = [T^{(i-1)}]^{-1}S^{(i-1)}T^{(i)}, \quad i = 1, 2, \dots, m-1 \text{ (93).}$$

Closer to the ordinary concept of change of variable would be instituting an ordinary change of variable in each space $\Sigma^{(i)}$. This would then correspond to changing variables according to an $(m-1)$ -ad. As before, if S is an m -adic linear transformation, τ equivalent to an $(m-1)$ -ad of m -adic linear transformations, the result of changing variables in S according to τ will be an m -adic linear transformation R with $R = \tau^{-1}S\tau$. The corresponding formula for transforming the m -adic matrix $S = [S', S'', \dots, S^{(m-1)}]$, by the $(m-1)$ -ad $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$ to yield the m -adic matrix $R = [R', R'', \dots, R^{(m-1)}]$ may again be obtained by following through the transformations involved, or, perhaps just as easily, by applying our formulas for operations on $(m-1)$ -ads. We thus obtain

$$R^{(i)} = [\tau^{(i)}]^{-1}S^{(i)}\tau^{(i+1)}, \quad i = 1, 2, \dots, m-1.$$

While our m -adic matrix notation is more convenient in most applications, our later generalization of characteristic equation requires rather the matrix of the corresponding ordinary linear transformation in the $(m-1)n$ variables.

(93) These equations can also be obtained from the equations defining the m -adic operation on m -adic matrices, and the original m -adic definition of transform.

With $A = [A', A'', \dots, A^{(m-1)}]$, the corresponding ordinary matrix then has the following form

$$\begin{pmatrix} 0 & A' & 0 & \cdots & 0 \\ 0 & 0 & A'' & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A^{(m-2)} \\ A^{(m-1)} & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

If then D is the determinant of this matrix, $D', \dots, D^{(m-1)}$ of the components $A', \dots, A^{(m-1)}$ of A , it follows that

$$D = (-1)^{mn} D' D'' \cdots D^{(m-1)}.$$

By contrast, for the $(m-1)$ -ad $\alpha = (\alpha', \alpha'', \dots, \alpha^{(m-1)})$, the corresponding ordinary matrix has the components of α along its principal diagonal, zero's elsewhere, and the determinant of the matrix is always the product of the determinants of the components.

36. m -adic collineations and collineation-groups. If the variables of each space $\Sigma^{(i)}$ be considered homogeneous coordinates in a corresponding space $S^{(i)}$ of dimension $n-1$, our m -adic linear transformation A on $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$, may be said to define an m -adic collineation on $S', S'', \dots, S^{(m-1)}$. In fact, if we let the ratios $x_{i1}/x_{in}, \dots, x_{i(n-1)}/x_{in}$ be denoted by $y_{i1}, \dots, y_{i(n-1)}$, we are thus led to the m -adic linear fractional transformation $i=1, 2, \dots, m-1$:

$$y_{is} = \frac{a_{s1}^{(i)} y'_{(i+1)1} + \cdots + a_{s(n-1)}^{(i)} y'_{(i+1)(n-1)} + a_{sn}^{(i)}}{a_{n1}^{(i)} y'_{(i+1)1} + \cdots + a_{n(n-1)}^{(i)} y'_{(i+1)(n-1)} + a_{nn}^{(i)}}, \quad s = 1, 2, \dots, n-1.$$

Unlike the case of an m -adic linear transformation, our m -adic linear fractional transformation is in general not a special case of an ordinary linear fractional transformation on all the variables, since the denominators in general are not all the same. On the other hand it justifies our phrase m -adic collineation, since the equality of the denominators for each i insures our m -adic linear fractional transformation on the nonhomogeneous y 's carrying the straight lines of each $S^{(i)}$ into those of $S^{(i+1)}$. Moreover, the product of m m -adic linear fractional transformations of $S', S'', \dots, S^{(m-1)}$ will again be of that form, so that we can expect to have m -adic linear fractional groups, and hence m -adic collineation-groups.

Two m -adic linear transformations A_1 and A_2 on $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ will yield the same m -adic linear fractional transformation on $S', S'', \dots, S^{(m-1)}$ when and only when their m -adic matrices $A_1 = [A'_1, A''_1, \dots, A_1^{(m-1)}]$ and $A_2 = [A'_2, A''_2, \dots, A_2^{(m-1)}]$ are such that the elements of each component $A_i^{(i)}$ are a constant k_i times the elements of the corresponding component $A_2^{(i)}$.

This then is the condition that A_1 and A_2 represent the same m -adic collineation. Since the k_i 's need not be the same, A_1 and A_2 as ordinary linear transformations need not then represent the same collineation in the ordinary sense. If now we let τ be the $(m-1)$ -ad

$$((k_1, k_1, \dots, k_1), (k_2, k_2, \dots, k_2), \dots, (k_{m-1}, k_{m-1}, \dots, k_{m-1}))$$

whose components are all ordinary similarity-matrices, we see from the preceding section that $A_1 = \tau A_2$. We shall call an $(m-1)$ -ad each of whose components is a similarity-matrix a *similarity- $(m-1)$ -ad*. It follows that A_1 and A_2 represent the same m -adic collineation when and only when $A_1 A_2^{-1}$ is a similarity- $(m-1)$ -ad.

$A_2^{-1} A_1$ must then also be a similarity- $(m-1)$ -ad; but it will equal $A_1 A_2^{-1}$ when and only when the k_i 's are all equal. In fact, again by the preceding section, writing the above $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$, we find that $A_1 = A_2 \bar{\tau}$, where $\bar{\tau} = (\tau^{(m-1)}, \tau', \dots, \tau^{(m-2)})$, and hence $A_2^{-1} A_1 = \bar{\tau}$. Comparing these two results, we see that $A_2^{-1} \tau A_2 = \bar{\tau}$. Since A_2 is an arbitrary m -adic matrix, it follows that every m -adic matrix transforms a similarity- $(m-1)$ -ad $(\tau', \tau'', \dots, \tau^{(m-1)})$ into the similarity- $(m-1)$ -ad

$$(\tau^{(m-1)}, \tau', \dots, \tau^{(m-2)}).$$

By contrast, every similarity- $(m-1)$ -ad is transformed into itself by an $(m-1)$ -ad.

Consider now any m -adic linear group G . Since the product of two similarity- $(m-1)$ -ads is again a similarity- $(m-1)$ -ad, the similarity- $(m-1)$ -ads of G_0 , the associated ordinary group of G , will constitute a subgroup H_0 of G_0 . Since every m -adic matrix transforms a similarity- $(m-1)$ -ad into a similarity- $(m-1)$ -ad, H_0 will be invariant under G . We may therefore form the m -adic quotient group $K = G/H_0$. Each coset of G as regards H_0 can be written $H_0 A$ with A in G , and hence consists of elements of G representing the same m -adic collineation as A , and, in fact, of all such elements of G . The elements of K are thus in 1-1 correspondence with the distinct m -adic collineations represented by the elements of G . K may therefore be called the *m -adic collineation-group* corresponding to G .

An arbitrary m -adic collineation-group G may be given by arbitrarily representing each collineation by an m -adic linear transformation⁽⁹⁴⁾. If G is of order g , and written thus "on n variables," a modification of the ordinary treatment will yield an m -adic linear group of order $n^{m-1}g$ which is $(n^{m-1}, 1)$ isomorphic with G , and whose transformations have *components of determinant unity*. In fact let $S = [S', S'', \dots, S^{(m-1)}]$ be in G thus represented, with the determinants of its components $D', D'', \dots, D^{(m-1)}$ respectively. Let $\theta^{(i)}$ be any solution of the equation $[\theta^{(i)}]^n = [D^{(i)}]^{-1}$, and form the similarity-

(94) The product of m such representatives need not then be in the given set of representatives, but need merely represent the same m -adic collineation as some member of the set.

$(m-1)$ -ad $\tau = ((\theta', \theta', \dots, \theta'), (\theta'', \theta'', \dots, \theta''), \dots, (\theta^{(m-1)}, \theta^{(m-1)}, \dots, \theta^{(m-1)}))$. Then $A = \tau S = [(\theta', \theta', \dots, \theta')S', (\theta'', \theta'', \dots, \theta'')S'', \dots, (\theta^{(m-1)}, \dots, \theta^{(m-1)})S^{(m-1)}]$ represents the same m -adic collineation as S , and has all of its components of determinant unity. For each S there will thus be n^{m-1} A 's, and these constitute all of the m -adic linear transformations with components of determinant unity representing the same m -adic collineation as S . It then readily follows that the set of $n^{m-1}g$ m -adic linear transformations thus corresponding to the g elements of G constitute a linear m -group isomorphic with G . For let S_1, S_2, \dots, S_m be any m transformations in the original representation of G , $A_1 = \tau_1 S_1, A_2 = \tau_2 S_2, \dots, A_m = \tau_m S_m$ corresponding transformations with components of determinant unity. Then

$$A = A_1 A_2 \cdots A_m$$

has for its i th component

$$A^{(i)} = A_1^{(i)} A_2^{(i+1)} \cdots A_m^{(i)} = \tau_1^{(i)} \tau_2^{(i+1)} \cdots \tau_m^{(i)} S_1^{(i)} S_2^{(i+1)} \cdots S_m^{(i)} = \tau^{(i)} S^{(i)},$$

where $S = S_1 S_2 \cdots S_m$, and τ is a similarity- $(m-1)$ -ad. A therefore has components of determinant unity, and represents the same m -adic collineation as S . A is therefore in our set of $n^{m-1}g$ transformations, whence finally our result.

To compare the ordinary treatment with this modification of it, we introduce the following considerations. Given an m -adic linear group G , those similarity- $(m-1)$ -ads of G_0 which have equal components themselves constitute a subgroup H'_0 of G_0 invariant under G . We may therefore form the m -adic quotient group $K' = G/H'_0$. Each coset of the expansion of G as regards H'_0 consists of all transformations in G which as ordinary transformations on $(m-1)n$ variables correspond to the same ordinary collineation. We shall therefore call K' the collineation- m -adic group of G . If now an arbitrary collineation- m -adic group G be given by corresponding representative m -adic linear transformations, the ordinary treatment applies without modification; and if G is of order g , and on n variables, a linear m -adic group of order $(m-1)ng$ is thus obtained which is $[(m-1)n, 1]$ isomorphic with G , and whose members as ordinary transformations are of determinant unity. On the other hand, if an arbitrary m -adic collineation-group G be thus given, the ordinary unmodified treatment will in general be inapplicable. In fact, otherwise, the given representatives of the members of G must also be representatives of the members of a collineation- m -adic group. This will clearly not be so for random representations of the members of G . And the following example shows that the m -adic collineation-group G may be such that no representation thereof will represent a collineation- m -adic group. The triadic collineations corresponding to

$$A: [(1, 1), (1, -1)], \quad B: \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right]$$

generate a triadic collineation-group G of order 4. The most arbitrary representations of A and B are

$$A: \quad [(a, a), (b, -b)], \quad B: \quad \left[\begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix}, \begin{pmatrix} 0 & d \\ -d & 0 \end{pmatrix} \right].$$

By direct computation we find that $AAA = [(a^2b, -a^2b), (ab^2, ab^2)]$, $BBA = [(-acd, acd), (bcd, bcd)]$. As triadic collineations, AAA and BBA are identical, being the same as $[(1, -1), (1, 1)]$. As ordinary collineations, they can but be identified with $[(a, -a), (b, b)]$, $[(-a, a), (b, b)]$ which are never the same. Since any representation of G can have but one triadic linear transformation for each triadic collineation in G , no representation of this triadic collineation-group can also represent a collineation-triadic group.

If however G itself is an m -adic linear group, both methods are applicable. The unmodified treatment will then yield an m -adic linear group which is $[(m-1)n, 1]$ isomorphic with the collineation- m -adic group of G , and whose members as ordinary linear transformations have determinants unity. On the other hand, our modified treatment yields an m -adic linear group which is $(n^{m-1}, 1)$ isomorphic with the m -adic collineation-group of G , and whose members have components of determinant unity.

37. m -adic Hermitian invariants. A set of $m-1$ positive-definite Hermitian forms $J = [J', J'', \dots, J^{(m-1)}]$, one for each space $\Sigma^{(i)}$, will be said to be an m -adic (positive-definite) Hermitian form. Now

$$J^{(i)} = \sum_{k=1}^n \sum_{l=1}^n q_{kl}^{(i)} x_{ik} \bar{x}_{il}, \quad q_{kl}^{(i)} = \bar{q}_{kl}^{(i)},$$

can be transformed into

$$I^{(i)} = y_{i1} \bar{y}_{i1} + y_{i2} \bar{y}_{i2} + \dots + y_{in} \bar{y}_{in}$$

by a change of variables of the form

$$y_{ik} = \sum_{l=1}^k p_{kl}^{(i)} x_{il}, \quad k = 1, 2, \dots, n.$$

Hence $J = [J', J'', \dots, J^{(m-1)}]$ can be transformed into $I = [I', I'', \dots, I^{(m-1)}]$ by changing variables in Σ' , Σ'' , \dots , $\Sigma^{(m-1)}$ according to an $(m-1)$ -ad whose components, with $i = 1, 2, \dots, m-1$, are of the above form. The $(m-1)$ -ad, of course, is that obtained by solving for the x 's in terms of the y 's. It is further understood that in operating on J by this $(m-1)$ -ad, if x_{ij} is replaced by a certain expression, \bar{x}_{ij} is replaced by the conjugate of that expression.

If, on the other hand, J is transformed according to an m -adic change of variables, $J^{(i)}$, written on the variables of $\Sigma^{(i)}$, becomes an expression in the new variables not of $\Sigma^{(i)}$ but of $\Sigma^{(i+1)}$. We are thus led to define an m -adic Hermitian invariant of an m -adic linear group as an m -adic Hermitian form

$J = [J', J'', \dots, J^{(m-1)}]$ such that each transformation in the group carries $J' \rightarrow J''$, $J'' \rightarrow J'''$, \dots , $J^{(m-1)} \rightarrow J'$. It then readily follows that every m -adic linear group G has an m -adic Hermitian invariant. For let G'_0 be the Σ' constituent group of G_0 , the complete analogue of the G'_0 of an m -adic substitution group. The linear group G'_0 then has an Hermitian invariant J' on the variables of Σ' . Let S be in G , and let J'' be the result of transforming J' according to $S, \dots, J^{(m-1)}$ of transforming $J^{(m-2)}$ according to S . Then $J = [J', J'', \dots, J^{(m-1)}]$ will be an m -adic Hermitian form on $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$, and, as in §39 to come, is seen to be an m -adic Hermitian invariant of G .

By combining the above two results it follows that the variables of an m -adic linear group G may be so changed according to an $(m-1)$ -ad that $I = [I', I'', \dots, I^{(m-1)}]$, $I^{(i)} = x_{i1}\bar{x}_{i1} + x_{i2}\bar{x}_{i2} + \dots + x_{in}\bar{x}_{in}$, is an m -adic Hermitian invariant of the resulting transform of G .

An m -adic linear group G in n variables will be said to be linearly reducible⁽⁹⁵⁾ if by a suitable change of variables according to an $(m-1)$ -ad there will be in $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ subspaces⁽⁹⁶⁾ $\Sigma'_1, \Sigma''_1, \dots, \Sigma^{(m-1)}_1$ respectively on $v < n$ variables each such that $\Sigma'_1 \rightarrow \Sigma''_1 \rightarrow \dots \rightarrow \Sigma^{(m-1)}_1 \rightarrow \Sigma'_1$ under every transformation in the resulting transform of G . If for some such change of variables the subspaces $\Sigma'_2, \Sigma''_2, \dots, \Sigma^{(m-1)}_2$ on the remaining $n-v$ variables of $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ are also each transformed into the next, then G will be said to be intransitive. In the first case $\Sigma'_1, \Sigma''_1, \dots, \Sigma^{(m-1)}_1$ will be said to be a reduced set for G , in the second case a set of intransitivity of G . We then prove the theorem a linearly reducible m -adic linear group G is intransitive, and a reduced set constitutes one of the sets of intransitivity of G , subject, of course, to a change of variables in the reduced set according to an $(m-1)$ -ad thereon. We may assume the variables in the reduced set to be the first v variables of each $\Sigma^{(i)}$. Then G may be further transformed by an $(m-1)$ -ad so that it will have the m -adic Hermitian invariant I above. And this further change of variables, according to the form given above, merely transforms the reduced set according to an $(m-1)$ -ad on its variables. With G in this last form, consider its containing group G^* . Then $I^* = I' + I'' + \dots + I^{(m-1)}$ will be an ordinary Hermitian invariant of the ordinary linear group G^* , while the $(m-1)v$ variables constituting the reduced set for G form a reduced set for G^* without further transformation. But then G^* is in intransitive form with those $(m-1)v$ variables constituting a set of intransitivity of G^* . The same is then true of G .

An m -adic matrix $A = [A', A'', \dots, A^{(m-1)}]$ will be said to be in canonical form if each component $A^{(i)}$ is in the canonical form $(a_1^{(i)}, a_2^{(i)}, \dots, a_n^{(i)})$. Then the corresponding ordinary theorem generalizes, i.e., if A is of finite

(95) To distinguish between this extension of the ordinary concept and the totally unrelated polyadic concept we have termed reducibility.

(96) Strictly, a misnomer, but a convenient one.

m-adic order, then it can be reduced to canonical form by transformation by an $(m-1)$ -ad⁽⁹⁷⁾. We shall prove this result in the next section more expeditiously. However we here give the analogue of the ordinary proof for the sake of the concepts thus introduced.

We prove then that we can always find $m-1$ linear functions

$$y_{ii} = b_1^{(i)} x_{i1} + b_2^{(i)} x_{i2} + \cdots + b_n^{(i)} x_{in}, \quad i = 1, 2, \dots, m-1,$$

such that each y_{ii} is transformed into a constant θ_i times $y_{(i+1)1}$ by A . These $m-1$ functions may then be said to constitute a *relative m-adic invariant* of A . With A the transformation

$$x_{is} = \sum_{t=1}^n a_{st}^{(i)} x'_{(i+1)t}, \quad s = 1, 2, \dots, n; i = 1, 2, \dots, m-1,$$

we find that $(y_{ii})A = \theta_i y_{(i+1)1}$ provided the following equations are true:

$$\theta_i b_t^{(i+1)} = \sum_{s=1}^n b_s^{(i)} a_{st}^{(i)}, \quad t = 1, 2, \dots, n.$$

By successive substitution, with $i = 1, 2, \dots, m-1$, we obtain from these equations

$$\theta_1 \theta_2 \cdots \theta_{m-1} b'_t = \sum_{s=1}^n b'_s a_{st}^{(0)}, \quad t = 1, 2, \dots, n,$$

where the ordinary matrix $(a_{st}^{(0)}) = A_0 = A'A'' \cdots A^{(m-1)}$ ⁽⁹⁸⁾. A set of solutions b'_1, b'_2, \dots, b'_n , not all zero, of this last set of equations can always be found provided $\theta_1 \theta_2 \cdots \theta_{m-1}$ is a root of the characteristic equation of A_0 . The preceding equations, with $i = 1, 2, \dots, m-2$, then determine the remaining b 's, while the equations for $i = m-1$ are then automatically satisfied.

Having thus found a relative *m-adic invariant* of A , the remainder of the proof follows the lines of the standard proof. That is, by a change of variables according to an $(m-1)$ -ad given in part by our relative *m-adic invariant* of A , the new variables $y_{11}, y_{21}, \dots, y_{(m-1)1}$ are transformed according to the equations $y_{ii} = \theta_i y_{(i+1)1}$, $i = 1, 2, \dots, m-1$, and hence constitute a reduced set for the *m-adic linear group* generated by A . If then A is of finite *m-adic order*, further change of variables according to an $(m-1)$ -ad will

(97) It might be thought that since A as ordinary linear transformation is then of finite ordinary order, the standard theorem would apply. But note that an *m-adic matrix* in canonical form is not in canonical form as ordinary matrix. And from the contrary point of view, while A as ordinary matrix could thus be reduced to ordinary canonical form, the resulting linear transformation would no longer be an *m-adic linear transformation*; and the transformation used to obtain it would be a linear transformation on all the $(m-1)n$ variables in a form constituting a meaningless jumble from the point of view of *m-adic linear transformations*.

(98) Or, more expeditiously, from $(y_{ii})A^{m-1} = \theta_1 \theta_2 \cdots \theta_{m-1} y_{11}$.

change $y_{11}, y_{21}, \dots, y_{(m-1)1}$ into a set of intransitivity of the group generated by A . A then determines an m -adic linear transformation on the remaining $n-1$ variables, and the process may be repeated until A appears in canonical form, and, indeed, as the result of a single change of its original variables according to an $(m-1)$ -ad.

Our proof of the existence of relative m -adic invariants of A might have taken a different turn. Our original $(m-1)n$ homogeneous linear equations in the $(m-1)n$ undetermined b 's will have a set of solutions not all zero, and hence, as shown by the equations themselves, not all zero for any i , provided the determinant of their coefficients is zero. We are thus led to one equation in the $m-1$ unknowns $\theta_1, \theta_2, \dots, \theta_{m-1}$ which may be called the m -adic characteristic equation of A . Its right-hand member is zero; left, the determinant of A as ordinary linear transformation with the elements of the principal diagonal, all zero in A , replaced by $-\theta_{m-1}, \dots, -\theta_{m-1}, -\theta_1, \dots, -\theta_1, \dots, -\theta_{m-2}, \dots, -\theta_{m-2}$. With $\theta_1 = \theta_2 = \dots = \theta_{m-1} = \theta$, the m -adic characteristic equation of A becomes the ordinary characteristic equation of A as ordinary linear transformation. We are thus, in fact, assured of relative m -adic invariants of A with θ 's all equal. However, comparison with the earlier treatment yields the following result. The solutions of the m -adic characteristic equation of $A = [A', A'', \dots, A^{(m-1)}]$ consist of all sets of values $\theta_1, \theta_2, \dots, \theta_{m-1}$ for which $\theta_1 \theta_2 \dots \theta_{m-1}$ is a root of the characteristic equation of $A_0 = A'A'' \dots A^{(m-1)}$.

38. Reduction to canonical form. If for two m -adic linear transformations A and B in n variables there is a third C such that $B = C^{-1}AC$, then A and B will be said to be *conjugate*. This is equivalent to there being an $(m-1)$ -ad γ such that $B = \gamma^{-1}A\gamma$, since C and $A^{m-2}C$ on the one hand, γ and $A\gamma$ on the other, yield the same transform of A . It follows that the relation "A and B are conjugate" is an equivalence relation. Likewise for m -adic linear groups.

The following easily proved theorem reduces the problem of conjugate m -adic linear transformations in n variables to that of conjugate ordinary linear transformations in n variables. *The necessary and sufficient condition that $A = [A', A'', \dots, A^{(m-1)}]$ and $B = [B', B'', \dots, B^{(m-1)}]$ are conjugate is that $A_0 = A'A'' \dots A^{(m-1)}$ and $B_0 = B'B'' \dots B^{(m-1)}$ are conjugate.* In fact, if $B = \gamma^{-1}A\gamma$, $\gamma = (\gamma', \gamma'', \dots, \gamma^{(m-1)})$, then by our formula for change of variables according to an $(m-1)$ -ad

$$B' = [\gamma']^{-1}A'\gamma'', B'' = [\gamma'']^{-1}A''\gamma''', \dots, B^{(m-1)} = [\gamma^{(m-1)}]^{-1}A^{(m-1)}\gamma'.$$

Hence

$$B'B'' \dots B^{(m-1)} = [\gamma']^{-1}A'A'' \dots A^{(m-1)}\gamma',$$

whence the necessity of our condition. Conversely, if A_0 and B_0 are conjugate, γ' may be chosen to satisfy the last of the above equations. If then $\gamma'', \gamma''', \dots, \gamma^{(m-1)}$ are determined in accordance with the first $m-2$ of the

change of variable equations, the last of those equations will be automatically satisfied. An $(m-1)$ -ad $\gamma = (\gamma', \gamma'', \dots, \gamma^{(m-1)})$ is thus determined which transforms A into B .

This result contrasts strongly with the corresponding result for $(m-1)$ -ads. We may define two $(m-1)$ -ads α and β to be conjugate if there is an $(m-1)$ -ad γ such that $\beta = \gamma^{-1}\alpha\gamma$. From our formula for the product of two $(m-1)$ -ads it follows that $\alpha = (\alpha', \alpha'', \dots, \alpha^{(m-1)})$ and $\beta = (\beta', \beta'', \dots, \beta^{(m-1)})$ are conjugate when and only when the corresponding components $\alpha^{(i)}$ and $\beta^{(i)}$ are conjugate for each i . Hence, while the question of conjugacy for an m -adic matrix in n variables depends on but one ordinary matrix in n variables, the same question for an $(m-1)$ -ad depends on $m-1$ independent ordinary matrices in n variables each. Intrinsically, therefore, an m -adic matrix is far simpler than an $(m-1)$ -ad. This is rather surprising in that apart from change of variables they are of equal generality; for if A is a fixed m -adic matrix the relation $S = \tau A$ gives a 1-1 correspondence between all m -adic matrices S and $(m-1)$ -ads τ .

A more symmetrical though less useful condition for the m -adic matrices A and B being conjugate is that the $(m-1)$ -ads A^{m-1} and B^{m-1} are conjugate. In fact, if $A^{m-1} = \alpha$, the equation $A^m = \alpha A$ yields

$$A^{m-1} = (A'A'' \dots A^{(m-1)}, A''A''' \dots A', \dots, A^{(m-1)}A' \dots A^{(m-2)}).$$

The first component of A^{m-1} is therefore the A_0 of our previous condition, while all the components are conjugate. The present condition then follows. We may note that all the components of an $(m-1)$ -ad being conjugate is sufficient as well as necessary for the $(m-1)$ -ad being the $(m-1)$ -st ordinary power of some m -adic matrix. Intrinsically, then, an m -adic matrix is of the same degree of generality as an $(m-1)$ -ad with conjugate components. Too much emphasis, however, must not be placed on the forms assumed by a single element under transformation, our present concern.

Returning to our first condition for the conjugacy of m -adic matrices, we have immediately that $A = [A', A'', \dots, A^{(m-1)}]$ is conjugate to $[A_0, E, \dots, E]$, with $A_0 = A'A'' \dots A^{(m-1)}$. If now A is of finite m -adic order, then A^{m-1} , and hence its first component A_0 , is of finite order. A_0 is then conjugate to a matrix in the canonical form (a_1, a_2, \dots, a_n) . Hence, if A is of finite m -adic order, it is conjugate to an m -adic matrix in the canonical form $[(a_1, a_2, \dots, a_n), E, \dots, E]$.

More generally, if A is of finite m -adic order, it is conjugate to those m -adic matrices in the canonical form $[(a'_1, a'_2, \dots, a'_n), (a''_1, a''_2, \dots, a''_n), \dots, (a^{(m-1)}_1, a^{(m-1)}_2, \dots, a^{(m-1)}_n)]$ for which $a'_1 a''_2 \dots a^{(m-1)}_n = a_{j_1}, a_{j_2}, a_{j_3}, \dots, a_{j_n}$ a permutation of a_1, a_2, \dots, a_n . Since a_1, a_2, \dots, a_n are the roots of the characteristic equation of A_0 , we may say, as a consequence of the last section, that an m -adic matrix A of finite order assumes those canonical forms for which each selection of corresponding elements chosen from its components

constitutes a solution of the m -adic characteristic equation of A , while the corresponding roots of the characteristic equation of A_0 are all of its roots each with the correct multiplicity. In particular, we may make $a'_i = a''_i = \dots = a_i^{(m-1)}$ for each i . Hence the useful special result if A is of finite m -adic order, it is conjugate to an m -adic matrix in canonical form having equal components.

The most satisfactory generalization of an ordinary similarity-matrix is our similarity- $(m-1)$ -ad. An m -adic matrix each of whose components is a similarity-matrix will not in general remain of that form under transformation by an m -adic matrix⁽⁹⁹⁾. We therefore define an m -adic similarity-matrix as one which is conjugate to an m -adic matrix whose components are all similarity-matrices. It readily follows from our criterion for the conjugacy of m -adic matrices that $A = [A', A'', \dots, A^{(m-1)}]$ is an m -adic similarity-matrix when and only when $A'A'' \dots A^{(m-1)}$ is a similarity-matrix. In particular, every first order m -adic matrix is an m -adic similarity-matrix. In fact, A is of m -adic order one when and only when $A'A'' \dots A^{(m-1)} = E$. Hence the first order m -adic matrices are the conjugates of $[E, E, \dots, E]$.

Our chief reason for introducing the above concept is the following theorem. If an m -adic linear group has an m -adic similarity-matrix as invariant element, it is conjugate to a group in which each element is an m -adic matrix with equal components. By an m -adic change of variable the invariant similarity-matrix can be transformed into an m -adic matrix A in canonical form in which the components are now equal similarity-matrices. If the given group is correspondingly transformed, a conjugate group having A as invariant element is obtained. For each element B of the transformed group we thus have $A^{-1}BA = B$, i.e.,

$$B^{(i)} = [A^{(i-1)}]^{-1}B^{(i-1)}A^{(i)}, \quad i = 1, 2, \dots, m-1.$$

Since $A^{(i)}$ and $A^{(i-1)}$ are the same similarity matrices, we thus have $B^{(i)} = B^{(i-1)}$ for $i = 1, 2, \dots, m-1$, whence our theorem.

An m -adic linear group which is reducible to a 2-group automatically satisfies the condition of this theorem via its invariant first order element. An interesting property of any m -adic linear group thus conjugate to an "equi-component" group is that its m -adic collineation-group is identical with its collineation- m -adic group. In fact, in the case of an equi-component group itself, the associated ordinary group consists of $(m-1)$ -ads with equal com-

(99) Nevertheless, the set of such m -adic matrices of an m -adic linear group do constitute a subgroup, if existent, though in general not an invariant subgroup, of the group—likewise, those of these matrices having equal components. On the other hand, the subset of m -adic similarity matrices, in the sense about to be defined, while constituting an invariant subset of the m -adic linear group by their very definition, do not in general constitute a subgroup thereof. They do, however, when existent, separate into a number of semi-invariant subgroups with the subgroup of similarity- $(m-1)$ -ads as common associated group.

ponents, and hence has no other similarity- $(m-1)$ -ads than those with equal components; while under transformation by an $(m-1)$ -ad the similarity- $(m-1)$ -ads are unchanged. An equi-component group clearly has the following two properties: (a), it is simply isomorphic with a group of ordinary matrices in the specified number of variables, (b), no two distinct elements of the group have a pair of corresponding components the same. Now these properties are invariant for transformation by an $(m-1)$ -ad; (a), by its very formulation, (b), by our formulas for transformation by an $(m-1)$ -ad. Hence they are satisfied by all groups conjugate to equi-component groups. The class of groups satisfying condition (a), as well as the class of groups satisfying condition (b), are therefore each at least as wide as the class of groups conjugate to equi-component groups. Actually each of the first two classes is wider than the third, for the following examples show that neither of the first two contains the other⁽¹⁰⁰⁾. Let G_0 be the axial group with elements $((1, 1), (-1, -1)), ((-1, -1), (1, 1)), ((-1, -1), (-1, -1)), ((1, 1), (1, 1))$; $S_0 = [(1, 1), (1, 1)]$. Then in terms of the present operations the conditions of the construction theorem of §8 are satisfied, and $G = G_0 S_0$ is a triadic linear group in two variables. Now let \bar{G}_0 be the axial group with elements $(1, -1), (-1, 1), (-1, -1), (1, 1)$;

$$\bar{S}_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then $\bar{G} = \bar{G}_0 \bar{S}_0$ is a 3-group of ordinary matrices in two variables. With elements of G_0 and \bar{G}_0 corresponding in order, S_0 corresponding to \bar{S}_0 , the conditions of the simple isomorphism theorem of §8 are satisfied, so that G is simply isomorphic with \bar{G} . Hence G satisfies condition (a), but clearly fails to satisfy condition (b), since condition (b) is equivalent to the same condition stated for G_0 . For our second example we consider the rather trivial case $n=1$. With G_0 the cyclic group whose elements are $((i), (-i)), ((-1), (-1)), ((-i), (i)), ((1), (1))$, and $S_0 = [(1), (1)]$, $G = G_0 S_0$ is a triadic linear group in one variable satisfying condition (b). But it cannot satisfy condition (a); for it is non-abelian, while any polyadic group of ordinary matrices in one variable is readily seen to be abelian.

We conclude this section with a proof of the following generalization of the corresponding ordinary theorem. *Any abelian m-adic linear group is conjugate to a group each of whose elements is in canonical form with equal components.* We first prove this result for the case of an abelian group G having an m -adic similarity-matrix A . By the proof of the theorem preceding the above digression, G is conjugate to an equi-component group \bar{G} in which \bar{A} , the correspondent of A , has for its components equal similarity-matrices. Now the constituent \bar{G}' of the associated ordinary group \bar{G}_0 of \bar{G} will be an ordi-

⁽¹⁰⁰⁾ Clearly these distinctions constitute but a first glance at a probably wide theory.

nary abelian linear group, and hence can be transformed by an ordinary matrix α' so that each of its elements appears in canonical form. Since \bar{G}_0 will consist of $(m-1)$ -ads with equal components, the $(m-1)$ -ad $\alpha = (\alpha', \alpha', \dots, \alpha')$ will transform \bar{G}_0 into a group in which each element appears with equal components in canonical form. As α transforms \bar{A} into itself, it will therefore transform $\bar{G} = \bar{G}_0 \bar{A}$ into the conjugate of G of our theorem.

Now let G be an arbitrary abelian m -adic linear group, A some fixed element thereof. By a previous result, we may assume the group to have been so transformed by an $(m-1)$ -ad that A appears in canonical form with equal components A' . The $(m-1)$ -ad A'^{m-1} then has the equal components A'^{m-1} , also in canonical form. It follows from the invariance of any element $B = [B', B'', \dots, B^{m-1}]$ of G under A'^{m-1} that

$$A'^{m-1}B^{(i)} = B^{(i)}A'^{m-1}$$

for each i . If then we separate the variables of each space $\Sigma^{(i)}$ into sets $\Sigma_1^{(i)}, \Sigma_2^{(i)}, \dots, \Sigma_l^{(i)}$ according to their distinct multipliers in A'^{m-1} , the proof of the corresponding ordinary theorem shows that $B^{(i)}$ transforms the variables of each $\Sigma_j^{(i)}$ into those of $\Sigma_j^{(i+1)}$. Each element B of G therefore transforms $\Sigma'_j \rightarrow \Sigma''_j \rightarrow \dots \rightarrow \Sigma_j^{(m-1)} \rightarrow \Sigma'_j$. That is, G appears in intransitive form with the l sets of intransitivity corresponding to $j = 1, 2, \dots, l$. Now for each set of intransitivity the corresponding partial transformations constitute any abelian m -adic linear group. Moreover, the corresponding partial transformation of A is an m -adic similarity-matrix, since the corresponding partial transformation of A'^{m-1} has but one distinct multiplier. Hence, by our special result, each of these constituent groups can be thrown into the desired form by transformation by an $(m-1)$ -ad on the corresponding set of intransitivity. Together, these l partial $(m-1)$ -ads constitute an $(m-1)$ -ad on $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ which transforms G into the conjugate group of our theorem.

Clearly, every m -adic linear group, each of whose elements is in canonical form with equal components, is abelian. On the other hand, unlike the ordinary case, an m -adic linear group each of whose elements is in canonical form need not be abelian. It is readily proved that the necessary and sufficient condition that such a group be abelian is that its associated ordinary group consist of elements with equal components.

39. m -adic invariants. In the theory of ordinary linear groups in n variables the concept of a function of those variables precedes that of an invariant. In our theory of m -adic linear groups G in n variables it is therefore natural to replace the concept of a function by a set of $m-1$ functions, one for each of the spaces $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$. If we transform such a set of functions $[f'(x_{11}, x_{12}, \dots, x_{1n}), f''(x_{21}, x_{22}, \dots, x_{2n}), \dots, f^{(m-1)}(x_{(m-1)1}, x_{(m-1)2}, \dots, x_{(m-1)n})]$ by an m -adic linear transformation T of $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$, each function $f^{(i)}(x_{i1}, x_{i2}, \dots, x_{in})$ will become a function of $x_{(i+1)1}, x_{(i+1)2}, \dots, x_{(i+1)n}$.

We therefore define $f = [f', f'', \dots, f^{(m-1)}]$ to be an (absolute) m -adic invariant of T if T transforms $f' \rightarrow f'', f'' \rightarrow f''', \dots, f^{(m-1)} \rightarrow f'$; of G , if f is an m -adic invariant of each element of G . Actually, the following analysis shows this definition to be too narrow for a real generalization of the ordinary concept. But how to widen it without destroying our basic concept of $m-1$ spaces $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ we do not at present know.

Our chief result involves the associated constituent groups $G'_0, G''_0, \dots, G^{(m-1)}_0$ of G already introduced in §37 as the complete analogues of the corresponding concepts for m -adic substitution groups. More specifically, we saw that if G is an m -adic linear group of m -adic matrices $T = [T', T'', \dots, T^{(m-1)}]$, G_0 , the associated ordinary group of G , may be concretely given by a group of $(m-1)$ -ads $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$. For each τ , $\tau^{(i)}$ represents a transformation of the space $\Sigma^{(i)}$ into itself; and the set of $\tau^{(i)}$'s constitute an ordinary group, the associated constituent group $G_0^{(i)}$ above. It is then fundamental that, as in the case of m -adic substitution groups, the associated constituent groups of G are conjugate, each element T of G in fact transforming $G'_0 \rightarrow G''_0, G''_0 \rightarrow G'''_0, \dots, G^{(m-1)}_0 \rightarrow G'_0$. To verify this fact we need only observe that T transforms G_0 into itself; while if we follow through the operations involved in $T^{-1}\tau T$, we see that the i th component of the resulting $(m-1)$ -ad is the transform of the $(i-1)$ -st component of τ by $T^{(i-1)}$.

Now let $f = [f', f'', \dots, f^{(m-1)}]$ be an m -adic invariant of G ; that is, each element of G transforms $f' \rightarrow f'', f'' \rightarrow f''', \dots, f^{(m-1)} \rightarrow f'$. Each element τ of G_0 may be written as the product $T_1 T_2 \cdots T_{m-1}$ of $m-1$ elements of G . By following through these $m-1$ transformations we see that τ transforms f' into itself. But τ can operate on f' only through its first constituent τ' . Hence each τ' transforms f' into itself, and f' is an ordinary invariant of the associated constituent group G'_0 .

Conversely, let f' be any invariant of G'_0 , T_0 some element of G . T_0 will transform f' , a function of the variables of Σ' , into a function of the variables of Σ'' . Call this function f'' , i.e., $f'' = (f')T_0$. Likewise write $f''' = (f'')T_0, \dots, f^{(m-2)} = (f^{(m-1)})T_0$. Now $(f^{(m-1)})T_0 = (f')T_0^{m-1}$. Since f' is an invariant of G'_0 , it will actually be transformed into itself by each element of G_0 , and hence by the $(m-1)$ -ad T_0^{m-1} . That is $(f^{(m-1)})T_0 = f'$, and $f = [f', f'', \dots, f^{(m-1)}]$ is an m -adic invariant of T_0 . We now show that it is also an m -adic invariant of every element T of G , that is, of G . Since G''_0 is the transform of G'_0 under T_0 , it follows that if τ'' is any element of G''_0 , then for some element τ' of G'_0 , $(f'')\tau'' = (f')T_0 T_0^{-1} \tau' T_0 = (f')\tau' T_0 = (f')T_0 = f''$. Hence, f'' is an invariant of G''_0 , and likewise f''' of $G'''_0, \dots, f^{(m-1)}$ of $G^{(m-1)}_0$. Each element $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$ of G_0 will therefore transform each function $f', f'', \dots, f^{(m-1)}$ into itself. Hence, by writing an arbitrary element T of G in the form τT_0 , with τ in G_0 , we see that T , along with T_0 , will transform $f' \rightarrow f'', f'' \rightarrow f''', \dots, f^{(m-1)} \rightarrow f'$.

We have thus proved the following theorem. *Given an m -adic linear group*

G with first associated constituent group G'_0 , then every m -adic invariant $f = [f', f'', \dots, f^{(m-1)}]$ of G is such that f' is an ordinary invariant of G'_0 ; and, conversely, every ordinary invariant f' of G'_0 yields an m -adic invariant $f = [f', f'', \dots, f^{(m-1)}]$ of G . Clearly, this correspondence between m -adic invariants of G and ordinary invariants of G'_0 is 1-1. A like correspondence of course exists between the m -adic invariants of G and the ordinary invariants of $G_0^{(i)}$ for any i .

The weakness of our concept of m -adic invariants, already apparent from this reduction to ordinary invariants, is conclusively demonstrated by a consideration of invariants as group determiners. While the groups in question will in general be infinite, no part of the above discussion involves the hypothesis of finiteness in a linear group. Suppose then that $f = [f', f'', \dots, f^{(m-1)}]$ is an m -adic invariant of at least one m -adic linear transformation T_0 , and let G be the set of all m -adic linear transformations with f as m -adic invariant. It is then readily verified that G is an m -adic linear group. By the proof of the above theorem, f' is an invariant of G'_0 , and, likewise, f'' of $G''_0, \dots, f^{(m-1)}$ of $G^{(m-1)}_0$. If then $\tau', \tau'', \dots, \tau^{(m-1)}$ is any selection from $G'_0, G''_0, \dots, G^{(m-1)}_0$, and $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$, then $T = \tau T_0$ has f for m -adic invariant. T is therefore in G , and hence τ in G_0 . That is, the m -adic linear group defined by a given m -adic invariant is of that special kind in which the associated ordinary group consists of all selections, written as $(m-1)$ -ads, that can be made from the associated constituent groups.

When the above definition is extended to relative m -adic invariant, entirely corresponding results obtain. However, by a device similar to that which gave us our m -adic alternating groups, we can enlarge somewhat the role of relative m -adic invariant as group determiner. $f = [f', f'', \dots, f^{(m-1)}]$ will be a relative m -adic invariant of an m -adic linear transformation T if T transforms f so that $f' \rightarrow \kappa_1 f'', f'' \rightarrow \kappa_2 f''', \dots, f^{(m-1)} \rightarrow \kappa_{m-1} f'$, the κ 's being constants depending on T . Each T having f as relative m -adic invariant thus determines a κ -sequence. Furthermore, if T_1, T_2, \dots, T_m have f as relative m -adic invariant, so also will $T = T_1 T_2 \cdots T_m$; and the κ -sequence of T is determined by the κ -sequences of T_1, T_2, \dots, T_m by the same equations that connected the δ -sequences of our alternating group theory. We are thus led to a complete m -adic κ -group; and corresponding to any subgroup thereof, the set of all T 's with κ -sequences in that subgroup will be an m -adic linear group. Furthermore, whenever the associated ordinary group of the κ -subgroup does not consist of all selections from its constituent associated subgroups, the corresponding m -adic linear group will also not be of this special type. However, with the $f^{(i)}$'s homogeneous polynomials in the corresponding variables, any T having f for relative m -adic invariant can be changed to a T having f for absolute m -adic invariant by multiplying it into a suitable similarity- $(m-1)$ -ad; and conversely, without qualification. Hence the T 's corresponding to any one κ -sequence represent the same m -adic collineations as the

T 's having f for absolute invariant. All the m -adic linear groups corresponding to the various κ -subgroups therefore have the same corresponding m -adic collineation-group as the G defined by f as absolute invariant, and our seemingly greater freedom is largely illusory.

An obvious, but probably superficial, remedy for the relative triviality of our concept of m -adic invariant would be to allow each of the functions $f', f'', \dots, f^{(m-1)}$ to be functions not of the variables of the corresponding Σ alone, but of all of the Σ 's. It may be mere prejudice that makes us object to thus uniting the $m-1$ spaces of n dimensions each into one space of $(m-1)n$ dimensions; for, certainly, arbitrarily to give $m-1$ points, one for each space, is equivalent to giving one point in the combined space. One qualification does suggest itself. Corresponding to the condition of homogeneity for the polynomial invariants of ordinary theory, §36 suggests that the $f^{(i)}$'s be polynomials homogeneous in the variables of each Σ separately. However, a finally acceptable form for a general concept of m -adic invariant will probably involve changes in our original idea both more specific and more drastic than here suggested.

40. Generalization of m -adic substitution and transformation groups. The concept of m -adic linear group is readily extended to that of an (m, μ) linear group, analogous to our earlier (m, μ) substitution group. However, both concepts admit of a far wider extension. We shall give this extension only for m -adic substitution groups, the generalization of m -adic linear group being entirely similar⁽¹⁰¹⁾. It is of interest to note that this generalization continues to be a generalization even when $m=2$. But the resulting ordinary groups are then essentially realizations of Specht groups, referred to in the introduction, or subgroups thereof⁽¹⁰²⁾.

The concepts of an m -adic substitution on the letters of classes $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ is associated with the cyclic substitution $(\Gamma_1 \Gamma_2 \dots \Gamma_{m-1})$ on the classes themselves; for, under the m -adic substitution, $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$. More generally then let $\Gamma_1, \Gamma_2, \dots, \Gamma_r$ be any finite set of classes, σ any substitution on those classes themselves as elements. s will then be said to be a *polyadic substitution corresponding to σ* if, whenever σ replaces class Γ_i by class Γ_j , s carries the members of Γ_i in 1-1 fashion into the members of Γ_j . Clearly, if polyadic substitutions s_1, s_2, \dots, s_m on the members of $\Gamma_1, \Gamma_2, \dots, \Gamma_r$ correspond to $\sigma_1, \sigma_2, \dots, \sigma_m$ respectively, $s_1 s_2 \dots s_m$, the result of performing these m polyadic substitutions in succession, is itself a

⁽¹⁰¹⁾ A corresponding generalization of our narrow concept of m -adic invariant immediately suggests itself.

⁽¹⁰²⁾ On the other hand, groups of the permutations of sets of variables considered by L. Weisner (*Generalization of Lagrange's theorem*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 629-630) are but a very special case of the present generalization with $m=2$. We may note that the associated and containing ordinary groups of m -adic substitution groups, and, indeed, of the present generalization thereof, also come under this generalization with $m=2$, and thus tie up with Specht groups, or subgroups thereof.

Polyadic substitution corresponding to $\sigma_1\sigma_2 \cdots \sigma_m$, the product of the m corresponding ordinary substitutions. It follows from our last result on homomorphisms given in §4 that if G is an m -group of polyadic substitutions s on the members of $\Gamma_1, \Gamma_2, \dots, \Gamma_r$, under the above m -adic operation, the corresponding ordinary substitutions σ form an m -group B of ordinary substitutions. Moreover, G is *homomorphic to* B . We shall call B the *basic m-group* corresponding to the *Polyadic substitution group* G . In the case of our m -adic substitution groups, and more generally our (m, μ) groups, the basic m -group is of first order, its sole substitution consisting of a single cycle the number of whose letters is $m-1$ in the first case, a divisor $\mu-1$ of $m-1$ in the second.

As a consequence of the homomorphism between an arbitrary polyadic substitution group G and its basic m -group B , we see that there are the same number of polyadic substitutions in G for each substitution in B . Hence, also, the order of G is always a multiple of the order of B . Again, the ordinary substitutions corresponding to the polyadic substitutions forming any subgroup of G will form a subgroup of B , if not B itself; while to each subgroup of B there is at least one corresponding subgroup of G , i.e., the one consisting of all the elements of G corresponding to the elements of the subgroup of B , and hence containing all such subgroups.

For simplicity, we now restrict ourselves to mutually exclusive classes $\Gamma_1, \Gamma_2, \dots, \Gamma_r$ of the same finite number of letters n each⁽¹⁰³⁾. Given any substitution σ on those classes as elements, there will then be a total of $(n!)^r$ polyadic substitutions corresponding to σ . If then B is a given m -group of substitutions on those classes as elements, and b is the order of B , the $(n!)^rb$ polyadic substitutions corresponding to the elements of B are readily seen to constitute a polyadic substitution group with B as basic group. It may be called the *m -adic symmetric group of degree n with basic m -group B* . We can now state that any polyadic group with basic m -group B is a subgroup of the corresponding m -adic symmetric group. On the other hand, a subgroup of that m -adic symmetric group may have but a subgroup of B for basic group.

Of the theory of m -adic substitution groups we shall redevelop here only the general aspects of the theory leading to m -adic alternating groups. Again form the Vandermonde determinants $\Delta_1, \Delta_2, \dots, \Delta_r$ for the letters of $\Gamma_1, \Gamma_2, \dots, \Gamma_r$, respectively. If now a substitution σ on the Γ 's as elements be written in the primitive form

$$\Gamma_1 \Gamma_2 \cdots \Gamma_r$$

$$\Gamma_{i_1} \Gamma_{i_2} \cdots \Gamma_{i_r},$$

a polyadic substitution corresponding to σ will transform the Δ 's as follows:

$$\Delta_1 \rightarrow \delta' \Delta_{i_1}, \Delta_2 \rightarrow \delta'' \Delta_{i_2}, \dots, \Delta^{(r)} \rightarrow \delta^{(r)} \Delta_{i_r}, \quad \delta', \delta'', \dots, \delta^{(r)} = \pm 1.$$

(103) When B is transitive, the number of letters in the several Γ 's must of necessity be the same.

To describe this transformation completely, we must therefore not only specify the δ -sequence $\delta = [\delta', \delta'', \dots, \delta^{(v)}]$, but the substitution σ . We therefore form the couple $\{\sigma, \delta\}$. Given then a polyadic substitution group G , each element thereof uniquely determines a $\{\sigma, \delta\}$ couple. Moreover, if s_1, s_2, \dots, s_m are any m elements of G , $\{\sigma_1, \delta_1\}, \{\sigma_2, \delta_2\}, \dots, \{\sigma_m, \delta_m\}$ the corresponding couples, then $s = s_1 s_2 \cdots s_m$ has a couple $\{\sigma, \delta\}$ completely determined by the couples of s_1, s_2, \dots, s_m . For clearly $\sigma = \sigma_1 \sigma_2 \cdots \sigma_m$. On the other hand, let $\delta = [\delta', \delta'', \dots, \delta^v]$, $\delta_i = [\delta_i', \delta_i'', \dots, \delta_i^{(v)}]$. For any substitution σ on the Γ 's as elements, if σ carries Γ_i into $\Gamma_{i'}$, write $i'_j = i\sigma$. Then we will have

$$\delta^{(i)} = \delta_1^{(i)} \delta_2^{(i\sigma_1)} \cdots \delta_{m-1}^{(i\sigma_1 \cdots \sigma_{m-2})} \delta_m^{(i\sigma_1 \cdots \sigma_{m-2}\sigma_{m-1})}.$$

It again follows from our last result on homomorphisms that the class of $\{\sigma, \delta\}$ couples corresponding to the elements of G constitutes an m -group under the resulting m -adic operation on $\{\sigma, \delta\}$ couples, and hence that G is homomorphic to this m -group. We shall call the latter the $\{\sigma, \delta\}$ subgroup corresponding to G . The homomorphism in question then again assures us that there are exactly the same number of elements of G for each $\{\sigma, \delta\}$ couple in its $\{\sigma, \delta\}$ subgroup, and again yields the many-one relation between the subgroups of G and those of its $\{\sigma, \delta\}$ subgroup.

Clearly the relationship between G and its $\{\sigma, \delta\}$ subgroup is intimately bound up with the relationship between G and its basic m -group B . In fact, the very form of a $\{\sigma, \delta\}$ couple yields a many-one correspondence between the elements of the $\{\sigma, \delta\}$ subgroup corresponding to G , and of B ; while our formulation of the m -adic operation on $\{\sigma, \delta\}$ couples shows this correspondence to be a homomorphism—hence again the sameness of the number of $\{\sigma, \delta\}$ couples corresponding to different σ 's, and the many-one correspondence between the subgroups of the $\{\sigma, \delta\}$ subgroup, and of the basic m -group B , corresponding to G . Much can now be said of the interrelations between G , its $\{\sigma, \delta\}$ subgroup, and its basic m -group B . But they are all implicit in the fact that the above homomorphism between G and B is the one determined by the homomorphism between G and its $\{\sigma, \delta\}$ subgroup, and the homomorphism between that $\{\sigma, \delta\}$ subgroup and B .

When G is the polyadic symmetric group of degree n corresponding to a given basic m -group B , then, as in the case of m -adic substitutions, G will have at least one polyadic substitution for each of the 2^v possible δ -sequences, and each substitution σ in B , provided $n > 1$. The " $\{\sigma, \delta\}$ subgroup" may now be called the complete $\{\sigma, \delta\}$ group corresponding to B . With B of order b , the corresponding complete $\{\sigma, \delta\}$ group is then of order $2^v b$. We thus have a division of the corresponding $(n!)^v b$ polyadic substitutions into $2^v b$ mutually exclusive classes of consequently $(n!/2)^v$ members each.

Now in the many-one relations between the subgroups of the polyadic symmetric group of degree n , the complete $\{\sigma, \delta\}$ group, and the basic

m-group B consider only those (proper) subgroups of the complete $\{\sigma, \delta\}$ group which correspond to B itself. For each of these $\{\sigma, \delta\}$ subgroups there is a unique largest subgroup of the polyadic symmetric group. These may then be called the *polyadic alternating groups* of degree n with basic *m*-group B . The corresponding $\{\sigma, \delta\}$ subgroups are of orders $2^{\nu_1}b$, $0 \leq \nu_1 < \nu$, and the polyadic alternating groups correspondingly of orders $(n!/2)^{\nu}2^{\nu_1}b$, each consisting of all the elements in each of $2^{\nu_1}b$ of the above mutually exclusive classes. Note that if B is considered as a substitution group on the symbols $\Gamma_1, \Gamma_2, \dots, \Gamma$, rather than on the classes they symbolize, then one and the same B will serve for arbitrary n . Hence also the complete $\{\sigma, \delta\}$ group will be independent of n ; and for each $n > 1$ there will be as many polyadic alternating groups of degree n and basic *m*-group B as the complete $\{\sigma, \delta\}$ group has subgroups also corresponding to B .

By considering an arbitrary polyadic group G of degree n , and with basic *m*-group B , a subgroup of the corresponding polyadic symmetric group, we see that the $\{\sigma, \delta\}$ subgroup for G is actually a subgroup, proper or improper, of the complete $\{\sigma, \delta\}$ group corresponding to B . But that subgroup also must correspond to B . That is, we have a many-one relation between all polyadic groups of degree n with basic *m*-group B , and those subgroups of the complete $\{\sigma, \delta\}$ group which themselves correspond to B .

COLLEGE OF THE CITY OF NEW YORK,
NEW YORK, N.Y.

The Two-Valued Iterative Systems Of Mathematical Logic

BY
EMIL L. POST

PRINCETON
PRINCETON UNIVERSITY PRESS
LONDON: HUMPHREY MILFORD
OXFORD UNIVERSITY PRESS

1941

Dedicated
to
CASSIUS J. KEYSER

in one of whose pedagogical devi-
ces the author belatedly recognizes
the true source of his truth-table me-
thod.

CONTENTS

	Page
Introduction	1
Part I. PRELIMINARIES	10
1. Iterative closedness and generation	10
2. Two-valued systems of functions and the related logic of classes	15
3. The Jevons notation; expansions of membership functions	20
4. Different expansions of the same function; relevant and irrelevant variables	26
5. Iterative aspects of the Jevons notation	28
6. Duality	30
7. Uniform terms	33
8. The [A:a], [AA:], and [:aa] conditions .	35
9. A and a-components	40
Part II. DERIVATION OF CLOSED SYSTEMS	43
10. First order functions and systems, and their general significance	43
11. Systems of functions reducible to first order.	47
12. $[\beta]$, $[\gamma]$, and $[\beta,\gamma]$ systems	49
13. Logical sum systems; logical product systems .	50
14. $[\alpha,\beta,\gamma]$ systems	53
15. Alternating systems	55
16. Failure of the alternating condition	61
17. $[\alpha,\beta,\gamma,\delta]$ systems	62
18. $[\alpha]$ systems; preliminary discussion	63

CONTENTS

	Page
Part II (Continued)	
19. $[\alpha, \beta]$ and $[\alpha, \gamma]$ systems; preliminary discussion	67
20. $[\alpha]$ systems of self-dual functions	71
21. $[\alpha, \delta]$ systems	76
22. $[\alpha]$, $[\alpha, \beta]$, and $[\alpha, \gamma]$ systems; further discussion	77
23. $[\alpha]$, $[\alpha, \beta]$, and $[\alpha, \gamma]$ systems; concluding discussion -- the eight infinite families of systems	85
24. Summary	94
Part III. CO-ORDINATION AND APPLICATION	96
25. The inclusion relation	96
26. Sets of independent generators of the complete system	105
Bibliography	119

INTRODUCTION

In its original form the present paper was presented to the American Mathematical Society, April 24, 1920, as a companion piece to the writer's dissertation, henceforth referred to as Elementary Propositions [23]¹. Though submitted for publication the following year, at least a revision of the present paper was editorially suggested. Except for the finishing touches, including therein introduction and footnotes², the present version was effected in the years 1929-1932.

Though, as suggested by its title^{2a}, the main contribution of the present paper may be considered to be the complete solution of a certain problem in classical mathematics, the origin of the paper and, as far as the writer can see, its main interest lies in the study of two-valued propositional calculi. In the first of the four parts into which Elementary Propositions was divided the propositional calculus of Principia Mathematica [37], based on the primitives negation and disjunction, was studied; and, as a preliminary to that study, it was shown that if what we termed truth-tables were assigned to $\neg p$ and $p \vee q$ in accordance with their heuristic meanings, then not only was a truth-table thereby formally determined for every expression in the calculus, but every (two-valued) truth-table was thus obtained. The second part of that paper attempted to generalize the first part by allowing for an arbitrary finite number of primitive truth-functions, of arbitrary finite numbers of arguments, to each of which a truth-table was arbitrarily assigned. It was then immediately obvious that while each expression in such a calculus again had a corresponding truth-table, the set of such truth-tables might

¹ Numerals in brackets refer to the bibliography concluding the present paper.

² However, footnotes 47 and 48 (§25) were also written within the period stated.

^{2a} This rather refers to the original title of the present version, "Determination of all iteratively closed two-valued systems of functions."

now be but a proper subset of the set of all possible truth-tables. Two problems thus suggested themselves. First, what truth-tables could be assigned to the primitive functions so that the set of all possible truth-tables would result; secondly, allowing for arbitrary initial truth-tables, what are the various possible sets of truth-tables that can thus result. Added interest was lent to the first problem by Nicod's achievement [21] in basing a propositional calculus on but one primitive truth-function in accordance with the pioneering work of Sheffer [27]³. While the second problem was at first viewed as an aid to the solution of the first, its intrinsic interest soon became apparent, and made it our main problem.

A brief summary of the result of our solution of this main problem appears in Elementary Propositions⁴. In this connection it should be noted that the operation of substitution, which is the only operation used in building up the various expressions of a propositional calculus by means of variables and primitive functions, has its counterpart for the corresponding truth-tables. The problem can then be considered entirely in terms of truth-tables. The above summary is then at the moment best restated as it appears in the printed abstract of the present paper [22] ".....there are 66 different systems generated by primitive tables with no more than three arguments, and 8 infinite families of systems which require tables of four or more arguments." Such a generated system of truth-tables clearly has the property of being closed under the operation of substitution, or, as we shall say, iteratively closed. It is then significant, as was noted in our abstract, that the converse also is true.⁵ In

³ On the other hand, such a basis for a logic of propositions as given by Tarski [30] is outside the scope of our investigation; since, in addition to the primitive truth-function equivalence, it employs quantification of propositional variables.

⁴ See [23], §5, (pp. 173-4).

⁵ Provided we exclude the null system of functions, which vacuously satisfies the condition of iterative closedness, or else allow a null set of generators, thus adding one to the number of generated systems. While this omission is easily supplied (but see footnote 51 (§25)), more serious is our exclusion of functions of no variables, i.e., constants. In this omission we follow the usual form of a propositional calculus, e.g., such as that of Principia Mathematica. Actually, there would be little difficulty in modifying our count of systems in

terms of generated systems and logic, we may be said to have determined all the non-equivalent sub-languages of the language of the complete two-valued propositional calculus. Since in itself a truth-table is but a two-valued function, two-valued referring to the range of values of arguments and possible values of function, in terms of iteratively closed systems we are led to the title of this paper,^{5a} or, to use the terminology of Wiener,⁶ we have determined all iterative fields of two-valued functions.

The first version of the present paper was presented in terms of truth-tables, and followed the method of discovery via the concept of generated system.⁷ That every closed system of truth-tables could be generated by a finite set of tables only appeared after all such generated systems had been found. Since it was impossible to present all the calculations and tabulations, in addition to arguments, inherent in this procedure, that first version may be said to have been but a report on the complete solution. The present version centers around the concept of closed system, and merely pauses in each instance to verify that the closed system in question can be generated by a

the two-valued problem to allow for the two constants there occurring, since those constants play essentially the same rôle as two of the four two-valued functions of one variable. On the other hand, generalizing our membership-functions of classes, (§2), to allow for arbitrary constants would fundamentally alter the scope of the present paper, as well as break down the isomorphism with the two-valued problem.

Our count of systems would have been considerably simplified if we failed to distinguish, for example, between functions of one variable and functions of two variables whose values are independent of the value of one of those variables. That such distinctions are real can be illustrated in a variety of ways; and while the simplification of the count, if desired, can be immediately effected, not so the meticulous recovery of the lost distinctions.

5a See footnote 2a.

6 See [41], p. 7, for the definition of iterative field, [40], p. 159, for the class of 'operations' generated by a single operation. Where Wiener analyses the operation of substitution, which is at the bottom of these concepts, into at least semi-atomic parts, we have found it more convenient to use a more uniform definition. On the other hand, Wiener makes no reference to a specific class of variables to be employed, thus making the first of the above definitions ambiguous, the second misleading.

7 A brief outline of this method might be given. Since, neglecting the particular variables heading a truth-table, there are but four truth-tables of one argument, sixteen of two, while

finite set of generators. Its synthetic development unifies the argumentative portions of the original development, and avoids all of the latter's tabulations, and most of its calculations, so that it is complete in itself.⁸ Furthermore, following the suggestion of the referee of the original version, the truth-table has been replaced by logical expressions in the Jevons notation. To preserve the precision of the truth-table development these expressions are interpreted not as truth-functions of propositions but as what we call membership functions of classes. While the advantages are not all with the Jevons notation, its greater flexibility perhaps justifies the change.⁹

To avoid the unclarity of the first version of the present paper, we have devoted the first two sections of the present

the tables of one or two arguments generated by a set of tables similarly restricted can be found by using only such tables, mere systematic calculation and tabulation served to yield all the distinct sets of such tables that could be so generated. On the other hand, special methods adapted to the individual cases were needed to obtain a formula for an arbitrary table in each of the corresponding systems. The next step was a systematic procedure which, again by special arguments, yielded formulae for all tables of more than two arguments which, taken singly, would generate systems not identical with one of the previously found systems. By restricting these tables to those of three arguments, all systems that could be generated by one table of three arguments, but not by tables of one and two arguments, were found, whence it was an easy matter to find all systems generated by arbitrary sets of tables of no more than three arguments. One repetition of this weeding-out process then so cleared the field that it was possible to follow the procedure for 'third order' systems in the case of n -th order systems with arbitrary $n > 3$, and thus complete the solution.

⁸ On the other hand, much additional information yielded by the original solution is lost. But see footnote 50, (§25).

⁹ Particularly for those systems satisfying the '[A:a] condition', and thus admitting for their members the 'normal [A:a] expansion', (see §8). It should be mentioned that in 1920-21 Professor Oswald Veblen urged the writer to replace the truth-table solution by one which represented logical functions by algebraic polynomials modulo 2 (the arithmetical forms of B. A. Bernstein [2], whose importance has since been emphasized by Stone [28] in connection with his researches on Boolean algebras. For their connection with Veblen see [32], P. 9). However, it then seemed to the writer, and still seems so, that while the derivation of what we have called the alternating systems (§15) would be greatly simplified thereby, all of the rest of the solution would become correspondingly more complicated.

version essentially to the statement of our problem. The next three sections develop those aspects of the technique of the Jevons notation which are required in the solution of the problem. The last four sections of Part I form an integral part of that solution, but are included among the preliminaries to expedite the presentation of that solution.

Part II is devoted to the complete solution of what we have called our main problem. The various closed systems of membership functions of classes are presented either individually, or, in the case of the infinite families of closed systems, with aid of a parameter. In general we may say that each closed system is given by a condition on its members. These conditions are such that mere inspection is sufficient to tell of a given expression whether the corresponding function satisfies that condition. They are easily retranslated in terms of truth-tables.

The first of the two sections comprising Part III unifies the set of all closed systems of membership functions, or truth-tables, with respect to the relation of inclusion. By analysing this relation in terms of immediate inclusion, we are led to the inclusion diagram, which gives a graphical summary of the solution of our main problem. It should be noted that even this revised version of the section was written before the christening, and recent development, of the lattice concept.¹⁰ The last section gives our solution of the other of the two problems posed above. This problem was also solved in the first version of the paper under the added condition that the generators be independent, and then further specialized in such a way that the infinite number of independent sets of generators reduced to 36 essentially different sets. Actually, in the light of the solution of our main problem, the solution of the secondary problem without the condition of independence is shown in the present version to be very simple. A different and far easier solution than in the first version is here obtained of the problem of independent generators, but with corresponding added complications in the derivation of the specialized result. That the same 36 essentially different sets are then obtained is but a check on both solutions.

¹⁰ See for example, [5].

The writer is aware of but two papers directly touching on our two problems. Both papers restrict themselves to second order functions, i.e., functions of but two variables, not only as generators but as functions generated. There being but sixteen such functions, the doubly infinite character of our problem thus becomes completely finite, and manageable so. In 1925, E. Zylinski [42] specifically determined the sets of second order functions generated by single second order functions. In a recent investigation, W. Wernick [36] discusses the sets of second order functions generated by arbitrary sets of second order functions, and specifically determines what we would call the various sets of second order functions which are sets of independent generators of the sixteen second order functions, and hence, indeed, of all functions. Since our general solution of the problem of independent generators can but be given by conditions on those generators, while our specialization rules out certain sets even consisting of second order functions, these results may be said but to overlap our own. We might note that had Zylinski seen our Elementary Propositions, published four years earlier, he would have found the question ending his paper answered in the affirmative, the conjecture immediately preceding disproved.

In the second part of Elementary Propositions will be found the following paragraph. "We thus see that complete systems are equivalent to the system of 'Principia' not only in the truth table development but also postulationally. As other systems are in a sense degenerate forms of complete systems, we can conclude that no new logical systems are introduced". Perhaps the essentially new logics yielded by our postulational generalization blotted out our earlier idea that these subsystems of the complete system would constitute a sort of logical playground for the construction of what are now called propositional calculi. This earlier vision was renewed in the observation that such a system as that generated by implication, if admitting a postulational development making it (postulational) 'closed', would not be 'completely closed' — to use concepts of the third part of Elementary Propositions. Actually, quite a number of these systems should admit a postulational development in the ordinary sense. And, indeed, through developments of Łukasiewicz

wicz, Tarski, Bernays, Léśniewski, Mihailescu, and Wajsberg, four of these systems may be said to have been so treated. In particular, we may refer to developments in terms of equivalence only ([14], [33], and [16]),¹¹ which generates the system we symbolize L_2 , (§15), equivalence and negation [17], and equivalence and its dual [18], each pair of which functions generates L_1 , (§15),¹² implication ([34], [10]), which generates F_4^∞ , (§22), and implication and conjunction [10] and perhaps equivalence and disjunction [19], among others, which generate the important C_2 , (§19).¹³ We might add that where an ordinary postulational development is not possible, a complete logic might be given for a system in other ways.¹⁴

While our main problem and its solution, retranslated in terms of truth-tables, fits hand in glove with developments of two-valued propositional calculi based on a finite number of primitive truth-functions, difficulties in formulating a concept

¹¹ In connection with these developments, and also the two immediately succeeding, see [24].

¹² The reviewers' observation that not all enunciations in terms of these primitives capable of proof in ordinary logic can be deduced here makes it questionable whether we can say that the postulates are for system L_1 . Though here it may be just a case of incompleteness, in other instances it may be that not the two-valued logic is under consideration.

¹³ We may say that as system C_1 is the concern of the complete two-valued calculus of propositions, system C_2 is the concern of the 'positive logic', ([9], p. 68). In this connection C_3 , the dual of C_2 , is distinguished by the fact that under the functions of classes interpretation it consists of all such functions whose values, for given classes, are independent of the particular universal class containing the given classes. Returning to the postulational developments, we must admit that the intent of the authors is not to develop a postulational development for a given system via certain primitives, but merely to study the inter-relations of those primitives among themselves. Thus, in [10] again, we find a study of expressions built up by implication, conjunction, and negation, while either of the first two primitives with the last generate the complete two-valued system.

¹⁴ Thus, in the duals of the first class of systems a 'contradictory logic' is possible (see [11]), while in all cases it may be possible to set up a complete logic solely by means of rules of derivation (see [20]).

of identity of truth-functions, as against mere equivalence, prevents us from giving even a definitive formulation of the problem of determining all iteratively closed systems of truth-functions. That our solution of the corresponding problem for truth-tables may here be of little help is indicated by the following specialization of the general problem. In the propositional calculus of *Principia Mathematica* let two truth-functions be considered identical when and only when their expressions in terms of the primitives $\sim p$ and $p \vee q$ are identical. The concept of an iteratively closed system of (\sim, \vee) truth-functions is then precise. But while to each closed system of (\sim, \vee) truth-functions there corresponds the closed system of corresponding truth-tables, the problem of determining all such closed systems of truth-functions is obviously entirely independent of the corresponding problem for truth-tables, being a problem in symbol structure only. A solution of this problem in the same sense as that of the present paper is hardly to be expected, since the cardinal number of (\sim, \vee) closed systems is that of the continuum.¹⁵ Indeed, we hazard the guess that in some sense the problem is unsolvable. In particular, it seems likely that the problem of determining for any two finite sets of (\sim, \vee) truth-functions whether the systems they generate have a function in common is unsolvable in the sense of Church [7]. We may finally note that unlike iteratively closed systems of truth-tables, an iteratively closed system of (\sim, \vee) truth-functions need not be capable of generation by a finite set of generators.¹⁶

¹⁵ In fact, even those consisting of functions of one variable. For obviously such a function may have an arbitrary number of occurrences of that one variable. Furthermore, if $f(p)$ has m occurrences of $p, g(p), n, f[g(p)]$ has mn occurrences of p . It follows that for any given subset, finite or infinite, of the set of primes $2, 3, 5, 7, \dots$ the set of (\sim, \vee) functions of one variable for which the number of occurrences thereof factors into primes in the subset is iteratively closed.

¹⁶ This follows from the number of (\sim, \vee) truth-functions being denumerably infinite, of closed (\sim, \vee) systems non-denumerably infinite. In particular, each of the closed systems of the preceding footnote corresponding to an infinite number of primes is incapable of being generated by a finite set of generators. For a finite number of such functions will, by the reasoning of that footnote, at most generate a system corresponding to a finite number of primes. In the terminology of §1, we may thus say that not every iteratively closed (\sim, \vee) system is of finite

On the other hand, the m -valued logics of the fourth part of Elementary Propositions do suggest a natural extension of our present problem. Added interest is lent to such a possible extension by the fact that the m -valued logic of Łukasiewicz and Tarski seems to correspond to but a proper subset of the set of all possible m -valued truth-tables.¹⁷ In particular, for three-valued logics the corresponding problem of determining all iteratively closed three-valued systems of functions should be a feasible one. The immediate chances for a complete solution of the m -valued problem are small, for, as has been observed by Wiener in different phraseology [40], among the iteratively closed systems of m -valued functions of one variable are included all substitution groups of degree m .¹⁸ However, the various conditions defining the iteratively closed two-valued systems of our solution easily generalize, and in more than one way, to give certain iteratively closed m -valued systems. And it may be feasible to obtain a general solution of the m -valued problem for arbitrary functions in terms of the corresponding problem for functions of one variable.¹⁹

order. This continues to be so even with 'finite order' generalized in the sense of footnote 24 (§1). For the set of all (\sim, \vee) functions of the form $\sim P$ is clearly iteratively closed. By letting P involve \vee 's only, we can have such a function on an arbitrary number of arguments with but one occurrence of \sim . On the other hand, such functions of n or fewer arguments can only generate functions of more than n arguments which involve at least two \sim 's.

¹⁷ See [35].

¹⁸ The substitution groups correspond rather to the contracted closed systems (see §1). In the same sense, the iteratively closed m -valued systems of functions of one variable in their entirety are identical with the generalized substitution groups of degree m in the sense of Suschkewitsch [29].

¹⁹ In this connection, see §10.

Part I

PRELIMINARIES

§1

ITERATIVE CLOSEDNESS AND GENERATION

We briefly term function what Whitehead and Russell call 'the ambiguous value of a function' which, in turn, is essentially the 'dependent variable' of mathematics. We postulate a class of independent variables in terms of which each function is understood to be ultimately expressed. A particular method of defining a function of a given set of independent variables will be called an operation on those variables. Functions yielded by different operations on the same set of independent variables are then considered to be identical if for all admissible values of those variables the corresponding values of the functions are always identical.

Let V be a denumerably infinite class of independent variables x_1, x_2, x_3, \dots , and let F be an existent class, or, as we shall say, system of functions whose arguments, finite in number and existent for each function, are members of V .²⁰ Let the variables in V range in value over one and the same class of values v , and let the corresponding values of the functions in F also belong to v . If X_1, X_2, \dots, X_n represent variables in V or functions in F , those variables and functions will in their totality depend on a finite number of distinct variables $x_{j_1}, x_{j_2}, \dots, x_{j_m}$, belonging to V . If then $f(x_{j_1}, x_{j_2}, \dots, x_{j_m})$ is also a function in F , on distinct variables $x_{j_1}, x_{j_2}, \dots, x_{j_m}$, $f(X_1, X_2, \dots, X_n)$ will represent a function of $x_{j_1}, x_{j_2}, \dots, x_{j_m}$. This function will be said to be obtained from $f(x_{j_1}, x_{j_2}, \dots, x_{j_m})$ by replacing $x_{j_1}, x_{j_2}, \dots, x_{j_m}$ by X_1, X_2, \dots, X_n respectively. And the expression $f(X_1, X_2, \dots, X_n)$, which not only represents this function but indicates the manner in which it is to be obtained, will be referred to as an iterative process depending on the variables $x_{j_1}, x_{j_2}, \dots, x_{j_m}$ and yielding the function in

²⁰ See footnote 5 (Introduction).

question. We now define F to be iteratively closed with respect to V if replacing the arguments of a function in F by variables in V or functions in F always results in a function in F , i.e., if whenever $f(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ is a function in F on distinct variables $x_{i_1}, x_{i_2}, \dots, x_{i_n}$, and X_1, X_2, \dots, X_n represent variables in V or functions in F , then the function yielded by the iterative process $f(X_1, X_2, \dots, X_n)$ is also in F . A system of functions iteratively closed with respect to V will also be referred to as an iteratively closed system over V .²¹

A function of n distinct independent variables will be said to be of order n . We then immediately verify that an iteratively closed system F either consists wholly of first order functions or else possesses functions of every finite order.

For if F does not consist wholly of first order functions it will possess some function $f(x_{i_1}, x_{i_2}, \dots, x_{i_m})$ of order $m > 1$. Then, by the closedness definition, the first order function $f(x_{j_1}, x_{j_2}, \dots, x_{j_1})$ belongs to F , and if the n -th order function $g(x_{j_1}, x_{j_2}, \dots, x_{j_n})$ is in F , and $x_{j_{n+1}}$ is distinct from $x_{j_1}, x_{j_2}, \dots, x_{j_n}$, $f[g(x_{j_1}, x_{j_2}, \dots, x_{j_n}), x_{j_{n+1}}, \dots, x_{j_{n+1}}]$ will be an $(n+1)$ -th order function in F . The result follows by mathematical induction.

Our proof actually shows that if a closed system F does not consist wholly of first order functions, then it possesses at least one function of each finite set of variables belonging to V . Now if $x_{i_1}, x_{i_2}, \dots, x_{i_n}$, and $x_{j_1}, x_{j_2}, \dots, x_{j_n}$ are any two sets of n distinct variables in V , then, if $f(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ is in F , $f(x_{j_1}, x_{j_2}, \dots, x_{j_n})$ will also be in F , and conversely. It easily follows that the totality of functions in F on the first set of n variables is transformed in 1-1 fashion into the totality of functions in F on the second set of n variables by any 1-1 replacement of the variables in the first set by the variables in the second set. We likewise see that a closed F which does consist wholly of first order functions

²¹ Or iteratively closed system, or closed system.

possesses at least one function of each single variable in V ; and the totality of functions in F on one variable is transformed in 1-1 fashion into the totality of functions in F on any other variable by replacing the first variable by the second.

We may therefore say that the functions in F on any one set of n variables are merely repeated over all other sets of n variables belonging to V . This redundancy can be overcome as follows. Let V be simply ordered in the series x_1, x_2, x_3, \dots , and let \mathfrak{F} be composed of those functions in F whose arguments are $(x_1), (x_1, x_2), (x_1, x_2, x_3)$, etc. Then \mathfrak{F} possesses all of the essentially different functions to be found in F without the latter's duplications.²² \mathfrak{F} will be called the contracted closed system corresponding to F . Note that when F consists wholly of first order functions, \mathfrak{F} possesses only functions of x_1 ; otherwise, \mathfrak{F} possesses at least one function of (x_1, x_2, \dots, x_n) for every positive integral n . We shall often implicitly use the concept of a contracted closed system for the purpose of exhibiting the functions possessed by a closed system. On the other hand, the replacement process, which is the basis of our concept of iterative closedness, demands the very duplication of the 'same functions' over different sets of variables found in closed systems, and deleted therefrom in forming contracted closed systems. Our derivations will therefore explicitly refer to closed systems as originally defined.²³

²² Ultimate economy would be achieved by forming the set of all 'abstract functions' corresponding to functions in F , two functions being said to correspond to the same abstract function if they can be made identical by a 1-1 replacement of arguments. The resulting 'abstract closed system' corresponding to F is, however, not as practical as \mathfrak{F} . It would carry us too far afield to relate these ideas to the concept $f(x_1, x_2, \dots, x_n)$ of Principia Mathematica.

²³ The requirement that V be a denumerably infinite class is in agreement with the usual form of a propositional calculus. Note that were W any infinite class of variables, G an iteratively closed system over W , then, if each function in G is on a finite number of arguments, a system F iteratively closed with respect to a denumerably infinite class V could be constructed such that the abstract closed systems, in the sense of the last footnote, corresponding to G and F are identical. If then V be infinite, it may as well be denumerably infinite. V finite, however, leads to something new. Note that if F is iteratively closed with respect to V , then if V' is an existent subclass of V , F' the class of functions in F with ar-

We give an independent definition of the system of functions generated by a given existent finite set of functions. Variables and values are to be in V and v as heretofore. There is no loss of generality in writing the functions that are to generate the system in the form $f(x_1, x_2, \dots, x_{n_i})$, $i = 1, 2, \dots, v$, $n_i \neq 0$. We then define the system of functions generated by these generators over the given denumerably infinite class of variables V as the class of all functions that can be assigned to the system by the use of the following criterion. If x_1, x_2, \dots, x_{n_i} , $i = 1, 2, \dots, v$, represent variables in V or functions in the generated system, then the function represented by $f_1(x_1, x_2, \dots, x_{n_i})$ also belongs to the generated system. We shall say also that every function thus obtainable can be generated by the given generators. Note that in the initial applications of this inductive definition x_1, x_2, \dots, x_{n_i} can only be variables in V .

It follows from this definition that each function in the generated system can be expressed in finite form in terms of the generators of the system, and variables in V . These expressions, or operations as they may be called, can be classified according to their rank. It is convenient to refer to variables in V as operations of rank zero. Then if the highest rank of the operations x_1, x_2, \dots, x_{n_i} is p , the operation $f_1(x_1, x_2, \dots, x_{n_i})$ will be said to be of rank $p+1$. We also inductively define the components of $f_1(x_1, x_2, \dots, x_{n_i})$ to be x_1, x_2, \dots, x_{n_i} and their components, a variable in V being understood to have no components. Clearly, an operation of rank p has at least one component of each rank less than p . Thus, with $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$ as generators, $f_1[f_2(x_1, x_2), f_1[f_2(x_3, x_1), x_2]]$ is an operation of rank three; and among its components are the operations $f_1[f_2(x_3, x_1), x_2], f_2(x_3, x_1), x_3$ of ranks two, one, zero, respectively. Associated with the generated system of functions is thus a system of operations. Each operation of rank greater than zero in the latter system defines a unique function in the former. On the other hand, for most of the generated systems

guments in V' , then F' is iteratively closed with respect to V' . This result easily transforms the solution of the principal problem of the present paper, given for infinite V , into solutions of the same problem for finite V 's. In this connection see also the next footnote.

that we shall study, each function in the former system is yielded by an infinite number of different operations in the latter.

By means of these operations we readily verify that if $f(x_{11}, x_{12}, \dots, x_{1n})$ is a function in a generated system on distinct variables $x_{11}, x_{12}, \dots, x_{1n}$, and X_1, X_2, \dots, X_n represent variables in V or functions in the system, then the function represented by $f(X_1, X_2, \dots, X_n)$ also belongs to the system. That is, every generated system is iteratively closed. Clearly, if each function in a given set of generators belongs to a closed system F , then the system of functions generated by those generators is contained in F . Hence, if in addition, each function in F can be generated by those generators, the generated system is F . There is no a-priori reason, however, for an arbitrary closed system admitting of generation by a finite set of generators.

We define the order of a finite set of generators as the highest order of the several functions in the set. If, then, a closed system can be generated by a finite set of generators, we define the order of the system as the lowest order that a finite set of generators of the system can have. The phrase, 'closed system of finite order' is then equivalent to the phrase 'generated system' provided the number of generators is understood to be finite.²⁴ Since functions of one variable can only generate functions of one variable, it follows that a closed system of the first order consists wholly of functions of the first order. On the other hand, a closed system of finite order greater than one must possess at least one function of order greater than one, and hence possesses functions of every order.

²⁴ The concept of generated system, however, continues to be valid if an infinite number of generators is allowed. This less stringent concept leads to a correspondingly wider concept of 'closed system of finite order'. The following observation assumes this wider concept. Clearly, a closed system over V of finite order n is determined by its n -th and lower order functions. It follows that if V' consists of say the first n variables in V , then there is thus determined a 1-1 correspondence between the closed systems over V of order less than or equal to n and the closed systems over V' in the sense of the preceding footnote. See also footnote 39 (§10). Where, as in the problem of the present paper, there are but a finite number of functions on a given finite set of arguments, there is no difference between the two concepts of 'closed system of order n '.

Continuing to restrict our attention to functions whose arguments are in V and values in v , we define a condition imposed upon such functions to be iterative²⁵ if whenever $f(x_{1j}, x_{12}, \dots, x_{1n})$, on distinct variables $x_{1j}, x_{12}, \dots, x_{1n}$, satisfies the condition, and X_1, X_2, \dots, X_n represent variables in V , or functions satisfying the condition, then the function represented by $f(X_1, X_2, \dots, X_n)$ also satisfies the condition.

Clearly, if each of several conditions is iterative, then the combined condition that a function satisfies each of those conditions is iterative.

There are two important consequences of a condition being iterative. On the one hand, the system of all functions satisfying the condition is iteratively closed. On the other, if each function in a given set of generators satisfies the condition, then every function in the generated system satisfies the condition. We shall in every case say that if each function in a closed system satisfies a certain iterative condition, then the system satisfies that condition. The discovery of iterative conditions thus plays a very important rôle in the determination of iteratively closed, and generated, systems of functions.

§2

TWO-VALUED SYSTEMS OF FUNCTIONS and THE RELATED LOGIC OF CLASSES

Variables, functions, and systems of functions for which the class of values v consists of but two members will be said to be two-valued. The following discussion is restricted to a single class v whose two members are symbolized + and - respectively.

Our definition of the identity of functions on the same set of variables implies the adoption of the extensional point of view with respect to functions. Every two-valued function,

²⁵ We have borrowed the word "iterative" in these connections from Wiener (see footnote 6). But our "iterative condition" is not the same concept as Wiener's "iterative characteristic" [41].

$\phi(\beta, \delta, \dots, \eta)$ can therefore be uniformly exhibited in finite form by a table such as

$\beta, \delta, \dots, \eta$	$\phi(\beta, \delta, \dots, \eta)$
++...+	+
++...-	-
....	.
.....	.
....-	.
--...-	+

which associates with every possible assignment of values to the variables $\beta, \delta, \dots, \eta$ the corresponding value of the function. Conversely, every such table defines a two-valued function of $\beta, \delta, \dots, \eta$. Since the two values +, -, can be assigned to n variables $\beta, \delta, \dots, \eta$ in 2^n different ways, it follows that there are exactly 2^{2^n} different two-valued functions of those n variables.

Let V be a given denumerably infinite set of variables $\alpha, \beta, \gamma, \dots$. For each set of n variables belonging to V , with $n = 1, 2, 3, \dots$, construct the 2^{2^n} two-valued functions of those variables. The system composed of all of these two-valued functions will be called the complete two-valued system over V .

Clearly every iterative process built up out of two-valued variables and functions yields a two-valued function. It follows that the complete two-valued system over V is iteratively closed with respect to V .

The main problem of the present paper is to determine all of the closed two-valued systems over V . Since each of these closed systems is composed of functions belonging to the complete system, our problem can be restated as follows: to determine all of the closed subsystems over V of the complete two-valued system over V .

These systems are more easily visualized through their associated contracted systems. The 'complete contracted two-valued system over V ' possesses exactly 2^{2^n} functions of order n , i.e., it is composed of 4 functions of the first order, 16 of the

second order, 256 of the third order, and so on. A contracted closed two-valued system over V consisting wholly of first order functions will be a subset of the set of four first order functions in the complete contracted system. Every other contracted closed two-valued system over V , on the other hand, will possess at least one of the 4 first order functions, at least one of the 16 second order functions, at least one of the 256 third order functions, and so on, that make up the complete contracted system. It may therefore be said to cut out an infinite swathe of functions from the complete contracted system.²⁶

For reasons of presentation we replace the above abstract problem by an equivalent problem in the logic of classes. Let v' be the class of subclasses of a given existent class.²⁷ v' then has at least two members, the given existent class, which will be called the universal class and symbolized 1, and the null class, symbolized 0. Let B, D, \dots, H be a set of n independent variables whose admissible values are arbitrary classes in v' , and let $f(B, D, \dots, H)$ be a function of those variables with its values also in v' . Among such functions we shall then only be interested in those functions $f(B, D, \dots, H)$ below termed membership functions.

Let x represent an arbitrary member of the universal class, G a class belonging to v' . We define the membership value of x with respect to G to be + when x is a member of G , - when x is not a member of G . Each admissible assignment of values to the variables x, B, D, \dots, H thus determines the membership values of x with respect to B, D, \dots, H , and $f(B, D, \dots, H)$. We then define $f(B, D, \dots, H)$ to be a membership function of B, D, \dots, H if the membership value of x with respect to $f(B, D, \dots, H)$ is found to depend solely on the membership values of x .

²⁶ Actually these contracted systems are best visualized as the different tables of the functions in the corresponding two-valued systems when the headings of those tables are removed.

²⁷ For simplicity. It of course suffices for v' to be a class of subclasses of the given existent class closed with respect to the operations logical sum, and complement with respect to the given class.

with respect to B, D, \dots, H .²⁸ Now each of the 2^n ways in which membership values can formally be assigned to x with respect to the n variables B, D, \dots, H will be exhibited by actual values of x, B, D, \dots, H . For we can let x be any member of the universal class, and B, D, \dots, H the universal or null class, according as the corresponding membership value is desired to be + or -. It follows that every membership function $f(B, D, \dots, H)$ determines the membership table of the function, e.g.,

<u>B, D, ..., H</u>	<u>$f(B, D, \dots, H)$</u>
+	+
+	-
.	.
.	.
.	.
-	+

which gives the membership value of x with respect to $f(B, D, \dots, H)$ for each of the 2^n ways in which x may assume membership values with respect to B, D, \dots, H .²⁹ Conversely, every membership table on B, D, \dots, H determines a membership function $f(B, D, \dots, H)$ of which it is the membership table. For, given B, D, \dots, H , such a table tells of any x in the universal class that, according as it does or does not belong to the several classes B, D, \dots, H , it does or does not belong to the class, $f(B, D, \dots, H)$, and thus determines the class $f(B, D, \dots, H)$. A 1-1 correspondence has thus been established between the different membership functions of B, D, \dots, H , and the membership tables on B, D, \dots, H . Evidently there are 2^{2n} different membership tables on B, D, \dots, H . It follows that there are exactly 2^{2n} different membership functions of the n variables B, D, \dots, H .

This suggests a close relationship between two-valued functions and membership functions, which may be formulated precisely

²⁸ We cannot say 'Boolean' since the latter may involve in its definition constants in 'v' other than 1 and 0.

²⁹ Note that if the given universal class has at least 2^n members, then a single suitable choice of B, D, \dots, H can be made with respect to which x will assume all of the 2^n possible sets of membership values as it roams through the universal class. The one corresponding class $f(B, D, \dots, H)$ suffices, then, to determine the membership function $f(B, D, \dots, H)$. Hence, in part, the efficacy of the Venn diagrams.

ly as follows. Let n two-valued variables $\beta, \delta, \dots, \eta$ be set into 1-1 correspondence with the n membership variables B, D, \dots, H . Then, if in the table of a two-valued function $\phi(\beta, \delta, \dots, \eta)$ the variables $\beta, \delta, \dots, \eta$ are replaced by the corresponding variables B, D, \dots, H , the result can be interpreted as the membership table of a membership function $f(B, D, \dots, H)$, and conversely. Through their tables a 1-1 correspondence is thus set up between the 2^{2n} two-valued functions of $\beta, \delta, \dots, \eta$, and the 2^{2n} membership functions of B, D, \dots, H .

As in the case of two-valued systems, we let V' be a given denumerably infinite set of variables A, B, C, \dots , and define the complete membership system over V' as the system composed of all membership functions whose arguments, finite in number for each function, belong to V' . Here, as below, it is understood that the variables and membership functions in question all correspond to the above class of values v' . That the complete membership system over V' is iteratively closed with respect to V' follows from the more general result proved below. We then propose the problem of determining all iteratively closed 'membership systems' over V' , and observe that this is equivalent to determining all of the closed subsystems over V' of the complete membership system over V' .

Now set the denumerably infinite set of variables $\alpha, \beta, \gamma, \dots$ forming V into 1-1 correspondence with the denumerably infinite set of variables A, B, C, \dots forming V' . By setting up the above correspondence between two-valued functions and membership functions for each finite set of variables in V , and finite set of corresponding variables in V' , a 1-1 correspondence is established between the functions in the complete two-valued system over V and the functions in the complete membership system over V' . This correspondence furthermore is preserved under iteration. For let the iterative process $\phi(\mu, v, \dots, w)$, depending on variables $\beta, \delta, \dots, \eta$ in V , correspond to the iterative process $f(M, N, \dots, W)$, depending on the corresponding variables B, D, \dots, H in V' , i.e., let the two iterative processes be similarly built up out of variables and functions that correspond. Then, if membership values are assigned to x with respect to B, D, \dots, H which are respectively the same as values assigned to $\beta, \delta, \dots, \eta$, membership values of x with respect to

M, N, \dots, W will be determined which are respectively the same as the resulting values of μ, v, \dots, w , and hence a membership value of x with respect to $f(M, N, \dots, W)$ will be determined which is the same as the value $\phi(\mu, v, \dots, w)$. The iterative process $f(M, N, \dots, W)$ therefore yields a membership function of B, D, \dots, H which corresponds to the two-valued function of $\beta, \delta, \dots, \eta$ yielded by the corresponding iterative process $\phi(\mu, v, \dots, w)$.

Given then any closed two-valued system over V , the membership functions corresponding to the two-valued functions therein must constitute a closed membership system over V' , and conversely. A 1-1 correspondence is thus induced between the aggregate of all closed two-valued systems over V and the aggregate of all closed membership systems over V' . In terms of this correspondence a solution of the problem of determining all closed two-valued systems over V automatically constitutes a solution of the problem of determining all closed membership systems over V' , and conversely. As the problem for membership systems will allow us to use the notation and algebra of symbolic logic, we shall confine our attention to its solution, and consider the result a solution of the problem for two-valued systems.³⁰

§3

THE JEVONS NOTATION; EXPANSIONS OF MEMBERSHIP FUNCTIONS

Since our mathematical formulation of membership function does not quite see eye to eye with the usual developments of symbolic logic, we go to some length to bridge the resulting gap. Our concern is not with formal symbolic logic, but with its interpretation as a logic of classes. In this discipline we may be concerned with the formal structure of an expression

³⁰ Indeed our formulas for the functions in the various closed membership systems, when considered merely in connection with the about to be defined complete expansions of those functions, can be immediately translated into conditions on the membership tables of those functions, and hence automatically into conditions on the tables of the functions in the various closed two-valued systems. This is illustrated in footnote 37, (§8).

$f(A, B, \dots, I)$, or, when this expression is meaningfully interpreted, we may refer to the function $f(A, B, \dots, I)$, and, when A, B, \dots, I are assumed to take on certain classes in v' as values, our concern may then be with the corresponding class $f(A, B, \dots, I)$, also in v' . If, for example, $f(A, B, \dots, I)$ is the null class 0 for every assignment to A, B, \dots, I of values in v' , we shall say that $f(A, B, \dots, I)$ is identically 0, and write $f(A, B, \dots, I) = 0$. Our use of the word 'contain' will always refer to inclusion of classes represented, and not to the formal structure of corresponding representations. Exactly, if $f(A, B, \dots, I)$ and $g(A, B, \dots, I)$, with not all arguments necessarily present in each expression, are such that for each assignment of values in v' to A, B, \dots, I the class $f(A, B, \dots, I)$ contains the class $g(A, B, \dots, I)$, then we shall say that the function $f(A, B, \dots, I)$, or even the expression $f(A, B, \dots, I)$, contains the function, or expression, $g(A, B, \dots, I)$. If, on the other hand, for each assignment of values in v' to A, B, \dots, I the classes $f(A, B, \dots, I)$ and $g(A, B, \dots, I)$ have no members in common, we shall say that $f(A, B, \dots, I)$ excludes $g(A, B, \dots, I)$.³¹

We define the negative of an arbitrary class belonging to v' as the class which consists of those members of the universal class which are not members of the given class. With logical sum and logical product defined as usual, it follows that the logical sum of a class and its negative is always the universal class 1, the logical product, the null class 0. Furthermore, the negative of 1 is 0, of 0, 1.

Following Jevons we introduce the small letters a, b, c, \dots to represent the negatives of the classes represented by the corresponding capital letters A, B, C, \dots . By a complete term on n given letters A, B, \dots, I we shall mean the formal logical product of the letters belonging to any selection chosen from the n pairs of letters $(A, a), (B, b), \dots, (I, i)$. Thus $AbcDe$ is a complete term on the five letters A, B, C, D, E . For the n letters A, B, \dots, I there are 2^n such selections, and hence, provided we disregard the particular order of their 'factors', there are 2^n complete terms on A, B, \dots, I . They constitute the

³¹ The paragraph just concluded is a recent addition. Likewise a few minor changes in the rest of this section.

logical alphabet of Jevons for those letters.³²

If $t(A,B,\dots,I)$ represents a complete term on A,B,\dots,I , then each assignment of classes in v' as values of A,B,\dots,I results in a value of $t(A,B,\dots,I)$ which is also a class in v' . It is then significant that $t(A,B,\dots,I)$ is never 0 for all values of A,B,\dots,I . In fact, by letting the variables A,B,\dots,I assume the values 1 or 0 according as the corresponding letters are capital or small in $t(A,B,\dots,I)$, $t(A,B,\dots,I)$ assumes the value 1. The relationship between a class and its negative yields the formal identities $A+a = 1$, $Aa = 0$. Since for any two distinct complete terms on A,B,\dots,I there is at least one letter that is capital in one term and small in the other, it follows that:

The logical product of any two distinct complete terms on A,B,\dots,I is identically 0. On the other hand, by expanding the first member of the identity $(A+a)(B+b)\dots(I+i) = 1$, we see that the logical sum of all the complete terms on A,B,\dots,I is identically 1.

We have, of course, assumed here the ordinary formal laws of logical sum and logical product. Likewise in later developments.

Given a membership function $f(A,B,\dots,I)$, if we represent the class $f(A,B,\dots,I)$ by M , the negative of $f(A,B,\dots,I)$ may be represented by m . The membership table of $f(A,B,\dots,I)$ can then be rewritten in the form

A	B	...	I	M
A	B	...	i	m
.
.
.
a	b	...	i	M

where the last column assigns an arbitrary member of the universal class to M or m according as it belongs to A or a , B or b , ... I or i . From this table we see that if $t(A,B,\dots,I)$ is a complete term on A,B,\dots,I , then either the class $t(A,B,\dots,I)$ is contained in the class M for all values of A,B,\dots,I , or the class $t(A,B,\dots,I)$ is contained in the class

³² See, for example, [12].

m for all values of A, B, \dots, I . That is, we have identically in A, B, \dots, I either $t(A, B, \dots, I)M = t(A, B, \dots, I)$ or $t(A, B, \dots, I)m = t(A, B, \dots, I)$. Since $Mm = 0$, we also have in the first case $t(A, B, \dots, I)m = 0$, in the second $t(A, B, \dots, I)M = 0$.

Hence the complete term $t(A, B, \dots, I)$ is either contained in M and excluded from m , or contained in m and excluded from M .

As $t(A, B, \dots, I)$ cannot be identically 0, these two alternatives are mutually exclusive. Now let $\epsilon_1(A, B, \dots, I)$ be the formal logical sum of the complete terms on A, B, \dots, I contained in M , $\epsilon_2(A, B, \dots, I)$ of those contained in m . By multiplying both sides of the identity $AB\dots I + AB\dots i + \dots + ab\dots i = 1$ by M , and again by m , we see that $\epsilon_1(A, B, \dots, I)$ is identically equal to M , $\epsilon_2(A, B, \dots, I)$ to m .

The ordered pair of expressions $\epsilon_1(A, B, \dots, I): \epsilon_2(A, B, \dots, I)$ will be called the complete expansion of the membership function $f(A, B, \dots, I)$. The first expression, $\epsilon_1(A, B, \dots, I)$ of the complete expansion of $f(A, B, \dots, I)$ is then a formal representation of $f(A, B, \dots, I)$, the second, $\epsilon_2(A, B, \dots, I)$, of the negative of $f(A, B, \dots, I)$. If, on the other hand, we do not start with a given membership function of A, B, \dots, I , but arbitrarily form a pair of expressions $\epsilon_1(A, B, \dots, I): \epsilon_2(A, B, \dots, I)$ by assigning each of the 2^n complete terms on A, B, \dots, I to one or the other of these expressions, the result will be called a complete expansion on the n letters A, B, \dots, I . Thus $ABC+ABc+AbC+aBC: abc+abC+aBc+Abc$ is a complete expansion on the three letters, A, B, C . Clearly there are exactly 2^{2n} complete expansions on the n letters A, B, \dots, I provided we disregard the particular order of the terms within each expression of a complete expansion. With this proviso, we see that each membership function of A, B, \dots, I determines its complete expansion. Conversely, each complete expansion on A, B, \dots, I determines a membership table on A, B, \dots, I , and hence a membership function of A, B, \dots, I of which it is the complete expansion. A 1-1 correspondence is thus set up between the 2^{2n} membership functions of A, B, \dots, I , and the 2^{2n} complete expansions on A, B, \dots, I .

A complete term on any proper subset of a given set of let-

ters A, B, \dots, I will be called an incomplete term on A, B, \dots, I . Thus, bDe is an incomplete term on the five letters A, B, C, D, E . Clearly, an incomplete term on A, B, \dots, I also assumes a value in v' for each assignment of values in $v!$ to A, B, \dots, I . Given an incomplete term, and any complete term, on A, B, \dots, I , then either each of the letters present in the incomplete term is capital in both terms or small in both terms, or at least one letter is capital in one term and small in the other. In the first case the logical product of the two terms is the complete term, in the second the logical product is identically 0.

That is, any complete term on A, B, \dots, I is either contained in a given incomplete term on A, B, \dots, I or is excluded therefrom.

Clearly, every complete term on A, B, \dots, I contained in a given incomplete term on A, B, \dots, I is the logical product of the incomplete term and a complete term on those letters A, B, \dots, I which are missing from the incomplete term, and conversely. Since the logical sum of all the complete terms on those missing letters is identically 1, we see by multiplying this sum by the incomplete term that:

An incomplete term on A, B, \dots, I is identically equal to the logical sum of all the complete terms on A, B, \dots, I contained therein.

An ordered pair of expressions such that each is the formal logical sum of terms, complete or incomplete, on A, B, \dots, I , and such that either expression is identically equal to the negative of the other, will be called an expansion on the letters A, B, \dots, I . Given a membership function of A, B, \dots, I , $f(A, B, \dots, I)$, then an expansion on A, B, \dots, I will be called an expansion of the membership function provided the first expression of the expansion is identically equal to $f(A, B, \dots, I)$, and hence the second to the negative of $f(A, B, \dots, I)$. Thus, the logical sum of A and B , which is a membership function of A and B with the complete expansion $AB+Ab+aB: ab$ is easily seen to

admit the other expansions $A+aB:ab$, $Ab+B:ab$, $A+B:ab$.

The condition that one expression of an expansion on A, B, \dots, I is identically equal to the negative of the other is equivalent to the pair of conditions: the logical sum of the two expressions is identically 1, the logical product identically 0.

It follows that every complete term on A, B, \dots, I is contained in at least one term of an expansion on A, B, \dots, I ; as otherwise it would be excluded from every term of the expansion, and the logical sum of the two expressions of the expansion would not be 1. Furthermore, a complete term on A, B, \dots, I cannot be contained in terms belonging to different expressions of an expansion on A, B, \dots, I , or the logical product of the two expressions would not be 0.

If then each term of an expansion on A, B, \dots, I is replaced by the logical sum of all the complete terms on A, B, \dots, I contained therein, and identical terms are then united, the result will be a complete expansion on A, B, \dots, I whose first and second expressions are identically equal to the first and second expressions respectively of the given expansion. Since every complete expansion on A, B, \dots, I defines a membership function of A, B, \dots, I , we thus see that every expansion on A, B, \dots, I is an expansion of a membership function of A, B, \dots, I .

In the above general discussion there occur certain lacunae which we now fill in. We have described the expressions of both complete expansions, and expansions in general, to be formal logical sums of terms, whereas the number of terms to be assigned to an expression may be one or zero. In the first case the expression in question must clearly be the term itself. In the second case we can consistently write 0 for the expression. For the complete expansion then has all of its terms in the other expression, and this other expression is therefore identically equal to 1. Note that 0 must not be considered a term of the expansion. On the other hand, it is convenient nominally to consider 1 a complete term on no letters, and hence actually an incomplete term on every set of letters. We can then state

that a function whose complete expansion has 0 for one expression admits an expansion whose other expression is 1.

§4

DIFFERENT EXPANSIONS OF THE SAME FUNCTION;

RELEVANT AND IRRELEVANT VARIABLES

We saw in the preceding section that every complete term on A, B, \dots, I is contained in at least one term of an arbitrary expansion on A, B, \dots, I , and that the terms of this expansion containing a given complete term on A, B, \dots, I are all in the same expression of the expansion. We now readily see that the terms of different expansions of the same membership function $f(A, B, \dots, I)$ that contain a given complete term on A, B, \dots, I are in the first expression for all expansions of the function, or in the second expression for all expansions of the function. For otherwise the complete term would be contained both in $f(A, B, \dots, I)$ and in the negative of $f(A, B, \dots, I)$.

We may then say that the terms of different expansions of the same membership function $f(A, B, \dots, I)$ that contain a given complete term on A, B, \dots, I are in corresponding expressions of their respective expansions. It follows, in particular, that every term of either expression of the complete expansion of $f(A, B, \dots, I)$ is contained in at least one term of the corresponding expression of any other expansion of $f(A, B, \dots, I)$.

If a membership function $f(A, B, \dots, I)$ admits an expansion which does not involve an argument G of the function, G will be said to be an irrelevant variable of the function; otherwise relevant.³³

³³ Note that it is the membership table of a membership function that determines for us what the arguments of the function are. If, then, a membership function is given by an expansion thereof, we assume that it is explicitly stated what other variables, if any, than those which appear in the expansion are arguments of the function.

Every expansion of $f(A, B, \dots, I)$ therefore involves all the relevant variables of $f(A, B, \dots, I)$. We now prove that $f(A, B, \dots, I)$ admits an expansion involving only its relevant variables. In particular, if all the arguments of $f(A, B, \dots, I)$ are irrelevant, $f(A, B, \dots, I)$ must admit one of the two expansions
1:0, 0:1.

If G is an irrelevant variable of $f(A, B, \dots, I)$, then any two complete terms on A, B, \dots, I which differ only in the capitalization of G must be in the same expression of the complete expansion of $f(A, B, \dots, I)$. For in any expansion of $f(A, B, \dots, I)$ which does not involve G , the same term which contains one of such a pair of complete terms must contain the other. By successive application of this result it follows that any two complete terms on A, B, \dots, I which differ only in the capitalization of irrelevant variables of $f(A, B, \dots, I)$ will be in the same expression of the complete expansion of $f(A, B, \dots, I)$. Let B, D, \dots, H be the relevant variables of $f(A, B, \dots, I)$. The complete terms on A, B, \dots, I can be grouped into mutually exclusive sets of terms, each set consisting of the complete terms on A, B, \dots, I contained in a given complete term on B, D, \dots, H . Since the terms in any one of these sets differ only in the capitalization of irrelevant variables of $f(A, B, \dots, I)$, they will all be in the same expression of the complete expansion of $f(A, B, \dots, I)$. Furthermore, their logical sum is identically equal to the complete term on B, D, \dots, H in which they are all contained. By thus uniting the terms in each of the above sets, the complete expansion of $f(A, B, \dots, I)$ is transformed into an expansion of $f(A, B, \dots, I)$ involving only the variables B, D, \dots, H .

A special consequence of the preceding argument is the following criterion for determining the relevant and irrelevant variables of $f(A, B, \dots, I)$. The necessary and sufficient condition that an argument G of a membership function $f(A, B, \dots, I)$ be irrelevant is that the complete expansion of $f(A, B, \dots, I)$ remain unchanged when G and g are interchanged, and hence:

That the function $f(A, B, \dots, I)$ remain unchanged when

the variable G is replaced by its negative.³⁴

§5

ITERATIVE ASPECTS OF THE JEVONS NOTATION

Let each function in a closed membership system over V' be given by some expansion of that function. In terms of these expansions the process of replacing the arguments of a function in the system by variables in V' or functions in the system assumes the following form.

In an expansion of the given function replace the capital and small letter of each argument of the function by the capital and small letter respectively of a replacing variable, or by the first and second expressions respectively of an expansion of a replacing function. The result, after due simplification by the rules of symbolic logic, will be an expansion of the function yielded by the given iterative process.

If an argument G of a function is thus to be replaced by a variable E or by a function with expansion $\epsilon_1:\epsilon_2$ we shall say that in the expansion of the function, $G:g$ is to be replaced by $E:e$ or $\epsilon_1:\epsilon_2$ respectively, i.e., G is to be replaced by E or ϵ_1 , g by e or ϵ_2 .

Our notation $f(M,N,\dots,W)$ for an iterative process depending on variables A,B,\dots,I does not reveal the identity of the original arguments of f . It is convenient in their place to use the auxiliary symbols M,N,\dots,W which represent the variables and functions that replace those arguments. Then, from an expansion of $f(M,N,\dots,W)$ in terms of M,N,\dots,W we obtain by the above method an expansion of $f(M,N,\dots,W)$ in terms of A,B,\dots,I , that is, an expansion of the function of A,B,\dots,I

³⁴ We have endeavored to develop a technique expressed entirely in terms of expansions of membership functions. Otherwise we would have defined an argument G of $f(A,B,\dots,I)$ to be an irrelevant variable thereof if the value of $f(A,B,\dots,I)$ is independent of the value of G . This is easily restated in terms of the membership table of $f(A,B,\dots,I)$.

yielded by the iterative process $f(M, N, \dots, W)$.

In particular we shall be interested in the relationship between the complete expansion of $f(M, N, \dots, W)$ in terms of M, N, \dots, W and the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, I . To study this relationship we must first extend a result of §3. We saw there that if M and m represent a membership function of A, B, \dots, I and its negative, then every complete term $t(A, B, \dots, I)$ is either contained in M and excluded from m , or contained in m and excluded from M . We now observe that this result continues to hold if M and m represent one of the variables A, B, \dots, I and its negative, or a membership function of only some of the variables A, B, \dots, I , and its negative. The first case is immediate. In the second, note that an expansion of the function can be considered to be an expansion on all the letters A, B, \dots, I . Since $t(A, B, \dots, I)$ is therefore contained in some term of one expression of the expansion and excluded from every term of the other expression it will be contained in one expression of the expansion and excluded from the other. Whence the above result.

For complete generality let $f(M, N, \dots, W)$ depend on no other variables than A, B, \dots, I . A complete term $t(A, B, \dots, I)$ will then be contained in one member of each pair $(M, m), (N, n), \dots, (W, w)$ and be excluded from the other. It therefore determines a selection from these pairs such that it is contained in each member of that selection but is excluded from at least one member of any other selection from those pairs. $t(A, B, \dots, I)$ is therefore contained in the complete term on M, N, \dots, W determined by this selection but is excluded from every other complete term on M, N, \dots, W . Clearly, the complete term on M, N, \dots, W containing $t(A, B, \dots, I)$ is in the first, or second, expression of the complete expansion of $f(M, N, \dots, W)$ with respect to M, N, \dots, W according as $t(A, B, \dots, I)$ is contained in $f(M, N, \dots, W)$, or in the negative of $f(M, N, \dots, W)$. It follows, in particular, that:

If $f(M, N, \dots, W)$ depends on all the variables A, B, \dots, I , each term of either expression of the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, I is contained in one and only one term of the cor-

responding expression of the complete expansion of $f(M, N, \dots, W)$ in terms of M, N, \dots, W .

Finally let $f(M, N, \dots, W)$ and $f(M', N', \dots, W')$ be two iterative processes with the same f which depend on no other variables than A, B, \dots, I . It will be convenient to refer to $G:g$ as an expansion of the variable G . By corresponding complete terms on M, N, \dots, W and M', N', \dots, W' will be meant two complete terms $T(M, N, \dots, W)$ and $T(M', N', \dots, W')$ with the same T . Recalling that a membership function and its negative are identically equal to the first and second expressions respectively of any expansion of the function, we are led by the above argument to the following result.

If a complete term $t(A, B, \dots, I)$ is contained in corresponding expressions of expansions of M, N, \dots, W and M', N', \dots, W' respectively, it will be contained in corresponding complete terms on M, N, \dots, W and M', N', \dots, W' , and hence in corresponding expressions of expansions of $f(M, N, \dots, W)$ and $f(M', N', \dots, W')$.

§6

DUALITY³⁵

Let us temporarily set aside the Jevons notation and symbolize the negative of A by $-A$. Since $-(-A) = A$ it follows that the function 'negation' is its own inverse. We now define the dual of a membership function $f(A, B, \dots, I)$ as the transform of the function under negation. The dual of $f(A, B, \dots, I)$ is

³⁵ The concept of duality goes back at least to Schröder. The writer came upon this concept early in the solution of the main problem of the present paper via the relativity of the signs, +, -, occurring in a truth-table when they are considered to be just marks. Zylinski [42] seems to have had the same experience. In the present introduction of this concept we follow the formulation of Wiener [39] and Frink [8]. The results of the present section in part overlap those of Bernstein in [3]. See also Frink [8], p. 485.

$\dots I$) is then $-f(-A, -B, \dots, -I)$. Clearly, the dual of the dual of $f(A, B, \dots, I)$ is $f(A, B, \dots, I)$. We can therefore refer to a membership function and its dual as a pair of dual functions.

In connection with iteration it is convenient to call a variable its own dual. If in an iterative process $f(M, N, \dots, W)$ we replace f, M, N, \dots, W by their duals, the resulting iterative process will be termed the dual of the given process. It follows from the general property of transforms that dual iterative processes yield dual functions.

Hence, if we replace each function in a closed membership system by its dual, the resulting system of functions will be closed. This closed system will be called the dual of the given system. Clearly, the dual of the dual of a closed system is the system itself. We can therefore talk of a pair of dual systems.

If each function in a set of generators is replaced by its dual, the resulting set of generators will be called the dual of the given set. Since dual iterative processes yield dual functions, it follows by induction that dual sets of generators generate dual closed systems. This principle introduces a great economy in the study of closed membership systems.

A function, or closed system, which is its own dual will be termed self-dual. Clearly, a closed system consisting wholly of self-dual functions is self-dual. On the other hand, a self-dual system need not consist wholly of self-dual functions. All we can say is that the dual of each function in a self-dual system is also in the system.

The condition of self-duality as applied to membership functions is our first example of an iterative condition.

In fact, if each function used in building up an iterative process is self-dual, then the iterative process will be its own dual. Hence the function yielded by the process will be its own dual, that is, self-dual. It follows in particular that the closed system generated by a set of self-dual generators is a

self-dual system consisting wholly of self-dual functions.

When our definition of the dual of a function is translated into the Jevons notation, we obtain the following rule for writing down an expansion of the dual of a function given an expansion of the function. Replace each capital letter occurring in the expansion by the corresponding small letter and conversely, and interchange the resulting expressions. The particular expansion yielded by this rule may be called the dual of the given expansion.

Thus, the dual of the expansion $A+B:ab$ is $AB:a+b$; and hence the dual of the membership function 'logical sum of A and B' is the membership function 'logical product of A and B' .

If each capital letter occurring in a term is replaced by the corresponding small letter, and conversely, the resulting term will be called the dual of the given term. Thus the dual of $AbcDe$ is $aBcdE$. We can therefore restate the above rule for obtaining the dual of a given expansion as follows. Place the dual of each term of the given expansion in the opposite expression of the desired dual expansion. Now the complete terms on a given set of letters can be grouped into mutually exclusive pairs of dual terms. Hence, the above process, when applied to the complete expansion of a function, results in the complete expansion of the dual of the function. A function will therefore be self-dual when and only when its complete expansion is unchanged by this process.

It follows that in the complete expansion of a self-dual function dual terms are always in opposite expressions, while in the complete expansion of a function that is not self-dual at least one pair of dual terms are in the same expression.

Two complete terms on the same set of letters are evidently dual when and only when no letter is capital in both terms or small in both terms. Hence a function is non-self-dual when and

only when there are two terms in the same expression of its complete expansion which have neither a capital nor small letter in common.³⁶ We now show that this criterion of non-self-duality applies to any expansion of a function, provided 1 is not a term of this expansion. In fact, if two terms in the same expression of the complete expansion of the function have neither a capital nor a small letter in common, two terms of the given expansion which respectively contain them, and which therefore are in the same expression of the given expansion, will have neither a capital nor a small letter in common. Conversely, if two terms in the same expression of the given expansion have neither a capital nor a small letter in common, by supplying missing letters in the two terms with opposite capitalizations two complete terms in the same expression of the complete expansion are obtained which have neither a capital nor a small letter in common.

We can therefore state that the necessary and sufficient condition that an expansion which does not have 1 for a term represent a self-dual function is that each pair of terms belonging to the same expression of the expansion have either a capital or a small letter in common.

§7

UNIFORM TERMS

A term will be said to be a uniform term of the first kind if all of its letters are capital, of the second kind if all of its letters are small.

1 is then the only term which is a uniform term of both kinds. A complete expansion on the letters A,B,...I has exactly one uniform term of the first kind and one of the second, namely, A,B,...I and a,b,...i. Since each term of the complete expansion of a function is contained in at least one term of any

³⁶ Note that 1:0 and 0:1 are never complete expansions.

other expansion of the function, while a term containing a uniform term of a given kind must be a uniform term of the same kind,

It follows that every expansion of a function has at least one uniform term of the first kind and at least one uniform term of the second kind; and if the complete expansion of the function has its uniform term of the i -th kind in the j -th expression, then every expansion of the function has all of its uniform terms of the i -th kind in the j -th expression.

The condition that the uniform terms of the first kind are in the first expression of an expansion of a function is iterative.

For let $f(M, N, \dots, W)$ be an iterative process depending on the variables A, B, \dots, I , and built up out of functions satisfying the condition. $AB\dots I$ will be contained in every uniform term of the first kind which involves no other letters than A, B, \dots, I . It will therefore be contained in M, N, \dots, W , and hence in $MN\dots W$. But $MN\dots W$ is in the first expression of the complete expansion of $f(M, N, \dots, W)$ in terms of M, N, \dots, W . Hence $AB\dots I$ will be in the first expression of the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, I .

By the dual argument it follows that the condition that the uniform terms of the second kind are in the second expression of an expansion of a function is also iterative. Therefore, the combined condition, the uniform terms of the first kind are in the first expression, the uniform terms of the second kind are in the second expression of an expansion of a function, is iterative.

§8

THE [A:a], [AA:] AND [:aa] CONDITION

We consider now the following three analogous conditions that an expansion of a function may satisfy.

The [A:a] condition: for any two terms in different expressions of the expansion there is a letter which is capital in the term that is in the first expression and small in the term that is in the second expression.

The [AA:] condition: any two terms of the first expression of the expansion have a capital letter in common.

The [:aa] condition: any two terms of the second expression of the expansion have a small letter in common.

In the last two conditions we must let the phrase 'two terms' cover the case of a single term chosen twice. This qualification is required only when the expression referred to in the condition consists of but a single term, in which case the condition is then equivalent to the term in question possessing a letter with the indicated capitalization. Note that when an expression referred to in any of the three conditions is 0, that expression has no terms, and hence the condition in question must be considered to be satisfied. Thus 1:0 satisfies both the [A:a] and [:aa] condition, 0:1 the [A:a] and [AA:] condition.

We first prove that if one expansion of a function satisfies any of these three conditions, then every expansion of the function satisfies that condition.

Since the proof is much the same for all three cases, we give it explicitly only for the [A:a] condition. If the [A:a] condition is not satisfied by the complete expansion of a function, there will be a term in the first expression of this com-

plete expansion, and a term in the second expression, such that no letter is at the same time capital in the first term and small in the second. Any other expansion of the same function will have a term in its first expression and a term in its second expression which respectively contain these terms of the complete expansion, and hence can be obtained from them by omitting some of their letters. There is therefore no letter which is capital in the first of these containing terms and small in the second. Hence the new expansion also fails to satisfy the [A:a] condition. Conversely, if any expansion of a function fails to satisfy the [A:a] condition, there will be a term in the first expression, and a term in the second expression, of the expansion which lead to this failure. And by supplying missing letters in the first term as small letters, and in the second term as capital letters, two complete terms are obtained which result in the complete expansion of the function failing to satisfy the [A:a] condition. It follows that the [A:a] condition is satisfied by all expansions of a function or by none.

Since these conditions on an expansion of a function are independent of the particular expansion of the function employed, they may be considered to be conditions on the function itself.³⁷

It is then significant that each of these conditions, considered as a condition imposed on a function, is iterative.

Again the proofs are similar so that we restrict our attention to a single condition, this time the [AA:] condition. Let $f(M, N, \dots, W)$ be an iterative process depending on the variables A, B, \dots, I and built up out of functions satisfying the

³⁷ Thus the [:aa] condition may be restated: whenever both x_1 and x_2 of the universal class are outside of the class $f(A, B, \dots, I)$, then x_1 and x_2 are both outside of at least one of the classes A, B, \dots, I . In terms of the membership table of $f(A, B, \dots, I)$ this becomes, whenever the membership value of x with respect to $f(A, B, \dots, I)$ is - for each of two assignments of membership values of x with respect to A, B, \dots, I , then the membership value of x with respect to at least one of the variables A, B, \dots, I is - in each assignment. In our original presentation in terms of truth-tables this condition was briefly written, "any two - configurations have a - in common".

[AA:] condition. Let t_1 and t_2 be any two terms of the first expression of the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, I . t_1 and t_2 will then be respectively contained in two terms T_1 and T_2 of the first expression of the complete expansion of $f(M, N, \dots, W)$ in terms of M, N, \dots, W . Since f satisfies the [AA:] condition, T_1 and T_2 must have a capital letter, say P , in common. Both t_1 and t_2 will therefore be contained in P . If P is one of the letters A, B, \dots, I , t_1 and t_2 have that capital letter in common. Otherwise, t_1 and t_2 will be respectively contained in two terms t'_1 and t'_2 of the first expression of some expansion of the function P . As this function satisfies the [AA:] condition, t'_1 and t'_2 will have a capital letter in common, and hence t_1 and t_2 have that capital letter in common. The function of A, B, \dots, I yielded by the iterative process $f(M, N, \dots, W)$ therefore satisfies the [AA:] condition.

The [A:a] condition assumes an added importance from the fact that a function satisfying this condition admits a characteristic form of expansion which is in general different from, and usually more useful than the complete expansion of the function.

We proceed to show that every function satisfying the [A:a] condition admits an expansion whose first expression involves no small letters and whose second expression involves no capital letters.

Consider any term of the first expression of the complete expansion of an [A:a] function $f(A, B, \dots, I)$ which involves at least one small letter. Every term obtainable from this term by changing one or more of its small letters into the corresponding capital letters will then also be in that first expression. For such a term can have no letter small that is capital in the given term. Now these terms, together with the given term, constitute all of the complete terms on A, B, \dots, I that are contained in the incomplete term on A, B, \dots, I formed from the given term by omitting all of its small letters. This incomplete term on A, B, \dots, I , which contains the given term, and involves no small letters, is therefore itself contained in the first ex-

pression of the complete expansion of $f(A,B,\dots,I)$. We can therefore replace every term of that first expression which involves at least one small letter by its associated incomplete term and thereby transform that first expression into one involving no small letters. By the dual method, the second expression of the complete expansion of $f(A,B,\dots,I)$ can be transformed into one which involves no capital letters.

Conversely, if a function admits an expansion whose first expression involves no small letters, and whose second expression involves no capital letters, the function must satisfy the $[A:a]$ condition. In fact, this must be true if either expression is of the indicated form. For in any expansion, corresponding to a term in the first expression and a term in the second expression there must be a letter that is capital in one term and small in the other; as otherwise the logical product of the two expressions would not be identically 0. And if either expression is of the indicated form, it must be the term in the first expression that involves the capital letter, the term in the second expression that involves the small letter. It follows that if either expression of an expansion of a function is of the indicated form, the function admits an expansion in which both expressions are of that form.

If in such an expansion of an $[A:a]$ function identical terms are united, and each term which is contained in another term of this simplified expansion is then removed, the result will be an expansion of the function, of the same form as the given expansion, in which no term is contained in another term. This expansion will be called the normal expansion of the $[A:a]$ function. It can be proved that an $[A:a]$ function admits but one normal expansion.³⁸ Furthermore, every variable appearing in the normal expansion of an $[A:a]$ function is relevant. Clearly, when one expression of the complete expansion of an $[A:a]$ function is 0, the normal expansion of the function is either 1:0 or 0:1. In every other case each expression of the normal expansion of an $[A:a]$ function has at least one term, and the first expression is actually written in terms of capital letters only, the second in terms of small letters only.

³⁸ A paper of Blake's [6] may here be relevant.

As examples we have $A+B:ab$ and $AB:a+b$ as the normal expansions of the $[A:a]$ functions 'logical sum of A and B ' and 'logical product of A and B ' respectively.

The three conditions discussed in the present section are intimately associated with the condition of self-duality. Note in general that if a function satisfies the $[A:a]$ condition, so will the dual of the function. And if a function satisfies one of the two conditions $[AA:]$, $[:aa]$ the dual of the function will satisfy the other. The $[A:a]$ condition may therefore be thought of as a self-dual condition, the $[AA:]$ and $[:aa]$ conditions as duals of each other. It follows immediately that if a self-dual function satisfies one of the two conditions $[AA:]$, $[:aa]$, it also satisfies the other. Furthermore, a self-dual function which satisfies one of the two conditions $[A:a]$, $[AA:]$ also satisfies the other. For let t_1 be any term in the first expression of the complete expansion of a self-dual function, t_2 any term in the second expression. t'_2 , the dual of t_2 , will therefore be in the first expression. If t_2 has no small letter which is capital in t_1 , t'_2 will have no capital letter in common with t_1 , and conversely. That is, if a self-dual function fails to satisfy the $[A:a]$ condition it also fails to satisfy the $[AA:]$ condition, and conversely. We therefore have that a self-dual function which satisfies any one of the three conditions $[A:a]$, $[AA:]$, $[:aa]$ satisfies the other two as well. Finally, if a function satisfies both the $[AA:]$ and $[:aa]$ condition, an expansion of the function cannot have 1 for a term, while any two terms in the same expression of the expansion will have either a capital or a small letter in common. The function will therefore be self-dual.

Hence, except for the pairs of conditions $[A:a]$, $[AA:]$, and $[A:a]$, $[:aa]$, if a function satisfies two of the four conditions, self-duality, $[A:a]$, $[AA:]$, $[:aa]$, it will also satisfy the other two.

In the development of Part II the failure of a function to satisfy one of the three conditions studied in the present section plays a significant rôle. Such failure is always due to the existence of a pair of terms in suitable expressions of an

expansion of the function which do not conform to the requirements of the condition. Now for any two terms of a complete expansion, the letters on which the expansion is written fall into at most four groups with respect to their capitalization in the terms in question, to wit,

first term: capital capital small small,
second term: small capital small capital.

In testing the complete expansion for the [A:a] condition, the first term will be understood to be in the first expression, the second in the second expression; while for the [AA:] condition both terms are to be in the first expression, for the [:aa] condition both terms are to be in the second expression. Then the failure of the complete expansion to satisfy the [A:a], [AA:], or [:aa] condition is equivalent to the existence of a pair of terms in which no letter belongs to the first group, second group and third group respectively.

§9

A AND a-COMPONENTS

The 2^n complete terms on n letters A,B,...I can be obtained by logically multiplying each of the 2^{n-1} complete terms on the $(n-1)$ letters B,...I by A, and again by a. If then in a complete expansion on A,B,...I only terms involving the capital letter A are retained, and that capital letter is then struck out, the result will be a complete expansion on B,...I. The function of B,...I defined by this complete expansion will be called the A-component of the function $f(A,B,\dots,I)$ defined by the original complete expansion. In like manner we define the a-component of $f(A,B,\dots,I)$. For an arbitrary expansion of $f(A,B,\dots,I)$ the corresponding rules are easily seen to be the following.

To obtain an expansion of the A-component of $f(A,B,\dots,I)$ strike out capital A and all terms involving small a. Likewise for the a-component.

Thus $AB+BC+AC:ab+bc+ac$ is an expansion with A-component $B+BC+C:bc$, a-component $BC:b+bc+c$, i.e., $B+C:bc$, $BC:b+c$ respectively.

If we symbolize the A and a-components of $f(A,B,\dots,I)$ by $f_A(B,\dots,I)$ and $f_a(B,\dots,I)$ respectively, we see from their definition that

$$f(A,B,\dots,I) = Af_A(B,\dots,I) + af_a(B,\dots,I).$$

Now in this identity substitute 1 for A, and hence 0 for a, and again 0 for A, and hence 1 for a. We then obtain the formulas

$$f_A(B,\dots,I) = f(1,B,\dots,I),$$

$$f_a(B,\dots,I) = f(0,B,\dots,I).$$

Our formula for $f(A,B,\dots,I)$ shows that $f(A,B,\dots,I)$ is determined by its A and a-components. Now a condition κ imposed upon functions may have the property that no two distinct κ -functions can have the same A-component. We may then say that a κ -function is determined by its A-component. Under certain circumstances, as when κ is iterative, this property of κ enables us to determine the system of functions generated by a set of κ -functions when the system of functions generated by the A-components of the given functions is known. Though the details of this method are best considered in connection with individual cases as they come up in Part II, we prove here a theorem which, with certain analogous results, constitutes the backbone of the method.

THEOREM. Let $f_1(A,B,C)$ and $f_2(A,B,C)$ have A-components $\phi_1(B,C)$ and $\phi_2(B,C)$ respectively. Then each function $\phi(B,C,\dots,I)$ generated by ϕ_1 and ϕ_2 is the A-component of a function $f(A,B,C,\dots,I)$ generated by f_1 and f_2 .

For consider any operation built up by ϕ_1 and ϕ_2 which yields the function $\phi(B,C,\dots,I)$. Since $\phi_1(B,C) = f_1(1,B,C)$,

and $\phi_2(B,C) = f_2(1,B,C,)$ we can rewrite this operation so that it is expressed in terms of $f_1, f_2, B, C, \dots I$ and the constant 1. In this transformed operation replace each 1 by A. The new operation then yields a function $f(A,B,C,\dots I)$ generated by f_1 and f_2 . From the manner in which this new operation was obtained we see then that $\phi(B,C,\dots I) = f(1,B,C,\dots I)$. That is $\phi(B,C,\dots I)$ is the A-component of $f(A,B,C,\dots I)$.

Clearly the above theorem is true if in place of $f_1(A,B,C)$ with A-component $\phi_1(B,C)$ we have a function $f_1(A,B)$ with A-component $\phi_1(B)$. Furthermore completely analogous results hold for a-components. We shall refer to the theorem of this section when any of these results are required. In certain cases we shall also require the following observation which is a direct consequence of the way in which the above operation for $f(A,B,C,\dots I)$ was constructed.

To wit, $f(A,B,C,\dots I)$ can be generated by f_1 and f_2 through iterative processes in which no replacement is effected upon the variable A.

PART II

DERIVATION OF CLOSED SYSTEMS

§10

FIRST ORDER FUNCTIONS AND SYSTEMS AND THEIR GENERAL SIGNIFICANCE

There are four complete expansions on a single variable A , to wit, $A:a$, $A+a:0$, $0:A+a$, $a:A$. The corresponding functions of A will be symbolized $\alpha_1(A)$, $\beta_1(A)$, $\gamma_1(A)$, $\delta_1(A)$ respectively. $\alpha_1(A)$ is therefore identically equal to A , $\beta_1(A)$ to 1, $\gamma_1(A)$ to 0, $\delta_1(A)$ to the negative of A . The functions $\alpha_1(A)$ and $\delta_1(A)$ are evidently self-dual, while $\beta_1(A)$ and $\gamma_1(A)$ are duals of each other. $\beta_1(A)$ and $\gamma_1(A)$ admit the expansion $1:0$, $0:1$ respectively in addition to their complete expansions.

If all of the arguments of a membership function $f(B,D,\dots,H)$ be replaced by A , the resulting first order function $f(A,A,\dots,A)$ must be one of the above four first order functions. It will be called the associated first order function; and according as it is $\alpha_1(A)$, $\beta_1(A)$, $\gamma_1(A)$, or $\delta_1(A)$, $f(B,D,\dots,H)$ will be said to be an α , β , γ , or δ -function. By referring to §7 we readily verify the following equivalences.

α -function: the uniform terms of the first kind are in the first expression, the uniform terms of the second kind are in the second expression of an expansion of the function;

β -function: the uniform terms of both kinds are in the first expression of an expansion of the function;

γ -function: the uniform terms of both kinds are in the second expression of an expansion of the function;

δ -function: the uniform terms of the first kind are in the second expression, the uniform terms of the second kind are in the first expression of an expansion of the function.

Clearly, the dual of a β -function is a γ -function and conversely, while the dual of an α -function

is an α -function, or a δ -function a δ -function.

It follows in particular that a self-dual function can only be an α -function or a δ -function.

Unless otherwise stated, by function we shall henceforth mean a membership function of a finite number of variables in V' , and by closed system a system of membership functions iteratively closed with respect to V' . The above definitions serve to divide the complete system of functions into four mutually exclusive classes of functions. In terms of this classification the iterative properties of uniform terms can be expressed as follows. The system of all α and β -functions is closed, the system of all α and γ -functions is closed, the system of all α -functions is closed. It turns out in the sequel that, except for the complete system, these three closed systems are the only ones which contain one or more of the above classes of functions in their entirety.

Since there are exactly four first order functions of any one variable, the system of all first order functions is constituted as follows.

$$[\alpha_1(A), \beta_1(A), \gamma_1(A), \delta_1(A), \alpha_1(B), \beta_1(B), \gamma_1(B), \delta_1(B), \alpha_1(C), \beta_1(C), \gamma_1(C), \delta_1(C), \dots].$$

This complete system of first order functions is evidently closed. For every iterative process involving only first order functions results in a first order function. Its associated contracted system is the finite set of functions $[\alpha_1(A), \beta_1(A), \gamma_1(A), \delta_1(A)]$. Now it is readily seen that the contracted systems associated with closed systems of first order functions are those subsets of this set which are themselves iteratively closed with respect to A. Furthermore, it can be directly verified that an iterative process $f(g(A))$ yields either the function $f(A)$, or $g(A)$, with the following exceptions.

$$\delta_1(\beta_1(A)) = \gamma_1(A), \quad \delta_1(\gamma_1(A)) = \beta_1(A), \quad \delta_1(\delta_1(A)) = \alpha_1(A).$$

It easily follows that of the fifteen existing subsets of the set

of functions $[\alpha_1(A), \beta_1(A), \gamma_1(A), \delta_1(A)]$ exactly nine are contracted closed systems. The corresponding closed systems of first order functions can be symbolized and described as follows:

$$o_1: [\alpha_1], \quad o_2: [\beta_1], \quad o_3: [\gamma_1], \quad o_4: [\alpha_1, \delta_1], \quad o_5: [\alpha_1, \beta_1], \quad o_6: [\alpha_1, \gamma_1], \\ o_7: [\beta_1, \gamma_1], \quad o_8: [\alpha_1, \beta_1, \gamma_1], \quad o_9: [\alpha_1, \beta_1, \gamma_1, \delta_1],$$

where $o_4: [\alpha_1, \delta_1]$, for example, means that o_4 consists of all functions $\alpha_1(X), \delta_1(X)$ with $X = A, B, C, \dots$. As we do not consider the null system of functions a closed system, o_1 to o_9 are the only closed systems of first order functions. Clearly, o_1, o_4, o_7, o_8 , and o_9 are self-dual systems, while o_2 and o_3 and o_5 and o_6 , are duals of each other. Since the contracted closed system associated with any of these systems itself constitutes a finite set of generators of these systems, these systems are all of order one.

Given any closed system, every iterative process which involves only first order functions that belong to the system must yield a first order function that also belongs to the system. It follows that the first order functions possessed by a closed system themselves constitute a closed system. This closed system, which must be one of the above nine first order systems, will be called the associated first order system. Evidently, the first order function associated with any function in a closed system must itself be in the system, and hence in the associated first order system.

The associated first order system therefore not only specifies the first order functions that belong to the given closed system but also indicates which of the four classes of functions $\alpha, \beta, \gamma, \delta$ are represented in the closed system.

Referring to the constitution of the nine first order systems, we see that every closed system can be described as an $[\alpha, \beta, \gamma, \delta]$, $[\alpha, \beta, [\alpha, \gamma]]$, $[\alpha, \beta, [\alpha, \gamma], [\beta, \gamma]]$, or $[\alpha, \beta, \gamma, \delta]$ system. Thus, to say that a closed system is an $[\alpha, \beta]$ system means that its associated first order system is $o_5: [\alpha_1, \beta_1]$,

and hence that it consists wholly of α and β -functions and actually possesses at least the first order α and β -functions.

It is interesting to note that for each closed first order system there is a maximal closed system with which it is associated, i.e., one containing all closed systems with the given associated first order system.³⁹

Thus, the set of all functions that belong to $[\alpha, \delta]$ systems may be considered to be an infinite set of generators which generate some closed system F . If $f(B, D, \dots, H)$ is any one of these generators, and M, N, \dots, W represent A , $\alpha_1(A)$, or $\delta_1(A)$ then $f(M, N, \dots, W)$ must represent either $\alpha_1(A)$ or $\delta_1(A)$. For $f(B, D, \dots, H)$ is in some closed $[\alpha, \delta]$ system. As $\alpha_1(A)$ and $\delta_1(A)$ must also be in this system, the same must be true of the function of A , $f(M, N, \dots, W)$. But the only functions of A in a closed $[\alpha, \delta]$ system are $\alpha_1(A)$ and $\delta_1(A)$. It follows by induction that the infinite set of generators can generate no other functions of A than $\alpha_1(A)$ and $\delta_1(A)$. F is therefore an $[\alpha, \delta]$ system; and by its definition it contains all $[\alpha, \delta]$ systems. The above found closed systems consisting of all α , α and β , α and γ -functions may be called the complete $[\alpha]$, $[\alpha, \beta]$, $[\alpha, \gamma]$ systems respectively. The complete $[\alpha]$ system must contain all closed systems consisting wholly of α -functions and hence is the maximal $[\alpha]$ system. Likewise the complete $[\alpha, \beta]$ and $[\alpha, \gamma]$ systems are the maximal $[\alpha, \beta]$ and $[\alpha, \gamma]$ systems respectively. The maximal $[\alpha, \beta, \gamma, \delta]$ system is of course the complete system itself. The composition of the remaining five

³⁹ This theorem and subsequent proof applies to any iteratively closed system within the generality of §1 provided the condition that the number of generators be finite is dropped from the definition of a closed system of given finite order. More generally and with the same widening of definition, the n -th and lower order functions of any iteratively closed system (in the sense of §1) will be the same as the n -th and lower order functions of a unique closed system of finite order less than or equal to n ; and of all closed systems thus associated for given n with a closed system of finite order, there will be a maximal closed system, i.e., one containing all. Though this maximal closed system theorem is of later origin, the relationship from which it stems was basic in our original solution of the main problem of the present paper.

maximal systems must be left to later developments.

This classification of all closed systems according to their associated first order systems is the backbone of the ensuing discussion. This discussion is considerably shortened by the use of the concept of duality. Note that the dual of an $[\alpha]$, $[\alpha, \delta]$, $[\beta, \gamma]$, $[\alpha, \beta, \gamma]$, and $[\alpha, \beta, \gamma, \delta]$ system is a system of the same type, while the dual of a $[\beta]$ system is a $[\gamma]$ system, and conversely, the dual of an $[\alpha, \beta]$ system an $[\alpha, \gamma]$ system, and conversely. The derivation of all closed $[\alpha]$, $[\alpha, \beta]$, and $[\alpha, \gamma]$ systems is particularly involved. There is therefore considerable value in the fact that after certain non-self-dual $[\alpha]$ and all $[\alpha, \beta]$ systems have been found, the duals of these systems immediately give us the remaining non-self-dual $[\alpha]$, and all $[\alpha, \gamma]$ systems.

§11

SYSTEMS OF FUNCTIONS REDUCIBLE TO FIRST ORDER

If a function admits the same expansion as a first order function, it will be said to be reducible to that first order function, or, more briefly, reducible to first order. If E is the argument of the first order function, then according as the given function is an α , β , γ , or δ -function, the first order function, and hence the given function, admits the expansion $E:e$, $1:0$, $0:1$, $e:E$ respectively. In the first and last case E is the only relevant variable of the given function, while in the second and third case all of the arguments of the given function are irrelevant. In the latter two cases the argument of the first order function can be chosen arbitrarily.

The condition that a function be reducible to first order is iterative. For if $f(M, N, \dots, W)$ is an iterative process depending on variables A, B, \dots, I , and built up out of functions reducible to first order, then $f(M, N, \dots, W)$ admits some expansion $R:r$, $1:0$, $0:1$, $r:R$ in terms of M, N, \dots, W , and hence some expansion $E:e$, $1:0$, $0:1$, $e:E$ in terms of A, B, \dots, I . The system of all functions reducible to first order is therefore closed.

This complete system of functions reducible to first order is composed as follows.

For each set of n variables it possesses n α -functions of those variables, one β -function, one γ -function, and n δ -functions. This follows from the fact that while all the variables are irrelevant in the case of a β and γ -function, any one of them can be chosen as the relevant variable of an α and δ -function. The associated contracted system therefore possesses exactly $2n+2$ functions of order n for every positive integer n .

Consider any closed system of functions reducible to first order which does not consist wholly of first order functions. Also consider an arbitrary finite set of variables. By §1 the system will have at least one function M on this set of variables. If the system has a β -function, it will also have the function $\beta_1(M)$. Hence, if the system has one β -function, it must have all β -functions reducible to first order. Likewise for γ -functions. If the system has an α -function of order greater than one, by replacing the irrelevant variables of this function by M , its relevant variable by any variable E in the given set of variables, an α -function on the given set of variables is obtained with relevant variable E . Hence, if the system has one α -function of order greater than one, it must have all α -functions reducible to first order. Since $\delta_1(M)$ is an α -function when M is a δ -function, and conversely, it also follows that if the system has a δ -function, and at least one α or δ -function of order greater than one, then it has all α and δ -functions reducible to first order. On the other hand, the system may have no α or δ -functions other than first order functions. For, if f is any β or γ -function reducible to first order, $f(M, N, \dots, W)$ is also a β or γ -function. While if f is a first order α or δ -function, $f(M)$ will be a β or γ -function if M is a β or γ -function, a first order α or δ -function if M is a variable, or a first order α or δ -function.

It follows that for each associated first order system there are at most two closed systems of functions reducible to first order, yet not of the first order, one system possessing all

functions reducible to first order of the types indicated by the associated first order system, the other possessing all such β and γ -functions, but only first order α and δ -functions. In describing these systems we shall use the symbols $\overset{*}{\alpha}_1$, $\overset{*}{\beta}_1$, $\overset{*}{\gamma}_1$, $\overset{*}{\delta}_1$, to mean any α , β , γ , δ -function respectively that is reducible to first order. If a closed system consists of functions reducible to first order, we shall say that the closed system is reducible to first order. By scanning the nine possible associated first order systems, we find that there are exactly thirteen closed systems reducible to first order, yet not of the first order. They may be symbolized and described as follows,

$$\begin{aligned} R_1 &: [\overset{*}{\alpha}_1], \quad R_2 : [\overset{*}{\beta}_1], \quad R_3 : [\overset{*}{\gamma}_1], \quad R_4 : [\overset{*}{\alpha}_1, \overset{*}{\delta}_1], \quad R_5 : [\overset{*}{\alpha}_1, \overset{*}{\beta}_1], \quad R_6 : [\overset{*}{\alpha}_1, \\ &\overset{*}{\beta}_1], \quad R_7 : [\overset{*}{\alpha}_1, \overset{*}{\gamma}_1], \quad R_8 : [\overset{*}{\alpha}_1, \overset{*}{\gamma}_1], \quad R_9 : [\overset{*}{\beta}_1, \overset{*}{\gamma}_1], \quad R_{10} : [\overset{*}{\alpha}_1, \overset{*}{\beta}_1, \overset{*}{\gamma}_1], \\ R_{11} &: [\overset{*}{\alpha}_1, \overset{*}{\beta}_1, \overset{*}{\gamma}_1], \quad R_{12} : [\overset{*}{\alpha}_1, \overset{*}{\beta}_1, \overset{*}{\gamma}_1, \overset{*}{\delta}_1], \quad R_{13} : [\overset{*}{\alpha}_1, \overset{*}{\beta}_1, \overset{*}{\gamma}_1, \overset{*}{\delta}_1]. \end{aligned}$$

Clearly, α and δ -functions reducible to first order are self-dual, while the dual of a β -function reducible to first order is a γ -function reducible to first order, and conversely. Hence R_2 and R_3 , R_5 and R_7 , R_6 and R_8 are dual closed systems, the others self-dual. Since in the above argument the function M can be generated by any second order function in the system, while the α or δ -function of order greater than one can be any second order α or δ -function in the system, it follows that each of the above systems can be generated by first and second order functions, and hence is itself of the second order. Together with the first order systems there are therefore 22 closed systems reducible to first order.

§12

 $[\beta]$, $[\gamma]$, AND $[\beta,\gamma]$ SYSTEMS

Suppose that a function $f(A, B, \dots, I)$ in a closed $[\beta]$ system has a term $t(A, B, \dots, I)$ in the second expression of its complete expansion. Since $f(A, B, \dots, I)$ can only be a β -function,

the uniform terms $AB\dots I$ and $ab\dots i$ must both be in the first expression of this complete expansion. $t(A,B,\dots I)$ therefore must have at least one capital letter, and at least one small letter. Now in the function $f(A,B,\dots I)$ replace each argument E by $\beta_1(A)$ or a according as E is capital or small in $t(A,B,\dots I)$. That is, in the complete expansion of $f(A,B,\dots I)$ replace $E:e$ by $1:0$ or $A:a$ in the manner indicated. $t(A,B,\dots I)$ then reduces to a , $AB\dots I$ to A , and hence the complete expansion of $f(A,B,\dots I)$ to $A:a$. That is, the above iterative process, which involves only functions in the given closed $[\beta]$ system, would yield the function $\alpha_1(A)$. $\alpha_1(A)$ would therefore belong to the given closed $[\beta]$ system, which is impossible.

Hence, all functions in a closed $[\beta]$ system have 0 for the second expressions of their complete expansions, and therefore admit the expansion $1:0$. Dually, all functions in a closed $[\gamma]$ system admit the expansion $0:1$. By the same argument the functions in a closed $[\beta,\gamma]$ system admit expansions $1:0$ and $0:1$.

That is, every closed $[\beta]$, $[\gamma]$, and $[\beta,\gamma]$ system consists wholly of functions reducible to first order. By referring to the last two sections, we thus see that the only closed $[\beta]$ systems are O_2 and R_2 , the only closed $[\gamma]$ systems are O_3 and R_3 , the only closed $[\beta,\gamma]$ systems are O_7 and R_9 . It follows incidentally that the maximal $[\beta]$, $[\gamma]$ and $[\beta,\gamma]$ systems are R_2 , R_3 , and R_9 respectively.

This result, though trivial in itself, is significant as marking the first step in our general plan of determining all closed systems through their associated first order systems.

§13

LOGICAL SUM SYSTEMS; LOGICAL PRODUCT SYSTEMS

The logical sum of two or more distinct variables $A,B,\dots I$ is a function having the simple expansion $A+B+\dots+I:ab\dots i$. If

a function admits the same expansion as such a 'logical sum function' it will be said to be reducible to that logical sum function. The arguments of the logical sum function are relevant variables of the given function, all other arguments of the given function, irrelevant. Obviously, all functions reducible to logical sum functions are α -functions.

Now it is easily verified that the system consisting of all functions reducible to first order α , β , and γ -functions, and to logical sum functions is closed.

We proceed to determine the closed subsystems of this complete 'logical sum system'.

If $f(A,B)$ is the function with expansion $A+B:ab$, then $f(A,A)$ has the expansion $A:a$, $f(A,f(B,\dots,f(H,I)\dots))$ the expansion $A+B+\dots+I:ab\dots i$. Furthermore, if M and N are variables or first order α -functions or logical sum functions, $f(M,N)$ is also a first order α -function or logical sum function. Hence, $f(A,B)$ generates S_1 , the closed system consisting of all first order α -functions and all logical sum functions.

Now any closed system of the type we are considering which does not consist wholly of functions reducible to first order must possess a function reducible to a logical sum function. If one of the relevant variables of this function be replaced by A , all of the remaining variables by B , the above function $f(A,B)$ with expansion $A+B:ab$ results. The closed system in question therefore contains S_1 .

If, furthermore, this closed system possesses an α -function not in S_1 , this α -function will have at least one relevant, and at least one irrelevant variable. If the relevant variables of this function are replaced by an arbitrary function M in S_1 , the irrelevant variables by an arbitrary function N in S_1 , the resulting function, which will also be in the closed system, is a function reducible to M . It follows that the closed system must then possess all functions reducible to functions in S_1 , i.e., all the α -functions possessed by the complete logical sum system.

There remains to see the effect of the presence of β and γ -functions. As in the case of the R systems, the closed system under discussion must possess all functions reducible to first order β and γ -functions therein. If the system does have a γ -function, by replacing B by $\gamma_1(B)$, i.e., $B:b$ by

0:1. $f(A,B)$ with expansion $A+B:ab$ becomes a function $g(A,B)$ with expansion $A:a$. The system therefore has the α -function $g(A,B)$ with relevant variable A , irrelevant variable B , and hence possesses all α -functions reducible to functions in S_1 . On the other hand, the presence of β -functions in the system is consistent with all of its α -functions being in S_1 . Thus, if any variable in a logical sum function be replaced by a β -function with expansion 1:0, the expansion of the logical sum function also reduces to 1:0, and but a β -function results.

The various possibilities are now easily seen. We shall continue the symbolism used for the R systems, and in addition use σ to symbolize an arbitrary logical sum function, σ^* an arbitrary function reducible to a logical sum function. Then the only closed subsystems of the complete logical sum system other than closed systems reducible to first order are the following.⁴⁰

$$S_1: [\alpha_1, \sigma], \quad S_2: [\alpha_1^*, \sigma^*], \quad S_3: [\alpha_1^*, \beta_1, \sigma], \quad S_4: [\alpha_1^*, \beta_1^*, \sigma^*], \quad S_5: [\alpha_1^*, \gamma_1^*, \sigma^*], \\ S_6: [\alpha_1^*, \beta_1^*, \gamma_1^*, \sigma^*].$$

The above derivation shows that each of these 'logical sum systems', can be generated by a suitable choice of generators from the functions $\beta(A)$, $\gamma_1(A)$, $f(A,B)$ with expansion $A+B:ab$, $g(A,B)$ with expansion $A:a$. Hence they are all of the second order.

The dual of the expansion $A+B+\dots+I:ab\dots i$ is the expansion $AB\dots I:a+b+\dots+i$. Hence the dual of a logical sum function is a 'logical product function', i.e., a function which is the logical product of two or more distinct variables, while the dual of a function reducible to a logical sum function is a function 'reducible to a logical product function'. By the principle of duality we may then immediately write down the closed systems which have no other functions than those reducible to first order α , β , and γ functions, and to logical product functions, and which are not themselves reducible to first order. Introducing the symbols π and π^* as duals of σ and σ^* , we thus have the following six 'logical product

⁴⁰ Allowing a closed system to be an improper subsystem of itself.

systems' which, in order, are the duals of the above six logical sum systems, and which are also of the second order.

$$P_1: [\alpha_1, \pi], \quad P_2: [\overset{*}{\alpha}_1, \overset{*}{\pi}], \quad P_3: [\alpha_1, \overset{*}{\gamma}_1, \pi], \quad P_4: [\overset{*}{\alpha}_1, \overset{*}{\gamma}_1, \overset{*}{\pi}], \quad P_5: \\ [\overset{*}{\alpha}_1, \beta_1, \pi], \quad P_6: [\overset{*}{\alpha}_1, \overset{*}{\gamma}_1, \overset{*}{\beta}_1, \overset{*}{\pi}].$$

§14

 $[\alpha, \beta, \gamma]$ SYSTEMS

By referring to §8 we see that the only functions satisfying the $[A:a]$ condition are β and γ -functions reducible to first order, and α -functions which admit a normal $[A:a]$ expansion. Since the $[A:a]$ condition is iterative, the system of all $[A:a]$ functions is closed. We thus have the closed $[\alpha, \beta, \gamma]$ system symbolized and described as follows.

A₁. All functions satisfying the $[A:a]$ condition.

Now suppose that a function belonging to a closed $[\alpha, \beta, \gamma]$ system does not satisfy the $[A:a]$ condition. Then, as seen in §8, there will be a term in the first expression and a term in the second expression of the complete expansion of this function with respect to which the arguments of the function fall into at most three groups as follows:

term in the first expression : capital small small,
term in the second expression : capital small capital.

Clearly, the third group cannot be empty, or the two terms would not be distinct. Now replace the variables in the first group by $\beta_1(A)$, in the second by $\gamma_1(A)$, in the third by A . The two terms become a and A respectively, hence the complete expansion $a:A$, and the function $\delta_1(A)$. The given closed system would therefore possess the function $\delta_1(A)$, which contradicts its being an $[\alpha, \beta, \gamma]$ system.

It follows that every closed $[\alpha, \beta, \gamma]$ system satisfies

the $[A:a]$ condition, i.e., consists wholly of functions satisfying the $[A:a]$ condition.

Hence, also, every closed $[\alpha,\beta,\gamma]$ system is contained in A_1 , which is therefore the maximal $[\alpha,\beta,\gamma]$ system.

Since the first expression of the normal expansion of an $[A:a]\alpha$ -function is the logical sum of logical products of capital letters, it follows that every $[A:a]\alpha$ -function can be generated by the two $[A:a]$ functions $A+B$ and AB . Hence A_1 can be generated by $\beta_1(A)$, $\gamma_1(A)$, $A+B$, and AB .

A_1 is therefore of the second order. It further follows that A_1 is the only closed $[\alpha,\beta,\gamma]$ system having both of the functions $A+B$ and AB .

Now let a closed $[\alpha,\beta,\gamma]$ system possess an α -function whose normal $[A:a]$ expansion has in its first expression a term $CE\dots F$ with more than one letter. Since $CE\dots F$ would not be in the normal expansion if it were contained in another term of the expansion, every other term of this first expression involves some capital letter other than C,E,\dots,F . By replacing every argument of the given function other than C,E,\dots,F by $\gamma_1(A)$, i.e., the corresponding capital letters by 0, the first expression in question reduces to $CE\dots F$. By further replacing C by A , and E,\dots,F by B , the first expression becomes AB . That is, the closed system possesses the function with expansion $AB:a+b$, i.e., the function AB . By the dual argument we see that if a closed $[\alpha,\beta,\gamma]$ system possesses an α -function whose normal expansion has in its second expression a term with more than one letter, it also possesses the function with expansion $A+B:ab$, i.e., the function $A+B$.

If then a closed $[\alpha,\beta,\gamma]$ system is not to be A_1 , and so is not to possess both AB and $A+B$, either the normal expansion of its α -functions all have first expressions with terms of one letter only, or second expressions with terms of one letter only. That is, either its α -functions are all reducible to first order functions and logical sum functions, or they are all reducible to first order functions and logical product functions.

As the β and γ -functions in the system must be reducible to first order, it follows that the system can then only be an O, R, S, or P system. We can therefore conclude that the only $[\alpha, \beta, \gamma]$ systems are O_8 , R_{10} , R_{11} , S_6 , P_6 , and A_1 .

§15

ALTERNATING SYSTEMS

A function will be said to satisfy the alternating condition if replacing an argument of the function by its negative either leaves the function unchanged, or changes it into its negative.⁴¹ By §4 the arguments of the first kind are the irrelevant variables of the function, those of the second kind its relevant variables. The set of relevant variables of an alternating function will be called its relevant subset. Clearly, if any number of arguments of an alternating function are changed into their negatives the function is left unchanged, or is turned into its negative, according as an even, or odd, number of variables in its relevant subset have been changed into their negatives. It follows that the alternating condition is iterative. For if $f(M, N, \dots, W)$ is an iterative process built up out of alternating functions and depending on variables A, B, ..., I, changing any one of these variables into its negative leaves M, N, \dots, W unchanged or turned into its negative, and hence leaves $f(M, N, \dots, W)$ unchanged or turned into its negative.

From the definition of an alternating function we see that a complete expansion corresponds to an alternating function when and only when interchanging any capital and corresponding small letter throughout the expansion either leaves the expansion unchanged, or has the effect of interchanging the two expressions of the expansion. Since in the first case the letter in ques-

⁴¹ For truth-functions a typical example of an 'alternating function' is $p = q$. Thus, reinterpreted as closed systems of truth-functions, L_2 below may be said to be generated by equivalence, L_3 by the dual, which is also the negative, of equivalence, L_1 by, for example, equivalence and negation.

tion corresponds to an irrelevant variable of the function, in the second to a relevant variable, it follows that two terms of the complete expansion of an alternating function which differ in the capitalization of but a single letter are in the same or different expressions of the expansion according as the letter corresponds to an irrelevant or relevant variable of the function. By successive application of this special result we obtain the following general result. Two terms of the complete expansion of an alternating function which differ in the capitalization of any number of letters are in the same or different expressions of the expansion according as an even or odd number of those letters correspond to variables in the relevant subset of the function.

In particular, a term of the complete expansion of an alternating function $f(A,B,\dots,I)$ is in the same expression as $AB\dots I$, or in the opposite expression, according as an even or odd number of the small letters in the term correspond to variables in the relevant subset of $f(A,B,\dots,I)$.

An alternating function $f(A,B,\dots,I)$ is therefore determined by its relevant subset, and the expression of its complete expansion in which $AB\dots I$ appears. Corresponding to any set of variables A,B,\dots,I , and subset of it chosen as the relevant subset, there are thus but two alternating functions, one with $AB\dots I$ in the first expression of its complete expansion, the other with $AB\dots I$, in the second expression.

Dual terms differ in the capitalization of all their letters. Hence, when the relevant subset of an alternating function has an even number of members, dual terms of the complete expansion of the function are always in the same expression, when odd, in different expressions.

Alternating functions with odd relevant subsets are therefore self-dual, and conversely. By considering the dual terms $AB\dots I$ and $ab\dots i$ we further see that alternating functions with even relevant subsets are β or γ -functions, those with odd relevant sub-

sets α or δ -functions.

It is interesting to note that if an expansion of an alternating function has an incomplete term, the letters missing in that term must correspond to irrelevant variables of the function. For among the terms of the complete expansion of the function which are contained in the incomplete term, and which are therefore in the same expression of the complete expansion, there will be at least one pair which differ only in the capitalization of any chosen one of those letters. Now we saw in §4 that a function admits an expansion involving only its relevant variables. In the case of an alternating function, therefore, each term of such an expansion must be a complete term on the relevant variables. If, then, we exclude the trivial case of expansions in which the same term appears more than once, we can conclude that an alternating function admits but one expansion involving only its relevant variables, and that is a complete expansion on those variables.

The same result of §4 shows that any function, and hence any alternating function, with not more than one relevant variable is reducible to first order. Conversely, it can be directly verified that any function reducible to first order is an alternating function with not more than one relevant variable.

That is, the class of functions reducible to first order is identical with the class of alternating functions having not more than one relevant variable. By examining the expansion of an alternating function which involves only the relevant variables of the function, we can then prove the following fact. An alternating function which satisfies any of the three conditions $[A:a]$, $[AA:]$, $[:aa]$ must be reducible to first order.

We turn now to the closed systems of alternating functions. As closed systems of functions reducible to first order have already been studied, it will be assumed that the systems under discussion have a function with relevant subset of more than one variable. Our derivation will fall into two parts according as the system has, or has not, a function with even relevant subset.

We shall constantly, though tacitly, be using the previously proved fact that an iterative process built up out of alternating functions yields an alternating function. The relevant subset of this function can in each instance be found by a direct application of the definition of an alternating function.

Consider then first a closed system of alternating functions which possesses a function with even relevant subset. By replacing all of the arguments of this function by A, a function $g(A)$ is obtained with null relevant subset. Now take a function in the system which has a relevant subset of more than one variable, and replace two of the variables in this subset by A and B respectively, all of the remaining arguments of the function by $g(A)$. We thus obtain a function $f(A,B)$, also in the system, with relevant subset (A,B) .

$f(A,A)$ is a function of A with null relevant subset, $f(A, f(A,A))$ with relevant subset (A) . If M is any alternating function of A, B, \dots, H , then $f(M, f(I, I))$ has the same relevant subset as M, $f(M, I)$ that of M with I added. By induction it follows that through $f(A, B)$ one of the two possible alternating functions corresponding to any set of variables A, B, \dots, I , and any subset of it chosen as the relevant subset, can be constructed.

If $f(A, B)$ has AB in the first expression of its complete expansion, then, by the iterative properties of uniform terms, this constructed function is the one with $AB \dots I$ in the first expression of its complete expansion. Hence, in this case, the system possesses all alternating functions of A, B, \dots, I which have $AB \dots I$ in the first expressions of their complete expansions. That is, the system possesses all α and β -alternating functions. Now γ and δ -alternating functions are the negatives of β and α -alternating functions respectively. Hence, if the system also possesses a δ -function, it will possess all γ and δ -alternating functions in addition to all α and β -alternating functions. The same is true if it possesses a γ -function. For, if we replace B by $\gamma_1(A)$, the complete expansion $AB+ab:Ab+aB$ of $f(A, B)$ becomes the complete expansion $a:A$ of $\delta_1(A)$. Hence, if the system does not consist of the α and β -alternating functions, it possesses all alternating functions. When $f(A, B)$ has AB in the second expression of

its complete expansion, it has ab also in the second expression, so that the dual results follow. We can therefore conclude that the following are the only closed systems of alternating functions which do not consist wholly of functions reducible to first order, and which possess a function with even relevant subset.

L_1 . All alternating functions.

L_2 . All α and β alternating functions.

L_3 . All α and γ alternating functions.

That these systems are closed follows from the iterative character of the alternating condition and the iterative properties of uniform terms.

Consider now a closed system of alternating functions which has only functions with odd relevant subsets. In a function in the system with odd relevant subset of more than one variable replace two of the relevant variables by A and B respectively, all the remaining arguments of the function by C. We thus obtain a function $f(A,B,C)$, also in the system, with relevant subset (A,B,C) .

$f(E,E,E)$ is a function of E with relevant subset (E).

If M is any alternating function with odd relevant subset, then if G is not an argument of M, $f(M,G,G)$ has the same relevant subset as M, while if H and J are not arguments of M, $f(M,H,J)$ has for relevant subset that of M with H and J added. It follows by induction that through $f(A,B,C)$ one of the two possible alternating functions corresponding to any set of variables, and any odd subset of it chosen as relevant, can be constructed.

We have seen that the alternating functions with odd relevant subsets are the α and δ alternating functions. Whatever function is constructed through $f(A,B,C)$ in the above manner, the negative of that function is the other possible alternating function corresponding to the given set of variables, and odd relevant subset thereof. Hence, if the system has a δ -function, it will have all alternating functions with odd relevant subsets, i.e., all α and δ alternating functions.

Otherwise the system can only have α -functions. $f(A,B,C)$ is then an α -function so that each of the above functions constructed by means of $f(A,B,C)$ is an α -function. But only one of the two possible alternating functions corresponding to a given set of variables and odd relevant subset thereof can be an α -function. The system therefore has all of the alternating functions with odd relevant subsets that are α -functions, i.e., it has all α alternating functions. Hence the only possible closed systems of alternating functions which do not consist wholly of functions reducible to first order, and which do not have a function with even relevant subset, are the following.

L_4 . All α alternating functions.

L_5 . All α and δ alternating functions.

That L_4 is closed is seen immediately. On the other hand, L_5 can be described as the system of all self-dual alternating functions, and hence is closed because of self-duality, and the alternating condition, are iterative.

Our derivation shows that L_1 , L_2 , and L_3 can be generated by the function $f(A,B)$ with the possible addition of $\delta_1(A)$.

Hence L_1 , L_2 , and L_3 are of the second order.

Since a second order alternating function with odd relevant subset is reducible to first order, L_4 and L_5 cannot be of the second order. They can however be generated by the above mentioned $f(A,B,C)$ with the possible addition of $\delta_1(A)$.

Hence L_4 and L_5 are of the third order.

Note that L_2 and L_3 are dual systems, while L_1 , L_4 and L_5 are self-dual, the last two in fact consisting wholly of self-dual functions.

L_1 , L_2 , L_3 , L_4 , L_5 are then the only closed systems of alternating functions which do not consist wholly of functions

reducible to first order. The O, R, and L systems therefore constitute all of the closed alternating systems, i.e., closed systems of alternating functions.

§16

FAILURE OF THE ALTERNATING CONDITION

An argument E of an alternating function $f(A,B,\dots,I)$ is evidently an irrelevant or relevant variable of the function according as the term whose only small letter is e is in the same expression of the complete expansion of $f(A,B,\dots,I)$ as AB...I or in the opposite expression. Hence, two terms of the complete expansion of an alternating function $f(A,B,\dots,I)$ which differ in the capitalization of but a single letter E are in the same or different expressions of the expansion according as AB...I and the term whose only small letter is e are in the same or different expressions. Conversely, a complete expansion which satisfies this condition for each of the letters on which it is written must correspond to an alternating function. For, on interchanging any capital and corresponding small letter in the expansion, either every term of the expansion is changed into a term of the same expression, or every term is changed into a term of the opposite expression, i.e., either the expressions are unchanged or interchanged.

In the complete expansion of a function $f(A,B,\dots,I)$ let t_0 and t_1 respectively be the uniform term AB...I and a term with but one small letter, t_2 and t_3 any other pair of terms which differ only in the capitalization of that letter. Then, if $f(A,B,\dots,I)$ does not satisfy the alternating condition, for some quadruplet (t_0, t_1, t_2, t_3) either t_0 and t_1 are in the same expression, t_2 and t_3 in different expressions, or the reverse. That is, for some quadruplet (t_0, t_1, t_2, t_3) three terms are in one expression, the fourth in the other. Now the variables A,B,...I fall into at most three groups with respect to their capitalization in those terms, to wit,

t_0	capital	capital	capital
t_1	capital	capital	small
t_2 or t_3	capital	small	capital
t_3 or t_2	capital	small	small

While no variable may belong to the first group, the last two groups cannot be empty or the four terms would not be distinct. Now in the function $f(A, B, \dots, I)$ replace any of the arguments that may fall in the first group by A , those in the second and third groups by B and C respectively. We thus either get a function $g(B, C)$ with three terms of its complete expansion in one expression, the fourth in the other, or a function $h(A, B, C)$ with an A -component of this type. Hence the following.

If a closed system possesses a function which fails to satisfy the alternating condition, then it also has either a second order function with an odd number of terms in each expression of its complete expansion, or a third order function whose A -component is of this type.

If, furthermore, the system possesses a β -function we can modify the above argument by replacing the variables that may fall in the first group by $\beta_1(B)$, and thus be sure to get the second order function in question.

§17

 $[\alpha, \beta, \gamma, \delta]$ SYSTEMS

The maximal closed $[\alpha, \beta, \gamma, \delta]$ system is the complete system

C_1 . All functions.

Since the first expression of an expansion of a function is the logical sum of logical products of variables and their negatives, we see that the functions with expansions $a:A$, AB : $a+b$, $A+B:ab$ constitute a set of generators of C_1 .

It follows that C_1 is a closed system of the second order.

We now prove that $\delta_1(A)$ and any function $f(A,B)$ whose complete expansion has a one term expression generate C_1 .

Note that $\delta_1[f(A,B)]$ has a complete expansion whose other expression consists of a single term. Now in the function with the one term first expression replace each argument that appears as a small letter in that term by its negative; in the function with the one term second expression replace each argument that appears as a capital letter in that term by its negative. The one term first expression thus becomes AB , the one term second expression ab . That is, by means of $\delta_1(A)$ and $f(A,B)$ we can construct the functions with expansions $AB:a+b$ and $A+B:ab$. As these functions together with $\delta_1(A)$ generate C_1 , $\delta_1(A)$ and $f(A,B)$ also generate C_1 .

Since an $[\alpha, \beta, \gamma, \delta]$ system must have a β -function, it follows from the preceding section that every closed $[\alpha, \beta, \gamma, \delta]$ system which does not consist wholly of alternating functions possesses such a second order function $f(A,B)$. Since it also possesses a δ -function, it must contain C_1 , and hence is C . Every closed $[\alpha, \beta, \gamma, \delta]$ system other than C_1 is therefore an alternating system. It follows that the only closed $[\alpha, \beta, \gamma, \delta]$ systems are O_9 , R_{12} , R_{13} , L_1 , and C_1 .

Pausing a moment to review our progress in determining all closed systems through their associated first order systems, we see that there remains the complete determination of all closed $[\alpha]$, $[\alpha, \beta]$, $[\alpha, \gamma]$, and $[\alpha, \delta]$ systems. The corresponding discussion constitutes the most significant part of our development.

§18

[α] SYSTEMS; PRELIMINARY DISCUSSION

An α -function $f(A,B,\dots,I)$ has the dual terms $AB\dots I$ and $ab\dots i$ in the first and second expressions respectively of its complete expansion. If $f(A,B,\dots,I)$ is not self-dual, it

will then have some other pair of dual terms in one and the same expression of its complete expansion. Now in the function $f(A, B, \dots, I)$ replace the arguments that appear as capital letters in one of these dual terms, and hence as small letters in the other, by A , those that appear as small letters in the first of these terms, and hence as capital letters in the second, by B . The two terms become Ab and aB respectively, and hence the complete expansion $AB+Ab+aB:ab$ or $AB:Ab+aB+ab$. The function $f(A, B, \dots, I)$, therefore, either reduces to the function with the expansion $A+B:ab$ or to the function with the expansion $AB:a+b$. It becomes convenient to use an expansion of a function which involves all of the arguments of the function as a symbol for the function.

We can then state that every closed $[\alpha]$ system which does not consist wholly of self-dual functions possesses either the function $A+B:ab$ or the function $AB:a+b$.

Note that $A+B:ab$ satisfies the $[:aa]$ condition but not the $[AA:]$ condition, while $AB:a+b$ satisfies the $[AA:]$ condition but not the $[:aa]$ condition. Now suppose that a closed $[\alpha]$ system which possesses the function $A+B:ab$ fails to satisfy the $[:aa]$ condition, i.e., possesses a function which fails to satisfy the $[:aa]$ condition. Then some function in the system has a pair of terms in the second expression of its complete expansion with respect to which the arguments of the function fall into at most three groups as follows:

capital	capital	small
capital	small	capital

The last two groups cannot be empty, or the term consisting of capital letters only would be in the second expression of this complete expansion, in contradiction to the function in question being an α -function. Now replace the arguments of this function that are in the second and third group by A and B respectively, those, if any, that are in the first group by $A+B$. When the first group is not empty, the above two terms become $(A+B)Ab$

and $(A+B)aB$, and hence, in any case, reduce to Ab and aB . The function therefore reduces to the function $AB:Ab+aB+ab$, i.e., to the function $AB:a+b$. The given closed $[\alpha]$ system therefore possesses the function $AB:a+b$. By the dual argument, if a closed $[\alpha]$ system possesses the function $AB:a+b$, but does not satisfy the $[AA:]$ condition, then it also must have the function $A+B:ab$.

Hence, every closed $[\alpha]$ system which does not consist wholly of self-dual functions and which does not satisfy either the $[:aa]$ or the $[AA:]$ condition possesses both the function $A+B:ab$ and the function $AB:a+b$.

By referring to §14 we see that the functions $A+B:ab$, $AB:a+b$ generate the second order closed system:

A_4 . All α -functions that satisfy the $[A:a]$ condition.

We now prove that the only other closed $[\alpha]$ system which possesses both of the functions $A+B:ab$, $AB:a+b$ is the complete $[\alpha]$ system:

C_4 . All α -functions.

The system in question must have an α -function which fails to satisfy the $[A:a]$ condition. There will therefore be a term in the first expression and a term in the second expression of the complete expansion of this function with respect to which the arguments of the function fall into at most three groups as follows:

term in the first expression : capital small small
term in the second expression : capital small capital

The third group cannot be empty, or the two terms would not be distinct; the first and second group cannot be empty, or the function would not be an α -function. Now replace the arguments of the function that are in the first group by $A+B+\dots+I$, those

that are in the second group by $AB\dots I$, those that are in the third group by any function M in the given system whose arguments constitute a subset of the set of variables (A, B, \dots, I) . The two terms in question become $(A+B+\dots+I)(a+b+\dots+i)m$, $(A+B+\dots+I)(a+b+\dots+i)M$ respectively, the given function a function of A, B, \dots, I which may be written $g(A, B, \dots, I, M)$, and which is also in the given system. Now every complete term on A, B, \dots, I other than $AB\dots I$ and $ab\dots i$ is contained in $(A+B+\dots+I)(a+b+\dots+i)$. Hence $g(A, B, \dots, I, M)$ has the property that every complete term $t(A, B, \dots, I)$ other than $AB\dots I$ and $ab\dots i$ is contained in corresponding expressions of expansions of $g(A, B, \dots, I, M)$ and $\delta_1(M)$.

By a result of the preceding section, we see that $\delta_1(A)$ and $\pi(A, B)$, the logical product of A and B , generate the complete system C_1 . Consider then any α -function $f(A, B, \dots, I)$, and any mode of building up this α -function by means of $\delta_1(A)$ and $\pi(A, B)$. Such a process is an operation in the sense of §1. We now inductively define a corresponding operation built up by means of functions in the system under discussion, and hence yielding a function in this system. Components of rank zero are to be the same in the new operation as in the old. Assume that components of the new operation have been defined which are to correspond to components of rank less than p of the old. Then to a component of rank p of the old operation which is of the form $\pi(M, N)$ we make correspond a component $\pi(M', N')$ of the new operation, where M' and N' correspond to M and N respectively; to a component of rank p of the form $\delta_1(M)$ we make correspond a component $g(A, B, \dots, I, M')$, where M' corresponds to M . We thus ultimately obtain the new operation that is to correspond to the old. Now let M, N, M', N' be the variables or functions yielded by the operations M, N, M', N' respectively. Then by §5 any complete term $t(A, B, \dots, I)$ which is contained in corresponding expressions of expansions of M and M' , N and N' , is contained in corresponding expressions of expansions of $\pi(M, N)$ and $\pi(M', N')$. And, as we have just seen, any complete term $t(A, B, \dots, I)$ other than $AB\dots I$ or $ab\dots i$ is contained in corresponding expressions of expansions of $\delta_1(M)$ and $g(A, B, \dots, I, M')$, and hence, if contained in corresponding expressions of expansions of M and M' , is contained

§19. $[\alpha, \beta]$ AND $[\alpha, \gamma]$ SYSTEMS: PRELIMINARY DISCUSSION 67

in corresponding expressions of expansions of $\delta_1(M)$ and $g(A, B, \dots, I, M')$. It follows by induction that every complete term $t(A, B, \dots, I)$ other than $AB \dots I$ and $ab \dots i$ is contained in corresponding expressions of expansions of $f(A, B, \dots, I)$ — the function yielded by the old operation — and the function of A, B, \dots, I yielded by the new operation. But the same must be true of the complete terms $AB \dots I$ and $ab \dots i$ since both functions are α -functions. It follows that the complete expansions of the functions are identical, and hence the functions are identical. The new operation therefore yields the same function $f(A, B, \dots, I)$ as the old.

Since $f(A, B, \dots, I)$ is any α -function of A, B, \dots, I , the closed $[\alpha]$ system under discussion possesses all α -functions, and hence is none other than C_4 .

This proof shows incidentally that C_4 can be generated by $A+B:ab$, $AB:a+b$, and any α -function that fails to satisfy the $[A:a]$ condition. Our observation that all three groups of variables associated with two terms resulting in this failure must be existent shows that while this non- $[A:a]$ α -function may be of the third order, every α -function of order less than three does satisfy the $[A:a]$ condition. It follows that C_4 is of the third order.

As a result of this preliminary discussion we can state that every closed $[\alpha]$ system other than A_4 and C_4 either consists wholly of self-dual functions, or else satisfies either the $[:aa:]$ or the $[AA:]$ condition.

§19

$[\alpha, \beta]$ AND $[\alpha, \gamma]$ SYSTEMS; PRELIMINARY DISCUSSION

As was seen in §10, the closed $[\alpha, \gamma]$ systems are the duals of the closed $[\alpha, \beta]$ systems. Our explicit development may therefore be restricted to closed $[\alpha, \beta]$ systems. Since $\beta_1(A)$, whose complete expansion is $A+a:0$, satisfies the $[:aa:]$ condition, but is not self-dual and does not satisfy the $[AA:]$ condition, a closed $[\alpha, \beta]$ system may satisfy the $[:aa:]$ condition,

but does not consist wholly of self-dual functions and cannot satisfy the $[AA:]$ condition. There results in consequence a certain analogy between the discussion of closed $[\alpha, \beta]$ systems, and closed $[\alpha]$ systems possessing the function $A+B:ab$. The purpose of our preliminary discussion is to derive all of the closed $[\alpha, \beta]$ systems which do not satisfy the $[:aa]$ condition.

As in the case of closed $[\alpha]$ systems some function in such a closed $[\alpha, \beta]$ system will fail to satisfy the $[:aa]$ condition, and for this function the arguments will fall into at most three groups

capital	capital	small
capital	small	capital

with respect to some pair of terms in the second expression of the complete expansion of the function. By replacing the arguments of the function that are in the second and third groups, again necessarily present, by A and B respectively, those, if any, that are in the first group by $\beta_1(A)$, the two terms become Ab and aB . As the resulting function may now be either an α or a β -function, the resulting complete expansion is either $AB:Ab+aB+ab$ or $AB+ab:Ab+aB$.

Hence, every closed $[\alpha, \beta]$ system which does not satisfy the $[:aa]$ condition possesses either the function $AB:a+b$ or the function $AB+ab:Ab+aB$.

Consider first a closed $[\alpha, \beta]$ system which possesses the function $AB:a+b$, but fails to satisfy the $[A:a]$ condition. Then for some function in the system there will be a term in the first expression and a term in the second expression of the complete expansion of the function with respect to which the arguments of the function fall into at most three groups as follows:

term in the first expression :	capital small small
term in the second expression :	capital small capital.

While the last two groups cannot be empty, the first group may

now be empty. By replacing the arguments that may fall in the first group by $\beta_1(A)$, those in the second by $AB\dots I$, those in the third by any function M in the system whose arguments constitute a subset of the set of variables (A, B, \dots, I) , the two terms in question become $(a+b+\dots+i)m$, $(a+b+\dots+i)M$ respectively. Since every complete term $t(A, B, \dots, I)$ other than $AB\dots I$ is contained in $a+b+\dots+i$, the resulting function $h(A, B, \dots, I, M)$ has the property that every complete term $t(A, B, \dots, I)$ other than $AB\dots I$ is contained in corresponding expressions of expansion of $h(A, B, \dots, I, M)$ and $\delta_1(M)$. As in the derivation of the closed $[\alpha]$ system C_4 , we therefore conclude that every function of A, B, \dots, I whose complete expansion has $AB\dots I$ in its first expression can be generated by means of $h(A, B, \dots, I, M)$ and $AB:a+b$, and hence is in the system under discussion.

Hence, the only closed $[\alpha, \beta]$ system which possesses the function $AB:a+b$, but fails to satisfy the $[A:a]$ condition, is the complete $[\alpha, \beta]$ system C_2 . All α and β -functions.

As the above function which fails to satisfy the $[A:a]$ condition may now be a β -function of order two, it follows that C_2 is of order two.

All other closed $[\alpha, \beta]$ systems that possess the function $AB:a+b$ therefore satisfy the $[A:a]$ condition. If the α -functions in such a system all have normal $[A:a]$ expansions with second expressions consisting of terms of one letter only, then the system must be the logical product system P_5 . Otherwise, the $[\alpha, \beta, \gamma]$ discussion of §14 reveals that from $\beta_1(A)$, and an $[A:a]$ α -function whose normal expansion has a term with more than one letter in its second expression, the function $A+B:ab$ can be derived. Since every β -function that satisfies the $[A:a]$ condition is reducible to first order, it follows that the system must now be

A_2 . All α and β -functions that satisfy the $[A:a]$ condition.

We may write $A_2 = A_4 + R_2$. A_2 is therefore of the second order.

P_5 and A_2 are thus the only closed $[\alpha, \beta]$ systems that possess the function $AB:a+b$ and satisfy the $[A:a]$ condition. Hence, the only closed $[\alpha, \beta]$ systems possessing the function $AB:a+b$ are P_5 , A_2 , and C_2 .

We turn now to closed $[\alpha, \beta]$ systems possessing the function $AB+ab:Ab+aB$. This function is the β alternating function of A and B with relevant subset (A, B) . It therefore generates the alternating system L_2 . Since L_2 possesses all α and β alternating functions, any other closed $[\alpha, \beta]$ system possessing the function $AB+ab:Ab+aB$ cannot consist wholly of alternating functions, and hence, by §16, must possess a second order function with a complete expansion having a one term expression. Now the only complete expansions of this type that correspond to α and β -functions of two variables A , B are $AB:Ab+aB+ab$, $AB+Ab+aB:ab$, $AB+aB+ab:Ab$, $AB+Ab+ab:aB$, which reduce to $AB:a+b$, $A+B:ab$, $a+B:Ab$, $A+b:aB$ respectively. If in $AB+ab:Ab+aB$ the variables $A:a$ and $B:b$ are replaced by $A+B:ab$ and $AB+ab:Ab+aB$ respectively, the function $AB:a+b$ results. The same is true if $A:a$ and $B:b$ are replaced by $a+B:Ab$ and $A:a$ respectively, and by $A+b:aB$ and $B:b$ respectively. We thus see that every closed $[\alpha, \beta]$ system other than L_2 that has the function $AB+ab:Ab+aB$ has the function $AB:a+b$. Since $AB+ab:Ab+aB$ does not satisfy the $[A:a]$ condition, it follows from the earlier discussion that the system must then be C_2 .

The only closed $[\alpha, \beta]$ systems possessing the function $AB+ab:Ab+aB$ are therefore L_2 and C_2 .

Returning to the initial result of this section we thus see that every closed $[\alpha, \beta]$ system other than P_5 , A_2 , L_2 , and C_2 satisfies the $[:aa]$ condition. The duals of C_2 and A_2 are the second order $[\alpha, \gamma]$ systems

C_3 . All α and γ -functions.

A_3 . All α and γ -functions that satisfy the $[A:a]$ condition.

Note also that $A_3 = A_4 + R_3$. As the dual of an $[:aa]$ function is an $[AA:]$ function, we can state immediately that every

closed $[\alpha, \gamma]$ system other than S_5 , A_3 , L_3 , and C_3 satisfies the $[AA:]$ condition.

§20

$[\alpha]$ SYSTEMS OF SELF-DUAL FUNCTIONS

Consider first a closed $[\alpha]$ system, consisting wholly of self-dual functions, which does not satisfy the $[A:a]$ condition. Again we take some function in the system that fails to satisfy the $[A:a]$ condition, and choose a term in the first expression and a term in the second expression of the complete expansion of the function that lead to this failure, and note the three fold grouping of the arguments of the function with respect to their capitalization in those terms;

term in the first expression	:	capital small small
term in the second expression	:	capital small capital.

As in the preliminary discussion of $[\alpha]$ systems, none of these groups can be empty. By replacing the arguments of the function by A, B, C according as they are in the first, second or third group respectively, the term in the first expression becomes Abc , in the second, AbC . Since the resulting function of A, B, C is self-dual, the duals of these terms, aBC and aBc , must be in the second and first expressions respectively of the complete expansion of this function. As it is also an α -function, it can only be one of the following two functions with complete expansions as indicated.

$$f_1(A, B, C); ABC + Abc + aBc + ABc : abc + aBC + AbC + abC.^{42}$$

$$f_2(A, B, C); ABC + Abc + aBc + abC : abc + aBC + AbC + ABC.$$

⁴² This function is Kempe's unsymmetrical resultant [13]. In view of the result following, Kempe's corresponding theory may be said to be about all self-dual α -functions. Likewise the function $h(A, B, C)$, shortly to be met, is Kempe's symmetric resultant. The corresponding theory is therefore that of all self-dual $[A:a]$ α -functions. On the other hand, triadic-rejection (see Frink [8], p. 487; also earlier pages) is seen by the discussion concluding the next section to generate all self-dual

Suppose that $f_1(A,B,C)$ is thus obtained. Then $f_1(A,B,C)$ is in the system under discussion, and hence $f_1(C,B,A)$ is also in the system. $f_1(A,B,C)$ has for A-component the function $BC+bc+Bc:bC$, i.e., $B+c:bC$; $f_1(C,B,A)$ has for A-component the function $BC:Bc+bC+bc$, i.e., $BC:b+c$. Now $B+c:bC$ is a β -function that does not satisfy the $[A:a]$ condition. Hence, by the preliminary $[\alpha,\beta]$ discussion, it, in conjunction with $BC:b+c$, generates the complete $[\alpha,\beta]$ system C_2 . It follows from the theorem of §9 that $f_1(A,B,C)$ and $f_1(C,B,A)$ can generate a function $f(A,B,C\dots I)$ with arbitrary α or β A-component $\phi(B,C,\dots I)$. Clearly $f(A,B,C,\dots I)$ must be a self-dual α -function. Now every α -function of $A,B,C,\dots I$ has for A-component some α or β -function of $B,C,\dots I$. Furthermore, the a-component of a self-dual function of $A,B,C,\dots I$ is easily seen to be the dual of its A-component, so that there can be but one self-dual function of $A,B,C,\dots I$ with given A-component. It follows that the above functions $f(A,B,C\dots I)$ constitute all self-dual α -functions of $A,B,C\dots I$. The given closed $[\alpha]$ system of self-dual functions must therefore be

D₁. All self-dual α -functions.

Suppose, on the other hand, that $f_2(A,B,C)$ was obtained. $f_2(A,B,C)$ is the α alternating function of A,B,C with relevant subset (A,B,C) . It therefore generates L_4 , which consists of all α alternating functions. If then the given closed system is not L_4 , it cannot consist wholly of alternating functions. It will therefore possess, in accordance with §16, a function $f_3(A,B,C)$ whose A-component has a complete expansion with a one term expression. That this is the only possible conclusion in the present case follows from the fact that a self-dual function of the second order would have two terms in each expression of its complete expansion. Now $f_2(A,B,C)$ has for A-component the function $BC+bc:bC+Bc$. This, with the A-compon-

functions. Interesting, but outside the scope of the present paper, is [25] by Royce. The sort of diagram given by Kempe, [13], p. 172, for symmetric resultant was invaluable to the writer in originally obtaining the constitution of the various generated systems.

ent of $f_3(A, B, C)$, would generate C_2 by the $BC+bc:bC+Bc$ discussion of the preceding section. Hence, as before, the given system would be D_1 .

It therefore follows that L_4 and D_1 are the only closed $[\alpha]$ systems of self-dual functions that do not satisfy the $[A:a]$ condition.

We turn now to closed $[\alpha]$ systems of self-dual functions that do satisfy the $[A:a]$ condition. It is interesting to observe in connection with our preliminary $[\alpha]$ systems discussion that by §8 these systems can also be described as the closed $[\alpha]$ systems that satisfy both the $[:aa]$ and $[AA:]$ condition.

Note first that if a self-dual α -function which satisfies the $[A:a]$ condition has a term with but one small letter in the second expression of its complete expansion, then it must be reducible to first order. For if that small letter be e , the normal $[A:a]$ expansion of the function will have e for a term of its second expression. As the function is self-dual, E will be a term of the first expression of that normal expansion which therefore must be $E:e$. Now O_1 and R_1 , the only closed $[\alpha]$ systems of functions reducible to first order, also happen to consist of self-dual functions that satisfy the $[A:a]$ condition. Every other closed $[\alpha]$ system of the type we are considering must therefore possess a function whose complete expansion has every term with but one small letter in its first expression.

This function cannot be of the first or second order, as the only self-dual α -functions of A and of A, B are $A:a$, $AB+Ab:ab+aB$, $AB+aB:ab+Ab$ which, in fact, are reducible to first order. If of the third order, with arguments A, B, C , the function can only be $AB+BC+AC:ab+bc+ac$, since the first expression of its complete expansion must be $ABC+ABC+aBC+AbC$. Suppose now that the function is of order n , $n > 3$. Its complete expansion will then have at least one term such that it, and hence also its dual, have at least two capital letters and at least two small letters. From such a pair of dual terms choose the term that is in the first expression of the complete expansion; and in the given function replace each argument that

appears as a small letter in that term by A, the remaining arguments be B,C...G respectively. The term in question becomes ABC...G. The remaining terms of the resulting complete expansion which have but one small letter are all derived from terms of the same type in the original complete expansion. Hence the resulting function of A,B,C,...G, which must be of order n' with $n > n' \geq 3$, also has a complete expansion in which every term with but one small letter is in the first expression. By repetition of this process we must finally obtain the third order function of this type.

Hence, every closed $[\alpha]$ system of self-dual functions that satisfies the $[A:a]$ condition, if not O_1 , or R_1 , possesses the function $AB+BC+AC:ab+bc+ac$.

We now prove that this function generates the closed system

D_2 . All self-dual α -functions that satisfy the $[A:a]$ condition.

It will then follow that O_1 , R_1 , and D_2 are the only closed $[\alpha]$ systems of self-dual functions that satisfy the $[A:a]$ condition, and hence that O_1 , R_1 , D_2 , L_4 , and D_1 are the only closed $[\alpha]$ systems consisting wholly of self-dual functions.

Let the function $AB+BC+AC:ab+bc+ac$ be symbolized $h(A,B,C)$. Since $h(A,B,C)$ satisfies the conditions defining D_2 , and those conditions are iterative, every function generated by $h(A,B,C)$ must be in the resulting closed system D_2 . Note that $h(A,A,B)$ admits the expansion $A:a$. Hence $h(A,B,C)$ can generate every function in D_2 that is reducible to first order. Since a function is identically equal to the first expression of any expansion thereof, we can write $h(T,S_1,S_2) = T(S_1+S_2)+S_1S_2$, and obtain by induction, for $n \geq 2$,

$$\begin{aligned} h_n(T,S_1,S_2,\dots,S_n) &= h(T,S_1,h(T,S_2,\dots,h(T,S_{n-1},S_n)\dots)) \\ &= T(S_1+S_2+\dots+S_n)+S_1S_2\dots S_n \end{aligned}$$

⁴³ Cf. Frink [8], p. 482.

Let M be any function in D_2 that is not reducible to first order, N any function on the same arguments that can be generated by $h(A,B,C)$. We shall call the first expression of the normal expansion of an $[A:a]$ function the normal form of the function. Then any term of the normal form of M must have at least two letters; for if that normal form had a one letter term E , self-duality would make the normal expansion $E:e$. Let $BD\dots G$ be any term of the normal form of M that is not wholly contained in N . As that term has v letters, $v \geq 2$, we can form the function N' , also generated by $h(A,B,C)$, as follows.

$$N' = h(N, B, D, \dots, G) = N(B+D+\dots+G)+BD\dots G.$$

Since D_2 must also satisfy the $[AA:]$ condition, every term of the normal form of M has a letter in common with the term $BD\dots G$, and hence is contained in $B+C+\dots+G$. It follows that N' contains every term of the normal form of M that N contains, and, in addition, the term $BC\dots G$. By repetition of this process we thus finally obtain a function N_0 , generated by $h(A,B,C)$, on the same arguments as M , and containing M . Now turn to the complete expansions of M and N_0 . Every term of the first expression of the complete expansion of M is then in the first expression of the complete expansion of N_0 . Since M and N_0 are self-dual, these first expressions must each consist of half the total number of terms of the corresponding complete expansions. The first expressions of the complete expansions of M and N_0 , are therefore identical, and hence M and N_0 are identical. $h(A,B,C)$ can therefore also generate every function in D_2 that is not reducible to first order. The closed system generated by $h(A,B,C)$ is therefore D_2 .

The new closed systems D_1 and D_2 found in this section are easily seen to be of the third order, for they have been generated by functions of order three, and cannot be generated by first and second order functions since, as we saw above, first and second order self-dual α -functions are reducible to first order, and hence can but generate O_1 or R_1 .

§21

[α, δ] SYSTEMS

We first observe that a closed [α, δ] system must consist wholly of self-dual functions.

For the negative of a non-self-dual δ -function is a non-self-dual α -function. By §18 such an α -function yields either the function $A+B:ab$ or the function $AB:a+b$. As either of these functions in conjunction with $\delta_1(A)$ generates the [$\alpha, \beta, \gamma, \delta$] system C_1 , the closed [α, δ] system cannot have either a non-self-dual α -function or a non-self-dual δ -function.

Secondly, the δ -functions possessed by a closed [α, δ] system are the negatives of the α -functions in the system. For the negative of an α -function is a δ -function, and every δ -function is the negative of that α -function which is the negative of the given δ -function. Now the α -functions in a closed [α, δ] system must themselves constitute a closed [α] system. As, by our first observation, this closed [α] system must consist wholly of self-dual functions, it follows that every closed [α, δ] system consists of the functions in a closed [α] system of self-dual functions and their negatives.

We saw, in the preceding section, that the only closed [α] systems of self-dual functions are O_1 , R_1 , D_2 , L_4 , and D_1 . Now D_1 consists of all self-dual α -functions. Since every self-dual function is either an α or a δ -function, and the condition of self-duality is iterative, it follows that the functions in D_1 with their negatives constitute the closed [α, δ] system

D_3 All self-dual functions.

D_3 is then the maximal [α, δ] system. As with D_1 it is easily verified that D_3 is of order three.

The functions in O_1 , R_1 , and L_4 with their negatives respectively constitute the known closed [α, δ] systems O_4 , R_4 , and L_5 . On the other hand, the functions in D_2 with

their negatives do not constitute a closed system. For if in the function $AB+BC+AC:ab+bc+ac$, which generates D_2 , we replace C by its negative, we obtain the function $AB+Bc+Ac:ab+bC+aC$ which is an α -function, yet, as it does not satisfy the $[A:a]$ condition, is not in D_2 . Hence, O_4 , R_4 , L_5 , and D_3 are the only closed $[\alpha, \delta]$ systems.

§22

$[\alpha]$, $[\alpha, \beta]$, AND $[\alpha, \gamma]$ SYSTEMS; FURTHER DISCUSSION

Since we have derived all closed $[\alpha]$ systems of self-dual functions, we see from our preliminary discussion of $[\alpha]$, $[\alpha, \beta]$, and $[\alpha, \gamma]$ systems that the derivation of all closed systems of these types, and hence of all types, will be completed when we have determined all closed systems that satisfy the $[:aa]$ or the $[AA:]$ condition.

As these conditions are duals of each other, it will be sufficient to derive explicitly all closed systems that satisfy the $[:aa]$ condition.

Since $AB:a+b$ does not satisfy the $[:aa]$ condition, it follows from our preliminary $[\alpha]$ systems discussion that every closed $[\alpha]$ system that satisfies the $[:aa]$ condition, but does not consist wholly of self-dual functions, possesses the function $A+B:ab$.

This useful result is made all the more useful by the following observation. The closed $[\alpha]$ systems of self-dual functions that satisfy the $[:aa]$ condition are at the same time the closed $[\alpha]$ systems of self-dual functions that satisfy the $[A:a]$ condition. Hence, by §20, the only closed $[\alpha]$ systems that satisfy the $[:aa]$ condition and do consist wholly of self-dual functions are O_1 , R_1 , and D_2 .

$\gamma_1(A)$ and $\delta_1(A)$ do not satisfy the $[:aa]$ condition. Hence, every closed system that satisfies the

$[:aa]$ condition is either an $[\alpha]$ or $[\alpha,\beta]$ system.

We can therefore classify all closed systems that satisfy the $[:aa]$ condition as follows.

1. $[:aa]$ $[\alpha]$ systems not satisfying the $[A:a]$ condition.
2. $[:aa]$ $[\alpha]$ systems satisfying the $[A:a]$ condition.
3. $[:aa]$ $[\alpha,\beta]$ systems satisfying the $[A:a]$ condition.
4. $[:aa]$ $[\alpha,\beta]$ systems not satisfying the $[A:a]$ condition.

This fourfold classification is fundamental both in the present and in the following section.

In the present section the following specialization of the $[:aa]$ condition will be required. A function will be said to satisfy the $[:a_\infty]$ condition if all the terms of the second expression of an expansion of the function have a small letter in common.⁴⁴ As in §8 it is readily verified first, that the $[:a_\infty]$ condition is independent of the particular expansion used, second, that the $[:a_\infty]$ condition is iterative. We can therefore also refer to a closed system satisfying the $[:a_\infty]$ condition. Clearly, the $[:a_\infty]$ condition for function or closed system implies the $[:aa]$ condition.

If the common small letter referred to in the $[:a_\infty]$ condition is a , it will be convenient to say the function satisfies the $[:a_\infty]$ condition with respect to a . This too is independent of the particular expansion used. Note that for such a function, every term of the complete expansion of the function that involves the capital letter A must be in the first expression of that complete expansion.

Hence, a function that satisfies the $[:a_\infty]$ condition with respect to a is determined by its a -component.

In applying the mechanism of a -components to closed sys-

⁴⁴ A typical and important example is the function $A + b : aB$ shown below to generate all functions satisfying the $[:a_\infty]$ condition. Reinterpreted as truth-function of propositions, this function is the familiar 'implication' of Principia Mathematica.

tems that satisfy the $[:a_\infty]$ condition, as is thus suggested, the following observation will prove to be indispensable.

If each function in a given set of generators satisfies the $[:a_\infty]$ -condition with respect to a , then every function that can be generated by those generators through iterative processes in which no replacement is effected upon the variable A also satisfies the $[:a_\infty]$ condition with respect to a .

We now take up each of the four cases under which a closed system satisfying the $[:aa]$ condition may be classified, and obtain for each case a result which in this section will yield all closed systems satisfying the $[:a_\infty]$ condition, and which at the same time will furnish the necessary starting point for the concluding work of the next section.

1. $[:aa]$ $[\alpha]$ systems not satisfying the $[A:a]$ condition.

As in §20, except for the interchange of A and B , we find that every closed system of this type possesses a function $f_1(A, B, C)$ with the terms aBC and aBC in the first and second expressions respectively of its complete expansion. Since abc must be in the second expression of this complete expansion, the a -component of $f_1(A, B, C)$ is a γ -function that fails to satisfy the $[A:a]$ condition. As the closed system under discussion cannot consist wholly of self-dual functions, it must have the function $A+B:ab$. It therefore also possesses the function $f_2(A, B, C)$ with expansion $A+B+C:abc$. $f_2(A, B, C)$ then has for a -component the function $B+C:bc$. By the dual of the explicitly given preliminary $[\alpha, \beta]$ discussion we see that these a -components generate the complete $[\alpha, \gamma]$ system C_3 . Hence, by the theorem of §9, $f_1(A, B, C)$ and $f_2(A, B, C)$ can generate a function $f(A, B, C, \dots, I)$ with arbitrary α or γ a -component $\phi(B, C, \dots, I)$.

Since $f_1(A, B, C)$ has the term aBC in the second expression of its complete expansion, and satisfies the $[:aa]$ condition, it follows that every term of this second expression must have the small letter a in common with the term aBC . That is, $f_1(A, B, C)$ satisfies the $[:a_\infty]$ condition with respect to a .

The same is evidently true of $f_2(A, B, C)$. Now by §9 $f(A, B, C, \dots, I)$ can be generated by $f_1(A, B, C)$ and $f_2(A, B, C)$ through iterative processes in which no replacement is effected upon the variable A. Hence $f(A, B, C, \dots, I)$ also satisfies the $[:a_\infty]$ condition with respect to a. Since every α -function of A, B, C, ... I has for a-component some α or γ -function of B, C, ... I, while a function that satisfies the $[:a_\infty]$ condition with respect to a is determined by its a-component, it follows that $f_1(A, B, C)$ and $f_2(A, B, C)$ can generate every α -function of A, B, C, ... I that satisfies the $[:a_\infty]$ condition with respect to a. By mere interchange of arguments every α -function of A, B, C, ... I that satisfies the $[:a_\infty]$ condition can then be obtained. Hence $f_1(A, B, C)$ and $f_2(A, B, C)$, which are α -functions satisfying the $[:a_\infty]$ condition, generate the closed system

F_1^∞ . All α -functions satisfying the $[:a_\infty]$ condition.

We can therefore conclude that every closed $[\alpha]$ system that satisfies the $[:aa]$ condition but does not satisfy the $[A:a]$ condition contains F_1^∞ .

2. $[:aa] [\alpha]$ systems satisfying the $[A:a]$ condition.

This is the case that includes the $[:aa] [\alpha]$ systems of self-dual functions O_1 , R_1 , and D_2 . Every other closed system of the type under consideration will then not consist wholly of self-dual functions, and hence will possess the function $A+B:ab$. If the first expression of the normal $[A:a]$ expansion of each function in such a system consists of terms of one letter, the system must be a logical sum system, and hence can only be S_1 or S_2 . Otherwise some function in the system will have a normal expansion with a term of at least two letters in its first expression. By the definition of a normal expansion, every other term of this first expression has some letter not in the term in question. Now in this function replace each argument that is not involved in that term by A, one of the remaining arguments by B, the rest by C. The term in question becomes BC, while

every other term of the first expression of the resulting expansion will involve the capital letter A. If then we let the function thus obtained replace the variable B in the function $A+B:ab$, the resulting function $f_1(A,B,C)$ will have $A+BC$ for the first expression of its normal expansion, and hence $A+BC:ab+ac$ for its normal expansion.

The system under discussion therefore possesses this function $f_1(A,B,C)$. Since it has the function $A+B:ab$, it will also possess the function $f_2(A,B,C)$ with expansion $A+B+C:abc$. $f_1(A,B,C)$ has the a-component $BC:b+c$, $f_2(A,B,C)$ the a-component $B+C:bc$. These a-components therefore generate the closed system A_4 consisting of all $[A:a]$ α -functions. Now from their normal expansions we see that the a-component of every $[A:a]$ α -function of A,B,C,\dots,I , which satisfies the $[:a_\infty]$ condition with respect to a and which is not reducible to $\alpha_1(A)$ is an $[A:a]$ α -function of B,C,\dots,I . $f_1(A,B,A)$, which admits the expansion $A:a$, can generate every function of A,B,C,\dots,I reducible to $\alpha_1(A)$. It then easily follows, as in the derivation of F_1^∞ , that $f_1(A,B,C)$ and $f_2(A,B,C)$, which are $[A:a]$ α -functions satisfying the $[:a_\infty]$ condition with respect to a, generate the closed system

F_2^∞ . All $[A:a]$ α -functions satisfying the $[:a_\infty]$ condition.

We can therefore state that, with the exception of O_1 , R_1 , D_2 , S_1 , and S_2 , every closed $[\alpha]$ system that satisfies the $[:aa]$ condition and also satisfies the $[A:a]$ condition contains F_2^∞ .

3. $[:aa]$ $[\alpha, \beta]$ systems satisfying the $[A:a]$ condition.

The β -functions to be found in such a system must be reducible to first order, and hence constitute either the closed $[\beta]$ system O_2 or the closed $[\beta]$ system R_2 . In the first case, the α -functions in the system must be of the first order, and hence the system can only be O_5 . In every other case the system consists of the functions in R_2 and the functions in some closed $[\alpha]$ system that satisfies both the $[:aa]$ and $[A:a]$

condition. We therefore merely have to see for each $[\alpha]$ system Σ of this type whether $\Sigma + R_2$ is closed, i.e., whether the functions in Σ together with the functions in R_2 constitute a closed system.

If Σ consists wholly of self-dual functions, it can only be O_1 , R_1 , or D_2 . We have $O_1 + R_2 = R_5$, $R_1 + R_2 = R_6$. On the other hand $D_2 + R_2$ is not closed. For if in $AB+BC+AC:ab+bc+ac$, which generates D_2 , we replace C by $\beta_1(A)$, we obtain $A+B:ab$ which is an α -function not in D_2 . If Σ does not consist wholly of self-dual functions, it either is S_1 or S_2 , or else contains F_2^∞ . We have $S_1 + R_2 = S_3$, $S_2 + R_2 = S_4$. Furthermore $F_2^\infty + R_2$ is easily seen to be the closed system

F_3^∞ . All $[A:a]$ functions satisfying the $[:a_\infty]$ condition.

We therefore have the following result. Every closed $[\alpha,\beta]$ system that satisfies the $[:aa]$ condition and also satisfies the $[A:a]$ condition, if not O_5 , R_5 , R_6 , S_3 , or S_4 , contains F_3^∞ .

4. $[:aa]$ $[\alpha,\beta]$ systems not satisfying the $[A:a]$ condition.

By the methods of §19 and §20 we see that every closed system of this type possesses a function $f_1(A,B)$ with ab in the first, aB in the second expression of its complete expansion. Both AB and Ab must be in the first expression of this complete expansion, since neither of these terms has a small letter in common with the term aB . Hence $f_1(A,B)$ has the complete expansion $AB+Ab+ab:aB$, which simplifies to $A+b:aB$.

By replacing $B:b$ in $A+b:aB$ by $A+b:aB$ we obtain the function $A+B:ab$.⁴⁵ Hence the closed system under consideration possesses the function $f_2(A,B,C)$ with expansion $A+B+C:abc$. Now $f_1(A,B)$ has the a -component $b:B$, $f_2(A,B,C)$ the a -component $B+C:bc$. These a -components generate the complete

⁴⁵ Cf. Hilbert-Bernays [9], p. 51.

system C_1 . Hence $f_1(A,B)$ and $f_2(A,B,C)$ can generate a function $f(A,B,C,\dots,I)$ with arbitrary a -component $\phi(B,C,\dots,I)$. As $f_1(A,B)$ and $f_2(A,B,C)$ satisfy the $[:a_\infty]$ condition with respect to a , it follows, as in the derivation of F_1^∞ , that they, and hence $f_1(A,B)$ alone, generate the closed system

F_4^∞ . All functions satisfying the $[:a_\infty]$ condition.

Hence every closed $[\alpha,\beta]$ system that satisfies the $[:aa]$ condition but does not satisfy the $[A:a]$ condition contains F_4^∞ .

Every closed system satisfying the $[:a_\infty]$ condition also satisfies the $[:aa]$ condition, and hence comes under one of the four cases under which all closed $[:aa]$ systems have been classified. From the definition of F_i^∞ , $i = 1, 2, 3, 4$, we see that every closed $[:a_\infty]$ system coming under case i is contained in F_i^∞ . As all of the closed systems specifically mentioned in the above derivations, with the exception of D_2 , satisfy the $[:a_\infty]$ condition, it follows as a partial consequence of the above results that the only closed systems satisfying the $[:a]$ condition are

under case 1, F_1^∞ ,
 under case 2, $O_1, R_1, S_1, S_2, F_2^\infty$,
 under case 3, $O_5, R_5, R_6, S_3, S_4, F_3^\infty$,
 under case 4, F_4^∞ .

The dual of the $[:a_\infty]$ condition, which we call the $[A_\infty:]$ condition, is the following. All the terms of the first expression of an expansion of the function have a capital letter in common.

The duals of the four new closed systems found in the present section are then the following.

F_5^∞ . All α -functions satisfying the $[A_\infty:]$ condition.
 F_6^∞ . All $[A:a]$ α -functions satisfying the $[A_\infty:]$ con-

dition.

F_7^∞ . All $[A:a]$ functions satisfying the $[A_\infty:]$ condition.

F_8^∞ . All functions satisfying the $[A_\infty:]$ condition.

It is unnecessary to state the duals of the general results found above. We but add that F_5^∞ , O_1 , R_1 , P_1 , P_2 , F_6^∞ are the only closed $[\alpha]$ systems satisfying the $[A_\infty:]$ condition, while O_6 , R_7 , R_8 , P_3 , P_4 , F_7^∞ , F_8^∞ are the only closed $[\alpha,\gamma]$ systems satisfying the $[A_\infty:]$ condition.

That the eight systems F_i^∞ , $i = 1, 2, \dots, 8$ are distinct is very readily verified. We may note that F_4^∞ contains both F_1^∞ and F_3^∞ each of which contains F_2^∞ . Dually for the second set of systems. On the other hand, the common part of any F_i^∞ , $i = 1, \dots, 4$, and any F_j^∞ , $j = 5, \dots, 8$, would be a closed system satisfying both the $[:aa]$ and the $[AA:]$ condition, and hence could only be O_1 , R_1 , or D_2 . As D_2 does not satisfy either the $[:a_\infty]$ or the $[A_\infty:]$ condition, while R_1 , which itself contains O_1 , is contained in each of the above eight systems, this common part is seen to be R_1 , i.e., the system of all α -functions reducible to first order.

Finally we consider the orders of these eight systems. As F_4^∞ was generated by a second order function, it, and hence also F_8^∞ , is of the second order. On the other hand, the only α -functions of A , and of A,B satisfying the $[:aa]$ condition are $A:a$, $AB+Ab:ab+aB$, $AB+aB:ab+Ab$, $AB+Ab+aB:ab$, which are all in S_2 . Since S_2 is contained in F_2^∞ , and yet is distinct from F_2^∞ , it follows that F_1^∞ , F_2^∞ , F_3^∞ , and hence also F_5^∞ , F_6^∞ , F_7^∞ cannot be generated by first and second order functions. Since they have been shown to be generated by third and lower order functions, it follows that they are of the third order.

§23

 $[\alpha]$, $[\alpha, \beta]$, AND $[\alpha, \gamma]$ SYSTEMS; CONCLUDING DISCUSSION:

THE EIGHT INFINITE FAMILIES OF SYSTEMS

Having thus determined all closed $[:aa]$ systems that satisfy the $[:a_\infty]$ condition, and all closed $[AA:]$ systems that satisfy the $[A_\infty:]$ condition, we turn now to closed $[:aa]$ systems that do not satisfy the $[:a_\infty]$ condition, and closed $[AA:]$ systems that do not satisfy the $[A_\infty:]$ condition. Again our explicit discussion can be restricted to closed $[:aa]$ systems. We then observe that the general results of the preceding section, which had for a partial application the determination of all closed $[:a_\infty]$ systems, have for the remainder of their content the following information about a closed $[:aa]$ system that does not satisfy the $[:a_\infty]$ condition.

1. If it is an $[\alpha]$ system that does not satisfy the $[A:a]$ condition, it contains F_1^∞ , the system of all $[:a_\infty]$ α -functions.
2. If it is an $[\alpha]$ system other than D_2 that satisfies the $[A:a]$ condition, it contains F_2^∞ , the system of all $[:a_\infty]$ $[A:a]$ α -functions.
3. If it is an $[\alpha, \beta]$ system that satisfies the $[A:a]$ condition, it contains F_3^∞ , the system of all $[:a_\infty]$ $[A:a]$ functions.
4. If it is an $[\alpha, \beta]$ system that does not satisfy the $[A:a]$ condition, it contains F_3^∞ , the system of all $[:a_\infty]$ functions.

This fourfold result, which corresponds to the fourfold classification of closed $[:aa]$ systems when that classification is restricted to systems not satisfying the $[:a_\infty]$ condition, is fundamental in the following development. Note that the $[:a_\infty]$ system contained in a given system as specified by these results is the largest $[:a_\infty]$ system that any system coming under the same case as the given system could contain, and hence is the largest $[:a_\infty]$ system contained in the given system.

Just as the preceding section required that specialization of the $[:aa]$ condition which we termed the $[:a_\infty]$ condition, so we now require further specializations of the $[:aa]$ condition.

A function will be said to satisfy the $[:a_m]$ condition, $m \geq 2$, if it admits an expansion with the following property. Any m terms, distinct or not, of the second expression of the expansion have a small letter in common. Note that when $m = 2$ this condition is the previously symbolized $[:aa]$ condition. Again as in §8, it is readily verified first, that the $[:a_m]$ condition is independent of the particular expansion used; second, that the $[:a_m]$ condition is iterative.

We can therefore refer to a closed system satisfying the $[:a_m]$ condition. From our definition we see that if a function or closed system satisfies the $[:a_m]$ condition, it also satisfies the $[:a_{m'}]$ condition for every m' with $m \geq m' \geq 2$.

If a function satisfying the $[:a_m]$ condition admits an expansion whose second expression has not more than m terms, those terms will all have a small letter in common, and hence the function also satisfies the $[:a_\infty]$ condition. If then a closed system satisfies the $[:a_m]$ condition for every finite m , with $m \geq 2$, every function in the system, and hence the system itself, must satisfy the $[:a_\infty]$ condition.

It follows that with every closed system that satisfies the $[:aa]$ condition but not the $[:a_\infty]$ condition there is associated a finite integer μ , $\mu \geq 2$, such that the system satisfies the $[:a_\mu]$ condition, but not the $[:a_{\mu+1}]$ condition.

We now prove that every closed system that satisfies the $[:a_\mu]$ condition, but not the $[:a_{\mu+1}]$ condition, possesses the function $h(A, B, \dots, E)$ of order $\mu+1$ with the following complete expansion. The first expression has all complete terms on A, B, \dots, E that

involve more than one capital letter, the second all complete terms on A, B, \dots, E with one or no capital letter.

This result is immediate for the exceptional system D_2 , the only closed system of self-dual functions that satisfies the $[:aa]$, but not the $[:a_\infty]$ condition. For μ is then 2, and the above function is the $h(A, B, C)$ of §20 which generates D_2 . In every other case we can therefore use the fourfold result stated above. The proof for all systems other than D_2 will then take the following form. The system in question must possess a function $f(A, B, \dots, I)$ that satisfies the $[:a_\mu]$ but not the $[:a_{\mu+1}]$ condition. Corresponding to A, B, \dots, I , the arguments of this function, we define functions M, N, \dots, W , each of which is a function of the $\mu+1$ variables A, B, \dots, E , such that $f(M, N, \dots, W)$, considered as a function of A, B, \dots, E , is the desired function $h(A, B, \dots, E)$. By proving M, N, \dots, W to be in the closed system under discussion, we complete the proof that $f(M, N, \dots, W)$, i.e., $h(A, B, \dots, E)$ is in the system.

$f(A, B, \dots, I)$ will have $\mu+1$ terms $t_1, t_2, \dots, t_{\mu+1}$ in the second expression of its complete expansion such that any μ of them have a small letter in common, but all do not. With these $\mu+1$ terms we set in 1-1 correspondence the $\mu+1$ complete terms $\tau_1, \tau_2, \dots, \tau_{\mu+1}$, on the $\mu+1$ letters A, B, \dots, E , which involve but one capital letter. When $f(A, B, \dots, I)$ is an α -function, our definition of the functions M, N, \dots, W that are to replace A, B, \dots, I respectively then takes the following form. Every complete term on A, B, \dots, E with more than one capital is to be in the first expression of the complete expansion of each function M, N, \dots, W , the term $ab\dots e$ with no capital is to be in the second expression of each complete expansion; on the other hand, a term τ_1 with but one capital is to be in the first or second expression of the complete expansion of the function M, N, \dots, W according as the variable A, B, \dots, I to be replaced by that function is capital or small in the term t_1 . The complete terms on A, B, \dots, E with more than one capital will then be contained in $MN\dots W$, and hence will be in the first expression of the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, E . The term $ab\dots e$ will be contained in $MN\dots W$, and hence will

be in the second expression of that complete expansion. Furthermore, every term τ_1 will be contained in that complete term on M, N, \dots, W which arises from t_1 when A, B, \dots, I is replaced by M, N, \dots, W respectively. As each t_1 is in the second expression of the complete expansion of $f(A, B, \dots, I)$, each τ_1 will be in the second expression of the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, E . If then we set $f(M, N, \dots, W) = h(A, B, \dots, E)$, $h(A, B, \dots, E)$ so defined has the desired complete expansion.

To complete our proof when $f(A, B, \dots, I)$ is an α -function, we need only show that M, N, \dots, W are in the system under discussion. Now none of the variables A, B, \dots, I can be small in each of the $\mu+1$ terms $t_1, t_2, \dots, t_{\mu+1}$, as those terms do not have a small letter in common. Hence none of the functions M, N, \dots, W can have all of the terms $\tau_1, \tau_2, \dots, \tau_{\mu+1}$ in the second expression of its complete expansion. Since each variable A, B, \dots, E is capital in but one of the terms $\tau_1, \tau_2, \dots, \tau_{\mu+1}$, any $v < \mu+1$ of those terms do have a small letter in common. It follows that M, N, \dots, W all satisfy the $[:a_\infty]$ condition. Clearly each of these functions is an α -function. Since no complete term on A, B, \dots, E other than τ_1 and $ab\dots e$ can have every letter small that is small in τ_1 , each of these functions is also seen to satisfy the $[A:a]$ condition. That is, M, N, \dots, W are $[:a_\infty][A:a]\alpha$ -functions, and hence belong to each of the systems $F_1^\infty, F_2^\infty, F_3^\infty, F_4^\infty$. As the system under consideration contains at least one of these systems, it possesses each of the functions M, N, \dots, W , and hence also the function $h(A, B, \dots, E)$.

This proof requires but a slight modification when $f(A, B, \dots, I)$ is a β -function. We can no longer place the term $ab\dots e$ in the second expression of the complete expansion of each function M, N, \dots, W , for, being contained in $MN\dots W$, $ab\dots e$ would then be in the first expression of the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, E . We therefore choose any term t_0 in the second expression of the complete expansion of $f(A, B, \dots, I)$, and place $ab\dots e$ in the first or second expression of the complete expansion of M, N, \dots, W according as the variable A, B, \dots, I to be replaced by that function is capital or small in t_0 . The term $ab\dots e$ will then be contained in that complete term on M, N, \dots, W which arises from t_0 when

A, B, \dots, I is replaced by M, N, \dots, W respectively, and hence will be in the second expression of the complete expansion of $f(M, N, \dots, W)$ in terms of A, B, \dots, E . By otherwise defining M, N, \dots, W as when $f(A, B, \dots, I)$ is an α -function, $f(M, N, \dots, W)$ again has the desired complete expansion in terms of A, B, \dots, E , and hence defines the function $h(A, B, \dots, E)$. While M, N, \dots, W will no longer all be $[A:a]$ α -functions, they still satisfy the $[:a_\infty]$ condition, and hence belong to F_4^∞ . As $f(A, B, \dots, I)$ is now a β -function not reducible to first order, the given system must come under case 4, and hence contains F_4^∞ . It therefore again possesses each of the functions M, N, \dots, W , and hence also the function $h(A, B, \dots, E)$.

The second expression of the complete expansion of $h(A, B, \dots, E)$ is

$$Ab \dots e + aB \dots e + \dots + ab \dots E + ab \dots e.$$

As in our previous discussion it easily follows that $h(A, B, \dots, E)$ is an $[A:a]$ α -function that satisfies the $[:a_\mu]$, but not the $[:a_{\mu+1}]$ condition. That it satisfies the $[A:a]$ condition can also be seen from the fact that the above second expression can be simplified to

$$bc \dots e + ac \dots e + \dots + ab \dots d,$$

i.e., the logical sum of the logical products of the $\mu+1$ small letters a, b, \dots, e taken μ at a time. This second expression of the normal expansion of $h(A, B, \dots, E)$ is the most useful in the following derivations.

Note and recall that every closed system which satisfies the $[:aa]$, but not the $[:a_\infty]$, condition comes under one and only one of the four cases under which all closed $[:aa]$ systems have been classified, and has associated with it a unique integer μ , $\mu \geq 2$, such that the system satisfies the $[:a_\mu]$, but not the $[:a_{\mu+1}]$, condition. We now prove that, with the exception of D_2 , there exists one and only one closed system for each case i , $i = 1, 2, 3, 4$, and associated integer μ , $\mu \geq 2$. This closed system will be designated F_i^μ . That there is at least one such system can be verified immediately. In fact define F_i^μ as follows.

$\mu \geq 2$:

- F_1^μ . All α -functions satisfying the $[:a_\mu]$ condition.
- F_2^μ . All $[A:a]$ α -functions satisfying the $[:a_\mu]$ condition.
- F_3^μ . All $[A:a]$ functions satisfying the $[:a_\mu]$ condition.
- F_4^μ . All functions satisfying the $[:a_\mu]$ condition.

We then first observe that each system F_i^μ is closed, since it is defined by means of iterative conditions. From its definition we see that F_1^μ possesses the function $h(A,B,\dots,E)$ of order $\mu+1$ which fails to satisfy the $[:a_{\mu+1}]$ condition. Hence F_1^μ , while satisfying the $[:a_\mu]$ condition, does not satisfy the $[:a_{\mu+1}]$ condition, and consequently has μ for its associated integer. Finally, F_1^μ belongs to case i since F_1^∞ is evidently the largest $[:a_\infty]$ system that it contains.

We now prove that F_1^μ is the only closed system belonging to case i, and associated number μ , with the sole exception $i = 2$, $\mu = 2$, for which there is also the closed system D_2 . To do so we examine each case in turn.

i. A closed system with associated number μ that comes under case i can now be described as a closed $[\alpha]$ system that satisfies the $[:a_\mu]$ condition, contains F_1^∞ , and possesses the function $h(A,B,\dots,E)$ of order $\mu+1$. Let $f(A,B,\dots,I)$ be any α -function of A,B,\dots,I that satisfies the $[:a_\mu]$ condition, and let n be the number of terms, apart from $ab\dots i$, that are in the second expression of the complete expansion of $f(A,B,\dots,I)$. When $n \leq \mu$, these n terms have a small letter in common as a consequence of $f(A,B,\dots,I)$ satisfying the $[:a_\mu]$ condition, and hence $f(A,B,\dots,I)$ satisfies the $[:a_\infty]$ condition. That is, when $n \leq \mu$, $f(A,B,\dots,I)$ is in F_1^∞ , and hence in the given system. We now prove inductively that when $n > \mu$, $f(A,B,\dots,I)$ will also be in the system. Assume that all $[:a_\mu]$ α -functions of A,B,\dots,I with $n = m$, $m \geq \mu$ are in the system, and let $f(A,B,\dots,I)$ have $n = m+1$. As $m+1 \geq \mu+1$, we can choose $\mu+1$ distinct terms other than $ab\dots i$ from the

second expression of the complete expansion of $f(A, B, \dots, I)$. Now form the $\mu+1$ expressions obtainable from this second expression by arbitrarily omitting one of those $\mu+1$ terms. By considering those $\mu+1$ expressions to be the second expressions of complete expansions on A, B, \dots, I , they serve to define $\mu+1$ functions M, N, \dots, Q of A, B, \dots, I . These functions are evidently $[:a_\mu] \alpha$ -functions of A, B, \dots, I with $m = n$. Hence, by our assumption, they are in the given system, and consequently $h(M, N, \dots, Q)$ is in the system. Now the second expression of the complete expansion of $h(M, N, \dots, Q)$ in terms of A, B, \dots, I consists of those terms which are in all, or all but one, of the second expressions of the complete expansions of M, N, \dots, Q . As every term in the second expression of the complete expansion of $f(A, B, \dots, I)$ is at most missing from one of the second expressions of the complete expansions of M, N, \dots, Q , while the latter second expressions have no terms that are not in the former, it follows that $h(M, N, \dots, Q)$ and $f(A, B, \dots, I)$ have complete expansions with identical second expressions, and hence are themselves identical. Since $h(M, N, \dots, Q)$ is in the given system, we thus see that $f(A, B, \dots, I)$ is in the system. That is, if all $[:a_\mu] \alpha$ -functions of A, B, \dots, I with $n = m$, $m \geq \mu$, are in the given system, the same is true for $n = m+1$. As this hypothesis was verified for $m = \mu$, our result follows. The given system therefore possesses all $[:a_\mu] \alpha$ -functions, and being an $[:a_\mu] [\alpha]$ system, can only be F_1^μ .

2. The fact that a closed system coming under case 2 satisfies the $[A:a]$ condition suggests using normal, instead of complete, expansions. If this be done, and exception is made of D_2 , the above derivation can be used here with but few modifications. The induction is now carried through with respect to the number of terms n in the second expression of the normal expansion of $f(A, B, \dots, I)$. F_2^∞ takes the place of F_1^∞ as the starting point of this induction. Finally, in obtaining the second expression of the normal expansion of $h(M, N, \dots, Q)$ in terms of A, B, \dots, I we first obtain an expression consisting of all the terms in the second expression of the normal expansion of $f(A, B, \dots, I)$, and in addition other terms contained in those terms. As these other terms can then be omitted, it follows

that $h(M, N, \dots, Q)$ and $f(A, B, \dots, I)$ have normal expansions with identical second expressions. We thus see that the system in question must be F_2^μ .

3. As a closed system coming under case 3 is an $[\alpha, \beta]$ system satisfying the $[A:a]$ condition, its β -functions must be reducible to first order. Since the system contains F_3^∞ which itself contains the $[\alpha]$ system F_2^∞ , the α -functions in the system must constitute a closed system, other than D_2 , which comes under case 2 and has the same associated integer μ as in the given system. This $[\alpha]$ system being then F_2^μ , the given system consists of the functions in F_2^μ and all β -functions reducible to first order, and hence is F_3^μ .

4. The argument of case 1 carries over completely with F_4^∞ in place of F_1^∞ , and n as the number of terms in the second expression of the complete expansion of $f(A, B, \dots, I)$, to show that the system must be F_4^μ .

We have therefore completely proved that except for D_2 , the only closed systems satisfying the $[:aa]$, but not the $[:a_\infty]$, condition are F_i^μ , $i = 1, 2, 3, 4$, $\mu = 2, 3, 4, \dots$. Since F_i^μ belongs to case i , and has μ for its associated integer, it follows that F_i^μ and F_j^ν are distinct unless $i = j$ and $\mu = \nu$. The systems F_i^μ thus constitute four distinct infinite families of systems, one family for each case i . Since the $[:a_\mu]$ condition is implied by the $[:a_{\mu+1}]$ condition, F_i^μ contains $F_i^{\mu+1}$, so that if the systems in each family be arranged in order of increasing μ , they form an infinite sequence with each system containing the following. As, also, a function which satisfies the $[:a_\mu]$ condition for every μ , $\mu \geq 2$, satisfies the $[:a_\infty]$ condition, and conversely, we see that the common part of all the systems in the i -th family is F_i^∞ .

Finally, we prove that each system F_i^μ is of order $\mu+1$. First, F_i^μ can be generated by the $(\mu+1)$ -th order function $h(A, B, \dots, E)$ and a set of generators of the system F_i^∞ . As the order of F_i^∞ does not exceed three, the order of F_i^μ is not greater than $\mu+1$. Now consider any set of generators of F_i^μ . This set will have to possess a function $f(A, B, \dots, I)$ that sat-

isfies the $[:a_\mu]$ but not the $[:a_{\mu+1}]$ condition. Let $t_1, t_2, \dots, t_{\mu+1}$ be a corresponding set of $\mu+1$ terms in the second expression of the complete expansion of $f(A, B, \dots, I)$ such that all do not have a small letter in common, but any μ of them do. For each such set of μ terms pick out one letter which is small in each term in that set. No two sets can yield the same letter, or that letter would be small in all $\mu+1$ terms. As there are $\mu+1$ such sets of μ terms, $f(A, B, \dots, I)$ must have at least $\mu+1$ arguments. That is, a set of generators of F_i^μ is at least of order $\mu+1$. The order of F_i^μ therefore is $\mu+1$. This result has the added interest of showing that closed systems exist of every finite order.

The dual of the $[:a_m]$ condition may be symbolized $[A_m:]$ and is to the effect that any m terms, distinct or not, of the first expression of an expansion of a function have a capital letter in common.

We can then state immediately that the only closed systems other than D_2 that satisfy the $[AA:]$, but not the $[A_\infty:]$ condition are the following.

$\mu \geq 2$:

- F_5^μ . All α -functions satisfying the $[A_\mu:]$ condition.
- F_6^μ . All $[A:a]$ α -functions satisfying the $[A_\mu:]$ condition.
- F_7^μ . All $[A:a]$ functions satisfying the $[A_\mu:]$ condition.
- F_8^μ . All functions satisfying the $[A_\mu:]$ condition.

We also have immediately that each of these systems with indicated μ is of order $\mu+1$, and that all the differently symbolized systems thus represented are distinct. They, too, therefore, form four infinite families of systems with the same abstract structure as the first set of families.

That the systems in this second set of four infinite families are distinct from the systems in the first set of four infinite families will follow from the discussion of the common part of two systems that belong one to an infinite family in

the first set, the other to an infinite family in the second set. To obtain the most complete result we shall consider F_i^∞ to be a member of the i -th family with $\mu = \infty$. Then, as in the preceding section, we see that the common part in question can only be O_1 , R_1 , or D_2 , of which the choice narrows down to R_1 and D_2 . Now D_2 is a closed $[A:a]$ $[\alpha]$ system that satisfies both the $[:a_2:]$ and $[AA:]$ condition, i.e., both the $[:a_2:]$ and $[A_2:]$ condition. It is therefore contained in each F_i^2 , and hence is the common part of any F_i^2 and F_j^2 that belong to families in different sets. On the other hand, D_2 cannot be contained in any F_i^μ for which $\mu > 2$, since it possesses the function $AB+BC+AC:ab+bc+ac$ which does not satisfy either the $[:a_3:]$ or $[A_3:]$ condition. Hence the common part of any F_i^μ and F_j^ν that belong to families in different sets is but R_1 , the system of all α -functions reducible to first order, when either μ or ν is greater than 2.

This concludes our derivation of all closed systems. As a consequence of the corresponding individual results we can state that every closed system is of finite order, i.e., can be generated by a finite set of generators. In the preceding sections only closed systems of the first three orders were disclosed. The present section, while adding to the third order systems previously discovered, enables us to state that there are exactly eight closed systems of every finite order greater than three.

§24

SUMMARY

For convenience of reference we here regroup the various closed systems according to their order, and also index them according to the section in which their definition occurs.

First Order — 9 Systems

$O_1, \dots, O_9.$

Second Order — 37 Systems

R_1, \dots, R_{13} , S_1, \dots, S_6 , P_1, \dots, P_6 ,

L_1, L_2, L_3 , F_4^∞, F_8^∞ , A_1, A_2, A_3, A_4 , C_1, C_2, C_3 .

Third Order — 20 Systems

D_4, D_5 , D_1, D_2, D_3 , $F_1^\infty, F_2^\infty, F_3^\infty, F_5^\infty, F_6^\infty, F_7^\infty, F_1^2, \dots, F_8^2$, C_4 .

n-th Order, $n > 3$, — 8 Systems

$F_1^{n-1}, \dots, F_8^{n-1}$

• •

$O_1, \dots, O_9 : \$10$

$R_1, \dots, R_{13} : \$11$

$S_1, \dots, S_6 : \$13$

$P_1, \dots, P_6 : \$13$

$L_1, \dots, L_5 : \$15$

$F_1^\infty, \dots, F_8^\infty : \22

$\mu = 2, 3, \dots, F_1^\mu, \dots, F_8^\mu : \23

$D_1, D_2 : \$20$, $A_1 : \$14$, $C_1 : \$17$

$D_3 : \$21$. $A_2, A_3 : \$19$, $C_2, C_3 : \$19$

$A_4 : \$18$, $C_4 : \$18$

PART III
CO-ORDINATION AND APPLICATION

§25

THE INCLUSION RELATION⁴⁶

The above determination of all closed systems, i.e., of all iteratively closed membership systems over ν' , acquires an added significance when these systems are co-ordinated through the relation of inclusion between systems. Note that system T_1 contains, or as we shall say includes system T_0 if each function in T_0 is also in T_1 . If, in addition, T_0 is distinct from T_1 , then T_1 properly includes T_0 . When the relation of inclusion is restricted to the set of all closed systems in the above sense, it leads to the significant relation of immediate inclusion. A closed system T_1 immediately includes a closed system T_0 if T_1 properly includes T_0 , and there is no closed system T such that T_1 properly includes T and T properly includes T_0 .

The set of all closed systems, which may also be described as the set of all closed systems included in the complete system C_1 , is an infinite set. Consider, on the other hand, a closed system which includes but a finite number of closed systems. Then for such a finite set of closed systems the usually redundant relation of inclusion is completely analysable in terms of the non-redundant relation of immediate inclusion. In this connection we define a finite inclusion chain joining a closed system T_1 to a closed system T_0 as a finite sequence of closed systems $T_1, T_2, \dots, T_n, T_0$ such that each system in the sequence,

⁴⁶ Cf. G. Birkhoff [5]. As pointed out in the introduction, even this version of the present section, including footnotes 47 and 48, was written before the recent work on lattices.

except the last, immediately includes the next. It is then easily proved, and in a purely abstract manner, that if T_0 and T_1 are any two closed systems in a finite set of closed systems of the above kind, then, if T_1 properly includes T_0 , there is at least one finite inclusion chain joining T_1 to T_0 .⁴⁷

The same result holds for two closed systems T_0 , T_1 in our infinite set of all closed systems provided there are but a finite number of closed systems T such that T_1 properly includes T and T properly includes T_0 . If, on the other hand, there are an infinite number of such closed systems T 'between' T_0 and T_1 , this result need no longer be valid. To state our positive result for this case we extend both the notion of inclusion chain and immediate inclusion.

We first define an open inclusion chain to be an infinite sequence of closed systems T_1, T_2, T_3, \dots such that each system in the sequence immediately includes the next. An open inclusion chain is then said to immediately include a closed system T' if each system in the open chain includes T' , and there is no closed system T properly including T' which also is included in each system in the open chain. Finally, a well ordered series, of closed systems, of the form $T_1, T_2, T_3, \dots T', T'', \dots T^{(n)}, T_0$ is called a simple transfinite inclusion chain joining T_1 to T_0 if each system in the series, except the last, immediately includes the next, and the resulting open inclusion chain T_1, T_2, T_3, \dots immediately includes T'' .

It can then be proved by a direct consideration of the eight infinite families of closed systems that if a closed system T_1 properly includes a closed system T_0 , and there are an infinite number of closed systems between T_0 and T_1 , then there exists a simple transfinite inclusion chain joining T_1 to T_0 . We thus have the following general result.

⁴⁷ Since, in this finite case, the number of systems in such a joining chain must be bounded, we may define the interval from T_1 to T_0 to be the largest value of n , i.e., the largest number of 'links' in a chain joining T_1 to T_0 . Symbolizing this interval by $I(T_1, T_0)$, we observe that if T_1 properly includes T , and T properly includes T_0 , then

$$I(T_1, T_0) > I(T_1, T) + I(T, T_0),$$

which is reminiscent of the relation between timelike intervals in the special relativity theory!

If T_0 and T_1 are any two closed systems such that T_1 properly includes T_0 , then there exists either a finite or simple transfinite inclusion chain joining T_1 to T_0 .⁴⁸

Since the relation of proper inclusion is transitive, we immediately have the converse of the above result. It follows that for closed systems the relation of proper inclusion, and hence of inclusion, will be completely determined if the following two problems be solved.

(a). To determine for every closed system the closed systems immediately included therein.

(b). To determine for every open inclusion chain the closed systems immediately included therein.

In this connection it is of fundamental importance that, as was pointed out in Part II, all of the differently symbolized closed systems are in fact distinct.

The solution of problem (a) is at worst tedious. We merely give the discussion for the closed systems immediately included in the complete system C_1 , as this result is required in the next section. By §10, every closed system associated with a given closed first order system is included in the maximal closed system associated with that first order system. There are eight such maximal closed systems apart from C_1 itself, namely C_4 , R_2 , R_3 , D_3 , C_2 , C_3 , R_9 , A_1 , associated with O_1, O_2, \dots, O_8 , respectively. Of these, R_2 , R_3 , R_9 are included in A_1 , C_4 in C_2 . Hence, every closed system which is not an $[\alpha, \beta, \gamma, \delta]$ system is included in at least one of the four closed systems D_3 , C_2 , C_3 , A_1 . On the other hand, every closed $[\alpha, \beta, \gamma, \delta]$ system other than C_1 is an alternating system, and hence is included

⁴⁸ More generally any series of systems with T_1 and T_0 as first and last elements may be called an inclusion chain joining T_1 to T_0 if each system in the series includes all those that follow it, while no system may be inserted in the series without destroying this property. It can then be proved for our infinite set of systems that every inclusion chain is either a finite or simple transfinite inclusion chain.

in the complete alternating system L_1 .

We thus have that every closed system properly included in C_1 is included in at least one of the five closed systems D_3, C_2, C_3, A_1, L_1 . But it can be directly verified that no one of these five systems is properly included in another. It follows that they constitute all of the closed systems immediately included in C_1 .

The solution of problem (b) is easily obtained from our knowledge of the eight families of closed systems. It is readily seen that F_i^μ immediately includes $F_i^{\mu+1}$ for every $\mu \geq 2$, $i = 1, 2, \dots, 8$. Hence the eight infinite families give rise to eight open inclusion chains $F_i^2, F_i^3, F_i^4, \dots, i = 1, 2, \dots, 8$. Since F_i^∞ is the logical product of all the systems in the i -th family, it follows that F_i^∞ is immediately included in the open inclusion chain corresponding to the i -th family, and, in fact, is the only closed system immediately included therein. Now consider any open inclusion chain. As there are but a finite number of closed systems outside of the infinite families, it follows that from a certain system onward all of the systems in the chain belong to the infinite families. Furthermore, it is readily seen that as we pass down the chain, the systems can change from one family to another in only the following ways: from the fourth family to the first or third, from the first or third to the second; and dually for the other four families. Consequently, along the given chain the systems can change from one family to another at most twice, so that from a certain system in the chain onward all of the systems belong to the same family. If this family is the i -th, F_i^∞ is the only closed system immediately included in the given open inclusion chain.

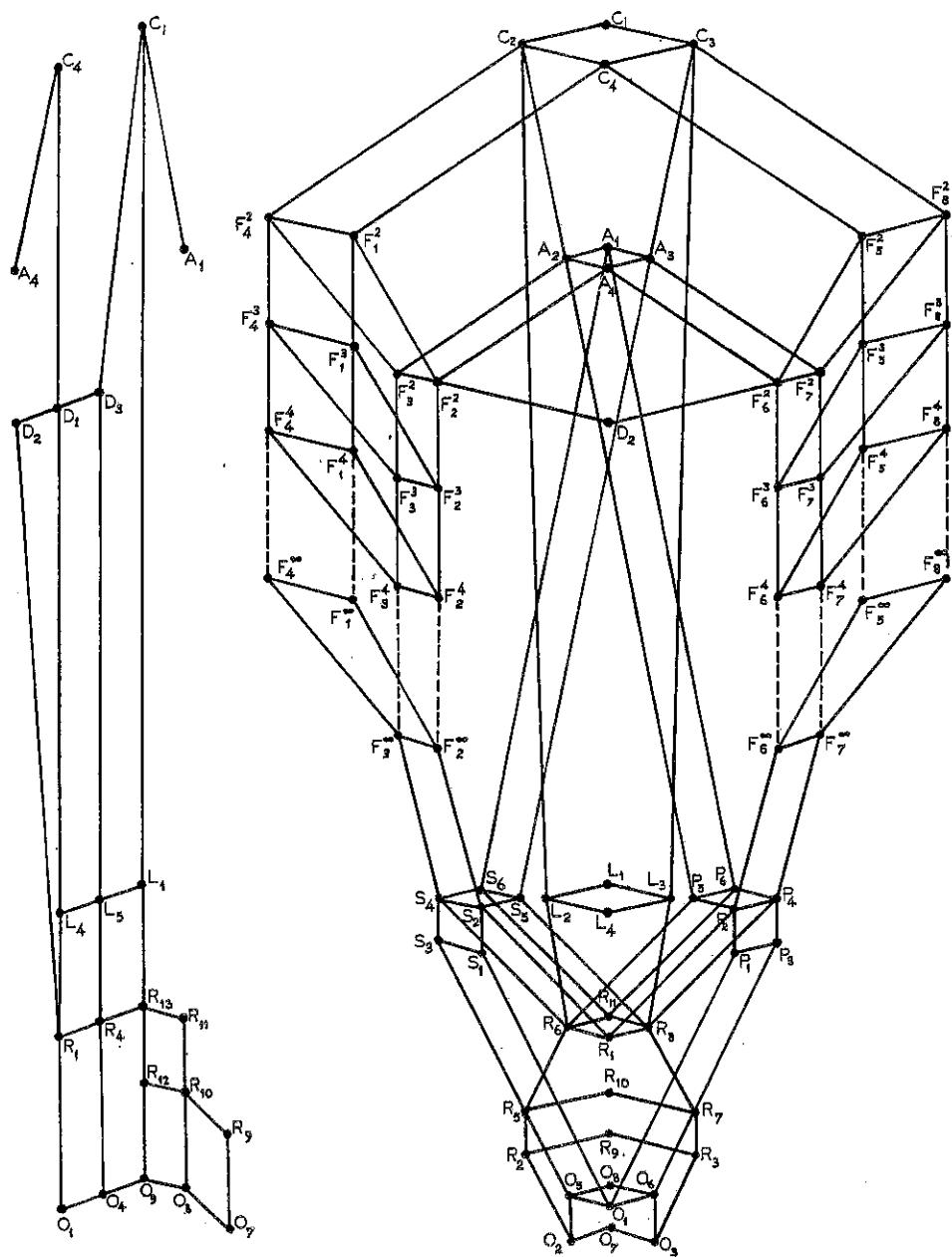
The eight statements ' $F_i^2, F_i^3, F_i^4, \dots$ immediately includes F_i^∞ ', $i = 1, 2, \dots, 8$, thus constitute a complete non-redundant solution of problem (b), since the analysis just given shows that the corresponding result for an arbitrary open inclusion chain can be deduced from these statements.

The analysis of the inclusion relation for closed systems thus achieved is presented pictorially in the accompanying inclusion diagram.⁴⁹ Dual closed systems are represented by points symmetrically placed with reference to a central vertical line, self-dual closed systems by points on this line. Due to the inherent difficulties of this requirement of symmetry, only those self-dual closed systems which immediately include, or are immediately included in non-self-dual closed systems are represented in the main diagram, while all self-dual closed systems are separately represented in the supplementary diagram at the left. For each pair of closed systems T_0, T_1 such that T_1 immediately includes T_0 we have joined the corresponding points by a straight line. The systems have been so arranged vertically, that this straight line always slopes downward from T_1 to T_0 . In the case of two self-dual closed systems T_0, T_1 , the relationship of immediate inclusion is thus represented only in the supplementary diagram. The eight dotted vertical lines serve the double purpose of indicating that the relation of immediate inclusion explicitly represented for systems in the infinite families with $\mu = 2, 3, 4$ be repeated indefinitely, and that the open inclusion chain $F_1^2, F_1^3, F_1^4, \dots, i = 1, 2, \dots, 8$, thus resulting immediately includes F_1^∞ . Now let the supplementary diagram be considered to form part of the main diagram, let all intersections at points other than those representing closed systems be disregarded, and let the diagram be sufficiently extended in the manner indicated by the dotted lines to include all closed systems with μ not exceeding those of the closed systems about to be mentioned, assuming that the latter belong to the infinite families. Then our result about inclusion chains has the following diagrammatic equivalent. A closed system T_1 properly includes a closed system T_0 when and only when there is a broken line, which may have a dotted segment, joining T_1 to T_0 which slopes downward at each point as it runs from T_1 to T_0 . The inclusion diagram thus completely determines the inclusion relation for closed systems.

The inclusion diagram can be put to a number of uses. First it can help determine the system generated by a given finite set of generators.⁵⁰ Note, either as a consequence of our chain re-

⁴⁹ Cf. G. Birkhoff [4], p. 436.

⁵⁰ And thus easily enable one to recover much specific infor-



THE INCLUSION DIAGRAM

sults, or by direct proof, that if T_0 is properly included in T_1 , then it is included in some closed system immediately included in T_1 . Now the formulas given in Part II for the closed systems enable us to determine whether a given function is or is not in a given system. We may therefore first determine whether the given set of generators is contained in some closed system T_1 immediately included in C_1 . If not, the generated system must be C_1 . Otherwise, we proceed in like fashion with respect to the systems immediately included in T_1 , and so on. This process must terminate unless it leads to some system F_1^{μ} with $\mu > 2$. In that case we pause to see if the given set of generators is contained in F_1^{∞} . If not, the original process will still terminate. Otherwise the process is re-begun with respect to F_1^{∞} , and must again terminate.

Another application of the diagram is found in the determination of the logical product and 'iterative sum' of a given set of closed systems. The logical product of a set of closed systems consists of the functions that are common to all of the systems of the set. Hence, if not null, this logical product itself constitutes a closed system. For an iterative process involving only functions which are in each of the given closed systems must yield a function which is also in each of the given systems. By the iterative sum of a given set of closed systems we mean that closed system which is generated by the logical sum of the given systems, i.e., by a set of generators, usually infinite in number, which consists of all of the functions in all of the given systems. Unlike the logical product, the iterative sum always exists for our closed systems.⁵¹

mation which formed an integral part of our original derivation of all closed systems of truth-tables, but was not required by the present derivation; e.g., the closed systems generated by the several second order functions taken one at a time, two at a time, etc.

⁵¹ Our chief reason for omitting a null closed system was to emphasize the difference between our inclusion diagram, and the inclusion diagram for the subgroups of a group, where an existent subgroup, the identity, is always common to all the subgroups. Were this omission supplied, then what we term below

Assuming their existence, we easily see that the logical product and iterative sum thus defined possess the following inclusion properties.

1. The logical product of a given set of closed systems is a closed system included in each of the given systems and including every closed system included in each of the given systems.
2. The iterative sum of a given set of closed systems is a closed system including each of the given systems and included in every closed system including each of the given systems.

In the case of the logical product the prerequisite existence depends on the condition

- a. There exists a closed system included in each of the given systems.

In the case of the iterative sum this existence follows from the property

- b. There exists a closed system including each of the given systems;

i.e., C₁. Clearly, properties 1 and 2 uniquely determine the closed system having that property. Hence, properties 1 and 2, as qualified by condition a and property b respectively, completely characterize the logical product and iterative sum in

an abstract (iterative) inclusion set is identical with a complete lattice ([5], p. 795). However, the writer is not convinced that the formulation in the present section is not to be preferred. Thus, in projective geometry, to preserve the relation $m+n = \mu+\nu$ between the dimensionality of two given spaces and that of their join and meet, it would be necessary to introduce ideal spaces of different negative dimensionality, for certain different cases of non-intersection, instead of fitting all non-intersections into the one mould of a common null-set as meet.

terms of the inclusion relation. They therefore enable us to determine the logical product and iterative sum of a given set of closed systems solely through the use of the inclusion diagram.

In fact, if the set of all closed systems be considered as an abstract inclusion set, i.e., a set of undefined 'closed systems' subject to the present inclusion relation, then 1 and 2 serve to define the logical product and iterative sum of a given set of closed systems. In that case, the fact that there exists a closed system having property 1 whenever condition a is satisfied, and that there always exists a closed system with property 2 as a consequence of b being satisfied, constitutes an important property of our inclusion set which serves to distinguish it from non-iterative inclusion sets. In this connection, however, note that 1 and 2 are not independent. For even with logical product and iterative sum thus abstractly defined, it is still readily verified that the iterative sum of a given set of closed systems is the logical product of all of the closed systems which include each of the given systems, while the logical product of a given set of closed systems is the iterative sum of all of the closed systems included in each of the given systems. That is, the existence of a closed system with property 2 is a consequence of 1 and b, while the existence of a closed system with property 1, subject to a, is a consequence of 2.

By the use of the iterative sum our previously described method of determining the closed system generated by a given set of generators can be modified as follows. We first determine by the original method the closed systems separately generated by each function in the given set of generators. Then the iterative sum of these closed systems will be the closed system generated by the entire set of generators. This modified method is very useful in the inverse problem of determining all the ways in which a given closed system can be generated by a set of independent generators. The solution of this problem for the complete system, given in the next section, constitutes our final application of the inclusion diagram.⁵²

⁵² The solution of this problem is of greater significance for propositional calculi than for Boolean algebras. For in

§26

SETS OF INDEPENDENT GENERATORS OF THE COMPLETE SYSTEM

We have proved in the preceding section that there are exactly five closed systems immediately included in the complete system, namely C_2 , C_3 , D_3 , A_1 , and L_1 . This proof, furthermore, reveals that every closed system other than the complete system is included in at least one of these five systems. It follows that the closed system generated by a given set of generators is the complete system when and only when each of the five systems C_2 , C_3 , D_3 , A_1 , and L_1 fails to contain the given set of generators.

If C_2 is not to contain a given set of generators, at least one of these generators must be a γ or a δ -function. Likewise, if C_3 is not to contain the given set of generators, at least one of these generators must be a β or a δ -function.

We thus immediately arrive at the important result that every set of generators of the complete system possesses either a δ -function, or both a β and a γ -function.

Either of these possibilities automatically prevents C_2

the latter, the primitive membership functions, using the logic of classes interpretation, may not by themselves generate all membership functions, though all membership functions may be defined with their help via the general logic in which a Boolean algebra is immersed. The reader may well feel that we overstress the condition that the complete system be generated by a set of independent generators. Thus, greater symmetry results if negation, disjunction, and conjunction are used as primitives instead of merely the first two. Without the condition of independence the problem becomes much easier, and is thus solved in the beginning of the next section. We should note that the method of solution suggested above is rather the method used in our original version of the present paper. In the present development a somewhat different method yields a much neater solution of the general problem of independent generators of the complete system. On the other hand, the specialization treated at the end of the next section was more easily arrived at via the original plan.

and C_3 from containing the given set of generators. Of the remaining three systems immediately included in the complete system, D_3 consists of all self-dual functions, A_1 of all $[A:a]$ functions, L_1 of all alternating functions. Now every δ -function fails to satisfy the $[A:a]$ condition, while every β and γ -function is non-self-dual. We can therefore restate the necessary and sufficient condition that a set of generators generates the complete system as follows.

If the set of generators possesses a δ -function, at least one generator must be non-self-dual and at least one non-alternating; if the set of generators possesses both a β and a γ -function, at least one generator must be a non- $[A:a]$ function and at least one generator must be non-alternating.

This condition, in conjunction with the result of the preceding paragraph, may be considered a solution of the problem of completely characterizing sets of generators of the complete system.

A set of generators of the complete system will clearly consist of independent generators when and only when no proper subset of the given set of generators is capable of generating the complete system.⁵³ This added restriction on a set of generators of the complete system results in a drastic limitation upon the number of generators in the set. In fact, if a set of generators of the complete system possesses a δ -function, by choosing from this set of generators the δ -function, one non-self-dual function, and one non-alternating function, we obtain a subset of the given set of generators which is capable of generating the complete system, and which consists of the δ -function and at most two other generators.

It follows that if a set of independent generators of the complete system possesses a δ -function, it can

⁵³ The definition of independence, of course, being that no generator in the set can be generated by the remaining generators.

have no more than two other generators. Likewise, if a set of independent generators of the complete system possesses both a β and a γ -function, it can have no more than two other generators.

As a result of this limitation on the number of generators in a set, it becomes possible to present a set of formulas which represent all sets of independent generators of the complete system. To achieve conciseness of statement, we introduce the following symbolism.

D : self-dual,	d : non-self-dual,
A : $[A:a]$,	a : non- $[A:a]$,
L : alternating	l : non-alternating,
R : reducible to first order,	r : not reducible to first order.

By $Dl\delta$, for example, we shall then mean any self-dual non-alternating δ -function.

In terms of this symbolism the proof that follows establishes the result that every set of independent generators of the complete system is a set of functions of one of the following forms, and conversely.

I : $\{d\delta\}$	
II : $\{D\delta, d\alpha\}$	
III : $\{Dl\delta, \beta\}$	IV : $\{Dl\delta, \gamma\}$
V : $\{D\delta, l\beta\}$	VI : $\{D\delta, l\gamma\}$
VII : $\{L\delta, L\beta, Dl\alpha\}$	VIII : $\{L\delta, L\gamma, Dl\alpha\}$
.....	
IX : $\{l\beta, \gamma\}$	X : $\{\beta, l\gamma\}$
XI : $\{L\beta, L\gamma, al\alpha\}$	
XII : $\{Lr\beta, L\gamma, l\alpha\}$	XIII : $\{L\beta, Lr\gamma, l\alpha\}$
XIV : $\{R\beta, R\gamma, Ar\alpha, Lr\alpha\}$.	

We might also interpret $Dl\delta$ as the class of all self-dual non-alternating δ -functions. Then every set of independent generators of the complete system would be a selection from one of

the above sets of classes, and conversely. These sets of classes are not mutually exclusive, i.e., certain sets have selections in common. There is no difficulty in making them mutually exclusive. But a certain uniqueness which could be shown to be possessed by the given solution would thereby be lost.

The proof of the above results divides itself into two parts according as there is, or is not, a δ -function among the generators. In the first of these two cases frequent reference is made to the following previously demonstrated properties.

1. Every α and δ alternating function is self-dual.
2. Every β and γ -function is non-self-dual.

We recall that in this case there can be no more than two generators besides the assumed δ -function.

If this δ -function is to be the only generator, it must be both non-self-dual and non-alternating. Hence, by 1, it can be any non-self-dual δ -function. We thus have I.

When there is just one generator besides the δ -function in question, this δ -function must be self-dual or it alone would generate the complete system under I, and the two generators would not be independent. The second generator therefore has to be non-self-dual and hence, for the same reason, cannot be assumed to be a δ -function. If this non-self-dual generator is an α -function, it is also non-alternating by 1, and so immediately gives II. Otherwise, by 2, the non-self-dual generator can be any β or γ -function. But then either the δ -function, or this β or γ -function must be assumed to be non-alternating. Hence III, IV, V, and VI.

Finally, there may be two generators besides the δ -function. If this set of generators is to generate the complete system, there must be among its members the δ -function, a non-self-dual function, and a non-alternating function. For independence, these three generators must be distinct, or they would constitute a proper subset of the given set capable of generating the complete system. Hence the δ -function has to be both self-dual and alternating, one of the other two generators must be non-self-dual and alternating, the other non-alternating and self-dual. By 1, the non-self-dual alternating function can only

be a β or γ -function, while by 2, and the fact that a δ -generator must in the present instance be alternating, the self-dual non-alternating function has to be an α -function. These conditions are easily simplified by 1 and 2 to yield VII and VIII.

When a set of independent generators of the complete system does not possess a δ -function, it must have both a β and a γ -function, and may have at most two other generators. We also recall that in this case at least one generator is to be a non-[A:a] function, and at least one non-alternating. The properties now useful are the following.

1. Every α , β , and γ -function reducible to first order is both [A:a] and alternating.
2. Every [A:a] β and γ -function is reducible to first order.
3. Every [A:a] alternating α -function is reducible to first order.

If there are to be no generators besides the β and γ -function, at least one of these two functions must be non-alternating. By 1, this non-alternating β or γ -function is not reducible to first order, and hence, by 2, is also a non-[A:a] function. Thus IX and X.

When there is but one generator besides the β and γ -functions, both the β and γ -function must be alternating to avoid cases IX and X. The remaining generator therefore has to be non-alternating, and consequently can only be allowed to be an α -function. Finally, one of the three generators must be non-[A:a]. If the α -function is non-[A:a], we immediately have XI. On the other hand, the β or γ -function being non-[A:a] is equivalent by 1 and 2 to its not being reducible to first order, so that we then have XII and XIII.

There thus remains the case where there are two generators besides the β and γ -functions. As in the corresponding δ -function case, the β and γ -functions must be both [A:a] and alternating, one of the remaining generators non-alternating and [A:a], the other non-[A:a] and alternating. The last two generators must therefore both be restricted to be α -functions. With the conditions on the β and γ -generators transformed by

1 and 2, and those on the α -generators by 1 and 3, we easily obtain XIV.

Before passing on to a certain specialization of the above general solution, we pause to note that the concept of the ϕ -subgroup of a group extends to our closed systems. More generally, consider an iteratively closed system S in the sense of §1, and suppose that S and each of its closed subsystems can be generated by a finite number of generators (as we have seen to be true of our closed systems of membership functions). On the one hand, it readily follows that if there be any function in S which cannot be one of a finite set of independent generators of S , the set of all such functions constitutes a proper closed subsystem of S which may then be called the ϕ -subsystem of S . An important consequence of our hypothesis on S is that each of its proper closed subsystems is included in a closed subsystem immediately included in S . Calling the latter the maximal subsystems of S (not to be confused with our previous use of this phrase), it follows as in finite group theory that the ϕ -subsystem of S is the cross-cut, i.e., logical product, of the maximal subsystems of S . In particular, the ϕ -subsystem of the complete system of membership functions C_1 is therefore the cross-cut of C_2, C_3, A_1, D_3 , and L_1 , and hence, either by the inclusion diagram, or more easily by property 3 above, is seen to be R_1 , the system of functions reducible to first order α -functions.⁵⁴

The requirement that a set of generators of the complete system consist of independent generators may be considered to be but a partial statement of the more general requirement that a set of generators of the complete system be not burdened with unnecessary complications. Various standards of irreducible simplicity may be adopted, each leading to a specialization of our general solution. We shall consider but one such specialization.

Only those sets of independent generators of the

⁵⁴ This paragraph is a recent addition. While it suffices for our present purpose, its extensions and ramifications are manifold. But such a wider treatment would have to be related to the recent developments of lattice theory.

complete system are to be admitted in which one cannot replace a generator by a function of lower order that can be generated by this generator, and thus obtain a new set of independent generators of the complete system.⁵⁵

An important consequence of this restriction, we shall see, is that it replaces the infinite number of sets of merely independent generators, infinite even when abstraction is made of the particular variables on which they are written, by a finite number of sets. Hence, while the solution of the unrestricted problem can only be given in terms of conditions imposed upon the generators, the solution of the restricted problem gives us the various individual sets themselves.

We have seen that every set of independent generators of the complete system is a selection from at least one of the sets of classes of functions I — XIV, and conversely. Our specialization clearly rejects every selection which chooses from a class a function that can generate a function of lower order in that class. We must therefore first restrict each class to those functions that cannot generate a function of lower order in that class. It will then be easy to see which of the selections from the resulting sets of restricted classes also fail to conform to our specialization.

This program can be carried out with far greater ease if we break up each of the classes $a\alpha$, $l\alpha$, and $A\alpha$ into two classes as follows:

$$a\alpha = d\alpha + Da\alpha, \quad l\alpha = d\alpha + Dl\alpha, \quad A\alpha = dA\alpha + DA\alpha,$$

correspondingly replace each of the sets of classes XI, XII, XIII, and XIV by two sets of classes, and apply the above plan to these modified classes and sets of classes. As a result of this modification, each class now consists entirely either of self-dual, or of non-self-dual functions.

⁵⁵ The new set, if a set of generators of the complete system, will be a set of independent generators.

This simplification enables us to state the uniform result that the restricted classes now consist of the functions of lowest order in each class.

To prove this result it is clearly sufficient to show for each class that every function in the class can generate one of the functions of lowest order in that class. This lowest order is easily seen to be the first order for the classes $D\delta$, $L\delta$, β , $L\beta$, $R\beta$, γ , $L\gamma$, $R\gamma$, the second order for the classes $d\delta$, $d\alpha$, $l\beta$, $l\gamma$, $Lr\beta$, $Lr\gamma$, $dA\alpha$, and the third order for the remaining classes $Dl\delta$, $Dl\alpha$, $da\alpha$, $DAl\alpha$, $DAr\alpha$, $Lr\alpha$. Our result then follows immediately for the $D\delta$, $L\delta$, β , $L\beta$, $R\beta$, γ , $L\gamma$, and $R\gamma$ classes since they consist of functions whose associated first order functions are in the same class. In the case of the $d\alpha$ and $dA\alpha$ classes, §18 shows that every function therein generates at least one of the second order functions $A+B:ab$, $AB:a+b$, both of which are in the classes concerned. A similar argument disposes of the $d\delta$ class. By §16, every function in the $Lr\beta$ class generates a second order $Lr\beta$ function. Dually for $Lr\gamma$. Each $Dl\delta$, $Dl\alpha$, $DAl\alpha$, $DAr\alpha$, and $Lr\alpha$ function generates a closed system which can be generated by a single third order function. This third order function is therefore in the same class as the given function, and is generated by the given function. For the remaining classes we have recourse to the inclusion diagram. This shows that each $l\beta$ function generates a closed system containing F_4^∞ , which in turn can be generated by the $l\beta$ function $a+B:Ab$. Dually for $l\gamma$ functions. Likewise each $da\alpha$ function generates a closed system containing F_1^∞ or F_5^∞ , each of which can be generated by a single third order $da\alpha$ function.

Each restricted class thus consists of functions of a single order not exceeding three. Before these functions are obtained explicitly, it is desirable to reconsider the final form we wish the solution of our specialized problem to take. We certainly do not wish to distinguish between sets of generators which differ solely in the particular variables used in their expression. More drastically, we may not wish to distinguish between two generators which can be obtained from each other by a mere 1-1 replacement of their arguments. If we consider such generators

to correspond to the same abstract function, our problem then takes the form of finding the different abstract sets of generators that correspond to sets of generators agreeing with our specialization. We shall adopt this point of view since it leads to the completest analysis, while if it be further desired to find all sets of generators corresponding to a given abstract set of generators, no new difficulty is encountered. In connection with the restricted classes, this requires us to know the different abstract functions that correspond to the functions therein. In fact, it will be convenient to consider the restricted classes as composed of these abstract functions, and in like manner, to reconceive our entire program in terms of abstract functions. Our main remaining problem therefore is to effect an analysis of abstract functions of the first three orders sufficient to describe the restricted classes from this new point of view.

Each of the four first order functions on a given argument gives rise to a distinct abstract function. Of these four possible abstract first order functions but three are seen to be members of the restricted classes. In terms of our previous symbolism they are β_1 , γ_1 , and δ_1 .

We will readily verify that the second order abstract functions which are members of the restricted classes comprise all second order abstract functions not reducible to first order. In counting these abstract functions, we first observe that of the sixteen second order functions on two arguments A, B, the eight functions which have Ab and aB in opposite expressions of their complete expansions give rise to but four different abstract functions. There are thus but twelve abstract second order functions. Of these, four are reducible to the four first order functions respectively. The number of abstract second order functions not reducible to first order is therefore eight. We symbolize and represent them as follows.

$$\begin{aligned} \alpha'_2, & \quad A+B:ab; \quad \beta'_2, \quad a+B:Ab; \quad \beta''_2, \quad AB+ab:Ab+aB; \quad \delta'_2, \quad ab:A+B \\ \alpha''_2, & \quad AB:a+b; \quad \gamma'_2, \quad aB:A+b; \quad \gamma''_2, \quad aB+Ab:ab+AB; \quad \delta''_2, \quad a+b:AB \end{aligned}$$

The third order abstract functions that are members of the restricted classes are perhaps most easily found by an enumeration of all abstract third order α -functions. Consider an arbitrary α -function of A, B, and C. Let the sign + or - be associated with a complete term on A, B, C according as that term is in the first or second expression of the complete expansion of the function. The terms ABC and abc then have the sign + and - respectively in all third order α -functions, and remain unchanged when A, B, and C are permuted. We therefore focus our attention on the remaining terms which we group in three pairs of dual terms as follows, (aBC, Abc), (AbC, aBc), (ABc, abC). A pair of signs can be assigned to each of these pairs of dual terms in four different ways, thus giving rise to the $4 \cdot 4 \cdot 4 = 64$ third order α -functions of A, B, and C. We observe, however, that permuting the variables A, B, C has the effect of correspondingly permuting the three pairs of dual terms. Hence, in counting abstract functions we can disregard the order in which the three pairs of signs are assigned to the three pairs of dual terms. There are thus 4 abstract functions corresponding to the three pairs of signs being all different, $4 \cdot 3 = 12$ abstract functions with two pairs of signs the same, the third different, 4 abstract functions with all three pairs of signs the same. The number of abstract third order α -functions is thus but 20. They may be represented as follows.⁵⁶

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
ABC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
aBC	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	+	-	-	
AbC	-	-	+	+	+	+	+	+	+	-	-	-	-	-	-	+	+	-	-	
ABc	-	-	-	+	-	-	+	-	+	+	-	+	+	-	+	+	-	-	-	
abC	+	+	+	-	+	-	+	-	+	+	-	+	-	-	-	+	-	+	+	
aBc	-	-	+	+	-	-	+	+	+	-	-	+	+	-	+	-	+	-	+	
Abc	+	-	-	-	-	-	+	+	+	-	-	+	+	+	+	-	+	-	-	
abc	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

⁵⁶ The key to the resulting tabulation is the above ordering of the pairs of dual terms coupled with the following ordering of the pairs of possible signs: +-, ++, --, -+, an ordering in keeping with our ordering $\alpha_1, \beta_1, \gamma_1, \delta_1$ of the four first order functions. The correctness of the tabulation can then be checked in a few minutes.

The only abstract third order functions that are members of the restricted classes are readily seen to be self-dual α -functions not reducible to first order, non-self-dual non-[A:a] α -functions, and self-dual non-alternating δ -functions. Of the 20 abstract third order α -functions, 7, 14, 17, and 20 only are self-dual, with 7 reducible to first order.

Hence, 14, 17, and 20, which we resymbolize $\alpha_3^!$, $\alpha_3^"$, and α_3^{**} respectively, are the only abstract third order self-dual α -functions not reducible to first order. The above tabulation also reveals that there are exactly ten abstract third order non-self-dual non-[A:a] α -functions, to wit, 1, 2, 3, 4, 9, 10, 12, 13, 15, and 16. We resymbolize them $\alpha_3^{(1)}$, $i = 4, 5, \dots, 13$. Finally, the abstract third order self-dual non-alternating δ -functions are clearly the negatives of the corresponding α -functions. They are therefore the negatives of $\alpha_3^!$ and $\alpha_3^"$, and may be symbolized $\delta_3^!$ and $\delta_3^"$ respectively.

We can now rapidly complete the solution of our specialized problem. With the help of the preceding analysis of abstract functions, and our proved result that the modified restricted classes consist of the functions of lowest order in each class, we readily find these modified restricted classes to have the following membership.

$$\begin{array}{llll}
 d\delta : \delta_2^!, \delta_2^" & \beta : \beta_1 & \gamma : \gamma_1 & d\alpha : \alpha_2^!, \alpha_2^" \\
 Dl\delta : \delta_3^!, \delta_3^" & l\beta : \beta_2^! & l\gamma : \gamma_2^! & Dl\alpha : \alpha_3^!, \alpha_3^" \\
 D\delta : \delta_1 & L\beta : \beta_1 & L\gamma : \gamma_1 & da\alpha : \alpha_3^{(1)}, i=4, 5, \dots, 13 \\
 L\delta : \delta_1 & Lr\beta : \beta_2^" & Lr\gamma : \gamma_2^" & Da\alpha : \alpha_3^! \\
 & R\beta : \beta_1 & R\gamma : \gamma_1 & dA\alpha : \alpha_2^!, \alpha_2^" \\
 & & & DAr\alpha : \alpha_3^" \\
 & & & Lr\alpha : \alpha_3^{**}
 \end{array}$$

The abstract sets of generators agreeing with our specialization will then be among the selections from the sets of restricted classes corresponding to the modified sets of classes I — XIV. Of these selections, the only ones still not conforming with our specialization are easily seen to be $\{\beta_2'', \gamma_1, \alpha_3^1\}$ and $\{\beta_1, \gamma_2'', \alpha_3^1\}$, where β_2'' and γ_2'' can be replaced by β_1 and γ_1 respectively.

We thus find that there are exactly thirty-six abstract sets of independent generators of the complete system that conform with our specialization.

Regrouped according to the original sets of classes I — XIV, they are as follows.

$$\text{I} : \{\delta_2^1\}, \{\delta_2^2\}$$

$$\text{II} : \{\delta_1, \alpha_2^1\}, \{\delta_1, \alpha_2^2\}$$

$$\text{III} : \{\delta_3^1, \beta_1\}, \{\delta_3^2, \beta_1\}$$

$$\text{IV} : \{\delta_3^1, \gamma_1\}, \{\delta_3^2, \gamma_1\}$$

$$\text{V} : \{\delta_1, \beta_2^1\}$$

$$\text{VI} : \{\delta_1, \gamma_2^1\}$$

$$\text{VII} : \{\delta_1, \beta_1, \alpha_3^1\}, \{\delta_1, \beta_1, \alpha_3^2\}$$

$$\text{VIII} : \{\delta_1, \gamma_1, \alpha_3^1\}, \{\delta_1, \gamma_1, \alpha_3^2\}$$

.....

$$\text{IX} : \{\beta_2^1, \gamma_1\}$$

$$\text{X} : \{\beta_1, \gamma_2^1\}$$

$$\text{XI} : \{\beta_1, \gamma_1, \alpha_3^1\}, \{\beta_1, \gamma_1, \alpha_3^{(1)}\}, \quad i = 4, 5, \dots, 13.$$

$$\text{XII} : \{\beta_2'', \gamma_1, \alpha_2^1\}, \{\beta_2'', \gamma_1, \alpha_2^2\},$$

$$\text{XIII} : \{\beta_1, \gamma_2'', \alpha_2^1\}, \{\beta_1, \gamma_2'', \alpha_2^2\},$$

$$\{\beta_2'', \gamma_1, \alpha_3^2\}$$

$$\{\beta_1, \gamma_2'', \alpha_3^2\}$$

$$\text{XIV} : \{\beta_1, \gamma_1, \alpha_2^1, \alpha_3^1\}, \{\beta_1, \gamma_1, \alpha_2^2, \alpha_3^1\}, \{\beta_1, \gamma_1, \alpha_3^2, \alpha_3^1\}.$$

Our notation immediately tells whether a generator is an α , β , γ , or δ -function, and, via the subscript of its symbol, what its order is. We thus see that of these thirty-six sets of

generators twelve are of the second order, twenty-four of the third order. Wernick [36] has observed for functions of two variables that there may be as many as three independent generators of the complete system, but not more than three. We may add that if functions of the first three orders are considered, there may be as many as four independent generators of the complete system, but that there can be no more than four no matter what the orders of the generators are.

Some of the occurrences of these sets of independent generators in the literature may be noted. Interpreted as truth-functions, we may note that $\{\delta_2'\}$ primarily, its dual $\{\delta_2''\}$ secondarily, termed rejection and incompatibility respectively, were introduced by Sheffer [27], the latter being the one used by Nicod [21]. $\{\delta_1, \alpha_2'\}$ is the negation and disjunction of Principia Mathematica [37], the dual set $\{\delta_1, \alpha_2''\}$, negation and conjunction, being employed by Lewis [15] for "material implication". Still earlier, the Principles of Mathematics [26], presumably following Frege, suggested negation and implication, $\{\delta_1, \beta_2'\}$, as primitives. These have since reappeared in the C — N calculus.⁵⁷ On the other hand, the recent C — O calculus of Wajsberg [34] may be said to use the set $\{\beta_2', \gamma_1\}$.

As pointed out elsewhere,⁵⁸ the applicability of these results to Boolean algebras is rather obscured. However, where an element interpretable as the universal class is postulated, we may perhaps consider it an added primitive; and by identifying that primitive with the constant function β_1 , the following additional occurrences may be noted. $\{\beta_1, \gamma_2'\}$ in postulates for 'exception' ([1], [31]), $\{\beta_1, \gamma_2'', \alpha_2'\}$ for "Boolean Rings with Unit", [28],⁵⁹ and perhaps $\{\delta_3'', \beta_1\}$ for 'ternary rejec-

⁵⁷ e.g., see [34]. Of course, C is β_2' , N is δ_1 .

⁵⁸ See footnote 52.

⁵⁹ Here, and in other cases, references to numerous papers of B. A. Bernstein other than those appearing in our bibliography could be given.

tion', ([8], [38]). While, as has been observed by B. A. Bernstein [1], one and the same set of postulates for a Boolean algebra will serve both for a given interpretation and its dual, in the case of propositional calculi the non-self-dual character of assertability requires entirely different postulational developments for dual sets of primitive truth-functions.

BIBLIOGRAPHY

1. B. A. Bernstein. A Complete Set of Postulates for the Logic of Classes Expressed in Terms of the Operation "Exception", and a Proof of the Independence of a Set of Postulates due to Del Re.
University of California Publications in Mathematics, Vol. 1, (1912 - '24), No. 4 (15th May, 1914), pp. 87 - 96.
2. B. A. Bernstein. Complete Sets of Representations of Two-element Algebras.
Bull. Amer. Math. Soc., Vol. 30, (1924), pp. 24 - 30.
3. B. A. Bernstein. The Dual of a Logical Expression.
Ibid., Vol. 33, (1927), pp. 309 - 311.
4. Garrett Birkhoff. On the Structure of Abstract Algebras.
Proceedings of the Cambridge Philosophical Society, Vol. 31, (1935), pp. 433 - 454.
5. Garrett Birkhoff. Lattices and Their Applications.
Bull. Amer. Math. Soc., Vol. 44, (1938), pp. 793 - 800.
6. Archie Blake. Canonical Expressions in Boolean Algebra.
Review of, Journal of Symbolic Logic, Vol. 3, (1938), p. 93.
7. Alonzo Church. An Unsolvable Problem of Elementary Number Theory.
Amer. Jour. Math., Vol. 58, (1936), pp. 345 - 363.
8. Orrin Frink. The Operations of Boolean Algebras.
Annals of Mathematics, 2 s. Vol. 27, (1925 - '26), pp. 477 - 490.

9. D. Hilbert and P. Bernays. *Grundlagen der Mathematik.*
Vol. 1, Berlin, 1934.
10. D. Hilbert and P. Bernays. *Grundlagen der Mathematik.*
Vol. 2, Berlin, 1939: supplement 3.
11. Tomoharu Hirano. *Die Kontradiktorische Logik.*
Review of, *Journal of Symbolic Logic*, Vol. 3, (1938),
p. 90.
12. W. S. Jevons. *Pure Logic.*
London, 1864.
13. A. B. Kempe. *On the Relation Between the Logical Theory
of Classes and the Geometrical Theory of Points.*
Proc. London Math. Soc., Vol. 21, (1889 - '90),
pp. 147 - 182.
14. Stanislaw Łesniewski. *Grundzüge Eines Neuen Systems der
Grundlagen der Mathematik.*
Fund. Math., Vol. 14, (1929), pp. 1 - 81.
15. C. I. Lewis and C. H. Langford. *Symbolic Logic.*
New York, 1932.
16. Eugen Gh. Mihailescu. *Recherches sur un Sous-système du
Calcul des Propositions.*
Review of, *Journal of Symbolic Logic*, Vol. 2, (1937),
p. 51.
17. Eugen Gh. Mihailescu. *Recherches sur la Négation et
l'équivalence dans le Calcul des Propositions.*
Review of, *Ibid.*, Vol. 2, (1937), p. 173.
18. Eugen Gh. Mihailescu. *Recherches sur l'équivalence et la
Réciprocité dans le Calcul des Propositions.*
Review of, *Ibid.*, Vol. 3, (1938), p. 55.
19. Eugen Gh. Mihailescu. *Recherches sur les Formes Normales
par Rapport à l'équivalence et la Disjonction, dans
le Calcul des Propositions.*
Review of, *Ibid.*, Vol. 4, (1939), pp. 91 - 92.
20. Ernest Nagel. Review by.
Ibid., Vol. 4, (1939), pp. 35 - 36.

21. J. G. P. Nicod. A Reduction in the Number of Primitive Propositions of Logic.
Proceedings of the Cambridge Philosophical Society,
Vol. 19, (Jan. 1917), pp. 32 - 41.
22. E. L. Post. Determination of all Closed Systems of Truth Tables.
Abstract, Bull. Amer. Math. Soc., Vol. 26, (1920),
p. 437.
23. E. L. Post. Introduction to a General Theory of Elementary Propositions.
Amer. Jour. Math., Vol. 43, (1921), pp. 163 - 185.
24. W. V. Quine. Review by.
Journal of Symbolic Logic, Vol. 2, (1937), p. 175.
25. Josiah Royce. An Extension of the Algebra of Logic.
Jour. Philos. etc., Vol. 10, (1913), pp. 617 - 633.
26. Bertrand Russell. The Principles of Mathematics.
2nd edition, London, 1937, New York 1938.
27. H. M. Sheffer. A Set of Five Independent Postulates for Boolean Algebras, with Application to Logical Constants.
Trans. Amer. Math. Soc., Vol. 14, (1913), pp. 481 - 488.
28. M. H. Stone. The Representation of Boolean Algebras.
Bull. Amer. Math. Soc., Vol. 44, (1938), pp. 807 - 816.
29. A. Suschkewitsch. Untersuchungen über Verallgemeinerte Substitutionen.
Atti-Congresso Bologna, Vol. 2, (1930), pp. 147 - 157.
30. Alfred Tarski. Sur le Terme Primitif de la Logistique.
Fund. Math., Vol. 4, (1923), pp. 196 - 200.
31. J. S. Taylor. A Set of Five Postulates for Boolean Algebras in Terms of the Operation "Exception".
University of California Publications in Mathematics,
Vol. 1, (1912 - 24), No. 12 (12th April 1920), pp.
241 - 248.

32. Oswald Veblen. The Cambridge Colloquium Lectures.
1916. Part II. (Analysis Situs).
33. Mordechaj Wajsberg. Beiträge zum Metaaussagenkalkul I.
Monatsh. Math. Phys., Vol. 42, (1935), pp. 221 - 242.
34. Mordechaj Wajsberg. Métalogische Beiträge.
Review of. Journal of Symbolic Logic, Vol. 2,
(1937), pp. 93 - 94.
35. Donald L. Webb. The Algebra of n-valued Logic.
Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie, Class III, Vol. 29, (1936 - '37), pp. 153 - 168.
36. William Wernick. Complete Sets of Logical Functions.
Abstract, Bull. Amer. Math. Soc., Vol. 46, (1940),
p. 235.
37. A. N. Whitehead and Bertrand Russell. Principia Mathematica.
2nd Edn., Vol. I, Cambridge, England, 1925. Part 1,
Section A.
38. Albert Whiteman. Postulates for Boolean Algebra in Terms
of Ternary Rejection.
Bull. Amer. Math. Soc., Vol. 43, (1937), pp. 293 -
298.
39. Norbert Wiener. Certain Formal Invariances in Boolean
Algebras.
Trans. Amer. Math. Soc., Vol. 18, (1917), pp. 65 - 72.
40. Norbert Wiener. Bilinear Operations Generating all Operations
Rational in a Domain Ω .
Annals of Mathematics, 2s., Vol. 21, (1920), pp.
157 - 165.
41. Norbert Wiener. Certain Iterative Characteristics of Bi-
linear Operations.
Bull. Amer. Math. Soc., Vol. 27, (1920), pp. 6 - 9.
42. Eustachy Żyliński. Some Remarks Concerning the Theory of
Deduction.
Fund. Math., Vol. 7, (1925), pp. 203 - 209.

Absolutely Unsolvable Problems and Relatively Undecidable Propositions— Account of an Anticipation¹

Emil L. Post

Introduction

There would be little point in publicizing the writer's anticipation of the existence of absolutely unsolvable problems in the sense of Church,² and, as a corollary thereof, of Gödel's theorem on the incompleteness of symbolic logics,³ merely as a claim to unofficial priority. Indeed, the present development is but fragmentary by comparison. But with the *Principia Mathematica* of Whitehead and Russell as a common starting point, the roads followed towards our common conclusions are so different that much may be gained from a comparison of these parallel evolutions. Less important is the bridge the present paper would provide between the writer's

¹ The phrase "absolutely unsolvable" is due to Church who thus described his problem in answer to a query of the writer as to whether the unsolvability of his elementary number theory problem was relative to a given logic (see footnote 2). By contrast, the undecidable propositions in this and related papers are undecidable only with respect to a given logic. A fundamental problem is the question of the existence of absolutely undecidable propositions, that is, propositions which in some a-priori fashion can be said to have a determined truth-value, and yet cannot be proved or disproved by any valid logic. An attempt at formulating such a proposition appears in the Appendix. The ideal candidate therefore would be a suitable arithmetic proposition in the sense of Gödel. For to the writer it is axiomatic that if the truth-value of $\phi(n)$ "is determined" for each natural number n , the truth-value of $(\exists n)\phi(n)$ and of $(n)\phi(n)$, n restricted to natural numbers, "is determined" — whether determined by us, or not.

The writer cannot overemphasize the fundamental importance to mathematics of the existence of absolutely unsolvable combinatory problems. True, with a specific criterion of solvability under consideration, say recursiveness (see footnote 6), the unsolvability in question, as in the case of the famous problems of antiquity, becomes merely unsolvability by a given set of instruments. And, indeed, the corresponding proofs for combinatory problems are almost trivial in comparison with the classic unsolvability proofs. The fundamental new thing is that for the combinatory problems the given set of instruments is in effect the only humanly possible set.

² Alonzo Church, *An unsolvable problem of elementary number theory*, American Journal of Mathematics, vol. 58 (1936), pp. 345–363.

³ Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik, vol. 38 (1931), pp. 173–198.

past and it is hoped future work.

The point of departure of the present paper is the formulation presented in the third part of the writer's dissertation of 1921 (see §1). Within its scope, the development of Part I is complete, and, indeed, whatever persuasiveness there is in the actual anticipation presented in Part II derives from it. The formative work which largely led to this anticipation falls within the writer's tenure as Procter Fellow in Princeton for the academic year 1920–21, and the summers immediately preceding and following. Actual dates are: for §1 and §2, October 1920; §4 and §5, August and September 1921; Part II, September and October 1921; the last entry of the Appendix, February 24, 1922.⁴

In keeping with the parallel evolution idea the present account tries to keep as close as it can to our original notes, even to the terminology employed. The footnotes then serve to give us a running critique. Because the work covered in Part I aimed at completeness as far as it went, we have there taken the liberty of correcting a number of errors, more or less serious, occurring in the originally unchecked notes, as well as adding more explanatory material than is there given. These departures are noted in the footnotes, and in every case attempt to go no further than a re-check on the notes at the time would have yielded. On the other hand, in the case of Part II we have deemed it essential not to alter or supplement the essential content of the notes, but merely connect and occasionally smooth out their presentations. Finally, except for the introductory paragraphs, the Appendix consists entirely of quotations from notes and concurrent diary,⁵ corrected only for spelling, obvious slips in words, and flagrant crimes against grammar.

Perhaps the chief difference in method between the present development and its more complete successors is its preoccupation with the outward forms of symbolic expressions, and possible operations thereon, rather than with logical concepts as clothed in, or reflected by, correspondingly

⁴ This work was significantly carried forward during June–July 1924. The first half of this month of work was devoted to developing an "operational logic" for sequences (see the next to the last paragraph of §8 and footnote 8). This development was completely detailed as far as it went. The consideration of quantifiers led to the second half of the month's work where a hierarchy of "mandates" (directions), "super mandates" etc. were introduced. At the time, this hierarchy was thought to exhaust all possibilities, but, at least by 1929, was later seen to but include mandates of finite type, admitting of extension into the transfinite. The more detailed discussion of mandates of lowest type all but anticipated the writer's published note of 1936 (see footnote 6). We might add that since February 1938 we have given an occasional week to a continuation of this work, and largely in the spirit of the Appendix. Our goal, however, is now an analysis of proof, perhaps leading to an absolutely undecidable proposition, rather than an analysis of finite process.

⁵ This diary, under the title "Time Accounts," was begun in the spring of 1916 and continued without interruption to the spring of 1922.

particularized symbolic expressions, and operations thereon. While this in part is perhaps responsible for the fragmentary nature of our development, it also allows greater freedom of method and technique. In particular, it reveals the "Gödel Representation" to be merely a case of *resymbolization*, and suggests that with the growth of mathematics such resymbolization will have to be effected again and again.

Apart from the indirect rôle the present paper may play as a different, if imperfect, pair of lenses with which to view recent developments, it has a direct contribution to make to present day literature in adding still another precise formulation to the list of general recursiveness, λ -definability, computability.⁶ Where in these formulations the informal basic idea is that of effective calculability, our own is that of a *generated set*. This derives from the idea of a symbolic logic rather than that of an algorithm, and may be described by saying that each member of the set is at some time generated⁷ by the continued application of a given method, while that method will at no time yield an individual on the primitives of the set which is not in the set. Our emphasis is on generated sets of sequences on a fixed set of symbols a_1, a_2, \dots, a_μ .⁸ The precise formulation is that of a *normal system* on a fixed set of symbols (see the beginning of §9), and our identification, not as definition, but as at least partially verified conclusion,⁹ is of

⁶ For the first two see footnote 2, for the third see A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, Proc. London Math Soc., vol. 42 (2) (1937), pp. 230–265, later referred to as "*computable numbers*." We might also add, E. L. Post, *Finite combinatory processes – formulation I*, Journal of Symbolic Logic, vol. 1 (1936), pp. 103–105.

In this connection we must emphasize the distinction between a formulation which includes an equivalent for every possible "finite process," and a description which will cover every possible method for setting up finite processes. It is toward the latter goal that the Appendix of the present paper strained, the first having been achieved in the generalization ending §7. While the Turing simplifications referred to in footnote 9 may make the detailed development envisioned in the Appendix unnecessary for the analysis of process, though retaining an intrinsic interest as added description, it is doubtful if Turing considerations can replace such a development in the analysis of proof.

⁷ Produced, created – in practice, written down.

⁸ More exactly, "strings" of such symbols to use a term of C. I. Lewis (*A Survey of Symbolic Logic*, Berkeley, CA 1918: Chapter VI, Section III). On the other hand, in the 1924 "opérational logic" for sequences, referred to in footnote 4, after many pros and cons, this concept was given up for essentially the *Principia Mathematica* concept of sequences of repeatable elements. Though this avoided the difficulties of "identity" for sequences as strings, and allowed a considerable development prior to the introduction of quantifiers, the ultimate wisdom of this decision is questionable.

⁹ In this connection see the last paragraph of the writer's note referred to in footnote 6. However, should Turing's finite number of mental states hypothesis (*computable numbers*, p. 250) bear up under adverse criticism, and an equally persuasive analysis be found for all humanly possible modes of symbolization, then the writer's position, while still tenable in an absolute sense, would become

generated set of sequences on a_1, a_2, \dots, a_μ with an *augmented normal system*¹⁰ on a_1, a_2, \dots, a_μ ; i.e., the subset of sequences a_1, a_2, \dots, a_μ only, of a normal system on a_1, a_2, \dots, a_μ , and additional letters $a'_1, a'_2, \dots, a'_{\mu'}$. The reductions of §§4,5, coupled with certain private correspondence with Church, make it certain that "generated" set in this precise sense is equivalent to "recursively enumerable set," while a "recursive set" of sequences on a_1, a_2, \dots, a_μ would be one for which both it and its negative, i.e., complement with respect to the set of all finite sequences on a_1, a_2, \dots, a_μ , are generated sets.¹¹ We thus reverse the usual order of these two concepts; and since our "generated set" is not burdened with the added ordering superimposed on the "recursively enumerable set," there may be certain advantages in making it the primary concept.

But perhaps the greatest service the present account could render would stem from its stressing of its final conclusion that *mathematical thinking is, and must be, essentially creative*.¹² It is to the writer's continuing amazement that ten years after Gödel's remarkable achievement current views on the nature of mathematics are thereby affected only to the point of seeing the need of many formal systems, instead of a universal one. Rather has it seemed to us to be inevitable that these developments will result in a reversal of the entire axiomatic trend of the late 19th and early 20th centuries, with a return to meaning and truth. Postulational thinking will then remain as but one phase of mathematical thinking. While in the Appendix we may be but following a will-o'-the-wisp, its very gropings may

largely academic.

¹⁰ Though we have had no occasion to introduce this phrase in our account proper, it occurs frequently in our notes, especially in the work of 1924. The concept, of course, is present in the generalization ending §7. In describing an augmented normal system as a set of sequences, we continue the phraseology of §9 where a normal system is so considered. Actually, apart from the purposes of §9, the notes' concept of normal system included also the methods of generating the sequences in question, and the same is true of augmented normal system. Thus, in 1924, the notes repeat that every set of finite sequences on a_1, a_2, \dots, a_μ that can be generated, can be generated by an augmented normal system.

¹¹ In accordance with footnote 96, the concept of recursive set of sequences appeared in our work in the following equivalent form: a generated set of sequences for which there is a finite-normal-test (see §9). But it was not then positively stressed. By 1924 its importance was beginning to be recognized, the formulation given in the body of the introduction then appearing. But its use was postponed for later applications, with generated set remaining the basic concept.

¹² Yet, as this account emphasizes, the creativeness of human mathematics has a counterpart inescapable limitation thereof – witness the absolutely unsolvable (combinatory) problems. Indeed, with the bubble of symbolic logic as universal logical machine finally burst, a new future dawns for it as the indispensable means for revealing and developing those limitations. For, in the spirit of the Appendix, Symbolic Logic may be said to be Mathematics become self-conscious. [Actually, the old dream of symbolic logic is finding partial realization in Tarski's recent positive work on decision problems.]

in some measure give an inkling of the creative unity that is Mathematics.

Part I. Formal Transformations

1. Canonical form A and its reduction to a canonical form B

In a previous paper¹³ the writer proposed the following formal postulational generalization of the (\sim, \vee) system of *Principia Mathematica*.¹⁴ For an arbitrary finite set of primitive functions of propositions

$$f_1(p_1, p_2, \dots, P_{m_1}), \dots, f_\mu(p_1, p_2, \dots, p_{m_\mu}).$$

of an arbitrary finite number of arguments each, assume a set of postulates of the following form.

- I. If p_1, \dots, p_{m_1} are elementary propositions, so is $f_1(p_1, \dots, p_{m_1})$.
.....
If p_1, \dots, p_{m_μ} are elementary propositions, so is $f_\mu(p_1, \dots, p_{m_\mu})$.
 - II. The assertion of a function involving a variable p produces the assertion of any function found from the given one by substituting for p any other variable q , or $f_1(q_1, \dots, q_{m_1})$, ..., or $f_\mu(q_1, \dots, q_{m_\mu})$.
 - III.

“ $\vdash g_{11}(P_1, P_2, \dots, P_{k_1})$ ”	...	“ $\vdash g_{\kappa 1}(P_1, P_2, \dots, P_{k_\kappa})$ ”
.....
“ $\vdash g_{1\kappa_1}(P_1, P_2, \dots, P_{k_1})$ ”	...	“ $\vdash g_{\kappa\kappa}(P_1, P_2, \dots, P_{k_\kappa})$ ”
produce	...	produce
“ $\vdash g_1(P_1, P_2, \dots, P_{k_1})$ ”	...	“ $\vdash g_\kappa(P_1, P_2, \dots, P_{k_\kappa})$ ”,

where the P 's are any combinations of f 's including the special case of the unmodified variable, while the g 's are particular combinations of this kind which need not have all the indicated arguments.

- $$\begin{aligned} \text{IV. } & \vdash h_1(p_1, p_2, \dots, p_{\ell_1}), \\ & \vdash h_2(p_1, p_2, \dots, p_{\ell_2}), \\ & \dots \\ & \vdash h_\lambda(p_1, p_2, \dots, p_{\ell_\lambda}) \end{aligned}$$

¹³ E. L. Post, *Introduction to a general theory of elementary propositions*, Amer. Journ. Math., vol. 43 (1921), pp. 163-185. See §8 p. 176 thereof.

¹⁴ That is, its propositional calculus. See A. N. Whitehead and B. Russell, *Principia Mathematica*, 2nd. edn., vol. I, Cambridge, England, 1925: Part I section A.

where the h 's are particular combinations of the f 's.¹⁵

As in the description of the (\sim, \vee) system given in that paper, we may observe that I may be said to determine the enunciations of the system under consideration.¹⁶ On the other hand, the repeated application of operations II and III to the primitive assertions in IV, and all derived assertions that may thus result, yield a subset of the set of enunciations consisting of the assertions of the system.¹⁷ In the case of the (\sim, \vee) system the writer proved, in the paper referred to, that by introducing the truth-table concept a finite method was thereby afforded for determining of any enunciation in the system whether it was or was not an assertion of the system. We shall say that we thus solved the *finiteness problem*¹⁸ for the (\sim, \vee) system. While this solution was purely formal, nevertheless it was suggested by the intuitive interpretation of " \sim " and " \vee ." For the above

¹⁵ by a combination of f 's is meant any expression built up out of the primitive functions and variables by the operation of substitution — strictly so for the P 's of III, with added abstraction of the final variables for the g 's and h 's. Where our notes always talk of "the capital P 's," we have allowed ourselves occasionally the phrase "the operational variables." These occur only in the basis of a system, while "variables" refer to the p 's, q 's etc. which are in the system itself. Note that our formulation is really that of a class of systems, each explicitly given basis yielding a corresponding system.

¹⁶ "Enunciation" is at least approximately equivalent to the more recent "well-formed formula."

¹⁷ We take this opportunity to point out an error in §10 of the above paper. Lemma 1 thereof requires the added condition that the expressions replacing the r 's do not involve any letter upon which a substitution is made in the given deductive process. This necessitates several minor changes in the proof of the basic theorem that follows. Actually, a little further analysis than was effected at the time allows the lemma 2 of the paper to be strengthened to

Lemma 2'. *The most general process of obtaining an assertion from a given set of assertions in accordance with II and III can be reduced to first asserting a number of functions in accordance with II, and then applying only operations III.* It then suffices to replace Lemma 1 by the simpler

Lemma 1' *If a given set of functions gives rise to some other function solely through the use of operations III, then the same deductive process will be valid if we have given the original functions with an arbitrary substitution on their letters as described in II, provided this substitution is also made throughout the process.* With these simpler lemmas the proof of the theorem is valid as it stands. In this connection see footnotes 21 and 24.

¹⁸ That is, the "deductibility problem" in the sense of Church. "Decision problem," if a translation of "Entscheidungsproblem," seems to have this as but one of two distinct meanings. (See Alonzo Church, *Correction to A note on the Entscheidungsproblem*, Journal of Symbolic Logic, vol. 1 (1936), pp. 101–102.) An alternative name for the "finiteness problem" in our notes is "the fundamental problem." This name occurs in an unpublished "Note on a Fundamental Problem in Postulate Theory" of the writer bearing the date June 4, 1921. This note was then left at Princeton, presumably for publication, but was withdrawn the following fall as a result of the nullifying of its program by the anticipation recorded in Part II.

generalizations such interpretations are not at hand. Nevertheless, even before the publication of the above paper, the writer solved the finiteness problem for those of the above systems in which the primitive functions are all functions of one variable, the resulting relative simplicity of the systems allowing a direct analysis of the formal processes involved.¹⁹ While considerable further labor produced but minor dents in the problem for the above systems not so restricted, impetus was lent to the work by our formally reducing the subsystem of *Principia Mathematica* treated in *10 and *11 thereof²⁰ to a system of the above type. For thereby a solution of the finiteness problem for all of the above systems would immediately lead to a solution of that problem for this important subsystem of *Principia Mathematica*. Actually, it is necessary to replace the above formulation by an equivalent formulation, the reduction in question being to a system in this modified form. To avoid the lengthy interruption of the direct proof of the equivalence of these two formulations, but one half of this equivalence is established in the present section, the rest following from Sections 4 and 5 with the help of the short additional §6.

Before this modification can be introduced, we must reexamine a certain dubious feature of the above original formulation. Formally, I should have been recast in operational form à la II, III, and IV to constitute a method for generating all enunciations of the system. A symbol for enunciation corresponding to “ \vdash ” for assertion could have been introduced. Of course, a vital difference between the two iterative processes thus set up would have been that whereas enunciations thus arise with, as it were, complexities monotonically increasing, not so, in general, for assertions. Now II is precisely enough stated to be independent of this assumed generation of the enunciations of the system. But III, as given, is so dependent. Now actually, the description of the P 's occurring in III could have been omitted for any P occurring in a premise therein. For once that premise has arisen as an assertion, and is written in the assumed form, the P 's thereof can only be variables or enunciations. But not so for a P occurring in the conclusion of a production, but in no premise thereof; and III as stated allows such productions. Furthermore, when such a production is used, it has the theoretical disadvantage that the premises do not determine the conclusion, but a class of conclusions made precise only by the precise generation of enunciations. In our new formulation we therefore replace III by

III' : III with the added restriction that each capital P of a conclusion is present in at least one premise of the corresponding production.²¹

¹⁹ An abstract of this as yet unpublished paper appears under E. L. Post, *On a simple class of deductive systems*, Bull. Amer. Math. Soc., vol. 27 (1921), pp. 396–7.

²⁰ That is, its restricted functional calculus.

²¹ This change in III is not made in our notes; but, in point of fact, all of the

Our principal aim however is to weaken operation II; for a formal system to which *10-*11 *Principia Mathematica* is to be reduced must not allow, for example, the replacing of a variable x in an expression involving $f(x)$ by anything except a variable, and that distinct from the variable f . We therefore replace II by the very weak

II' : II restricted to the *replacing of a variable by any other variable, and that not present in the given assertion.*

By iterating II' we can therefore first perform any 1-1 replacement of the variables in an assertion by variables not in the assertion, and hence, finally, any 1-1 replacement of variables by variables.

In saying that the formulations *A*: I, II, III, IV, and *B*: I, II', III', IV, are *equivalent* we mean that the solution of the finiteness problem for all systems coming under either formulation would lead to the solution of the finiteness problem for all systems coming under the other formulation. More specifically, a system S_1 will be said to be *reduced* to a system S_2 if a method is presented which would transform a solution of the finiteness problem for S_2 into one for S_1 .²² We shall say that a system falling under a formulation X is in *canonical form X*. A system S_1 will be said to be *reduced to canonical form X* if it is reduced to a system in canonical form X , and canonical form Y will be said to be reduced to canonical for X if a method is presented whereby each system in canonical form Y is reduced to canonical form X . Canonical forms X and Y are then equivalent if each is reducible to the other.

We proceed to prove that canonical form *A* is reducible to canonical form *B*.²³

productions occurring in our notes corresponding to the work of §2 on do satisfy the restriction present in III'. The notes attempt in part to remedy this defect in III by suggesting the convention that the general description given for the P 's only apply to P 's in the premise of a production, while a P in the conclusion thereof not present in any premise represent a variable only. Due to the presence of II, this convention does not alter the effectiveness of II and III combined. On the other hand, the notes explicitly use Lemma 2' as given in footnote 17, but this strengthening of Lemma 2 is not possible under the proposed convention. It may be that an implicit assumption of this convention was responsible for the use of Lemmas 1 and 2 instead of 1' and 2' in the paper referred to. We finally note that with III' replacing III, the proofs of the emended and modified lemmas are greatly simplified.

²² Actually, no definition is given in the notes. But the reference therein to transforming one system into part of another, coupled with the actual developments given, suggest that the following more precise definition was tacitly assumed. A system S_1 is reduced to a system S_2 if a 1-1 correspondence is (effectively) set up between the enunciations of S_1 , and certain of the enunciations of S_2 , so that an enunciation of S_1 is asserted when and only when its correspondent in S_2 is asserted.

²³ This, the easier part of the equivalence proof, was given incorrectly in the notes and probably for that very reason. The error was subsequently noted,

Let S_1 be any system in canonical form A . By a lemma stated in our earlier paper,²⁴ an arbitrary proof in S_1 can always be replaced by one which first merely obtains a set of assertions by the repeated use of II starting with assertions in IV, and then merely repeats the use of III starting with the set thus found. Now the assertions of S_1 that can be thus found via II and IV only are all enunciations of the form $h_i(P_1, P_2, \dots, P_{\ell_i})$ $i = 1, 2, \dots, \lambda$, where the P 's are arbitrary variables or enunciations in S_1 . It will be convenient in this connection to include variables in the class of enunciations of a system. We therefore introduce a new primitive function $e(P)$, and set up a system in canonical form B such that $e(P)$ will be asserted in that system when and only when P is an enunciation in S_1 . For that system, I is to be the I of S_1 with the additional postulate corresponding to the primitive function $e(P)$, II' the one II' of all systems in canonical form B . Our immediate purpose is then achieved by taking for III' and IV the following.

III'. " $\vdash e(P_1)$ "

.....

" $\vdash e(P_{m_i})$ " $i = 1, 2, \dots, \mu$.

produce

" $\vdash e(f_i(P_1, \dots, P_{m_i}))$ ".

IV. $\vdash e(p)$.

That the desired result is thus achieved may be seen from the fact that this IV and II' yield all " $\vdash e(P)$ "'s where P is a variable, that is, where P is any enunciation of rank zero in S_1 ;²⁵ and, assuming inductively that the same is true for all enunciations in S_1 of rank ρ , III' insures the result

but, while saying that it could easily be overcome, the further error noted in footnote 21 arose. Actually, the notes weaken operation II in two successive stages, the first allowing a variable to be replaced by another variable. The bulk of the work concerns the reducibility of each of the resulting canonical forms to canonical form A . In the case of the intermediate form, the proof depends on an unpublished method of the writer (the L. C. M. process referred to early in §3), but the method appears in simplified form in the second reduction of §6. In the case of the present canonical form B (except for III'), the method depended on replacing the variables p_1, p_2, p_3, \dots by

$$a(p), a(a(p)), a(a(a(p))), \dots$$

$a(p)$ being a new primitive function, and thus was a precursor of the method fully presented in §4.

²⁴ So say the notes. But while the paper states Lemma 2, this restates Lemma 2' (see footnotes 17 and 21). Whether this was an oversight on the part of the notes, or the result of further analysis, is not clear.

²⁵ The rank of an arbitrary enunciation $f_i(P_1, \dots, P_{m_i})$ is then inductively defined as one more than the maximum of the ranks of P_1, \dots, P_{m_i} .

for all enunciations in S_1 of rank $\rho + 1$. Hence every " $\vdash e(P)$ " with P an enunciation in S_1 appears in our new system; and a similar induction shows that no other " $\vdash e(P)$ "'s thus appear than with P an enunciation in S_1 .

We now add to the above III', keeping I, II', and IV unchanged. These additions are such that no conclusion of a production added to III' is of the form $e(P)$. Hence no new assertions of the form " $\vdash e(P)$ " result.

First add to III the productions

$$\begin{aligned} \text{III': } & \quad \vdash e(P_1) \\ & \dots \\ & \vdash e(P_{\ell_i}) \quad i = 1, 2, \dots, \lambda. \\ & \text{produce} \\ & \vdash h_i(P_1, \dots, P_{\ell_i}). \end{aligned}$$

We shall then have asserted in our system all assertions in S_1 obtained solely by the use of the II and IV of S_1 . The effect of III of S_1 is then easily reproduced in our new system as follows. Let the i -th production in III of S_1 have $P_{j_{i,1}}, \dots, P_{j_{i,\nu_i}}$ for those of its operational variables that are present in the conclusion of that i -th production, but in no premise thereof. The full effect of III of S_1 will then be achieved if we add to III',

$$\begin{aligned} \text{III': } & \quad \vdash g_{i1}(P_1, \dots, P_{k_i}) \\ & \dots \\ & \vdash g_{i\kappa_i}(P_1, \dots, P_{k_i}) \\ & \vdash e(P_{j_{i,1}}) \quad i = 1, 2, \dots, \kappa, \\ & \dots \\ & \vdash e(P_{j_{i,\nu_i}}) \\ & \text{produce} \\ & \vdash g_i(P_1, \dots, P_{j_{i,1}}, \dots, P_{j_{i,\nu_i}}, \dots, P_{k_i}), \end{aligned}$$

the g 's being those of the III of S_1 . In fact, with " $\vdash e(P)$ " read meaningfully, as is justified by our proved result concerning its occurrence in our new system, these additions to III' become exactly the III of S_1 .

Let then S_2 be the last found system, i.e., having for basis the I, II', IV as first given for our new system, and the III' consisting of the several parts thereof listed above. S_2 is then in canonical form B . Clearly the enunciations of S_1 are among the enunciations of S_2 . By the restatement of the way in which an arbitrary assertion of S_1 can be found, it also follows from the above that every assertion in S_1 is also an assertion in S_2 . In fact, it is immediately seen that apart from assertions of the form " $\vdash e(P)$ " the assertions in S_2 are the assertions in S_1 . The solution of the finiteness problem for S_2 thus immediately becomes a solution of the finiteness problem for S_1 . That is, S_1 in canonical form A has been reduced

to S_2 in canonical form B .

2. Reduction of *10-*11 Principia Mathematica to canonical form B .²⁶

Leaving aside for the moment the question of real versus apparent variables,²⁷ we observe that *10 allows for three distinct classes of variables: propositional variables p, q, r, \dots , individual variables x, y, z, \dots , functional variables f, g, h, \dots . On the other hand, a system in canonical form B makes no distinction between its variables, permitting, indeed, any 1-1 replacement of variables in an assertion. Actually, this distinction may also be disregarded in *10. For a propositional variable p always appears in some context $\sim p, p \vee P, P \vee p$, a functional variable f in some context $f(x_{i_1}, x_{i_2}, \dots, x_{i_n})$, an individual variable x in some context $f(\dots, x, \dots)$, so that the type of the variable is determined by the context, and needs no other distinguishing mark. For purposes of presentation we shall continue to use the symbol differentiations suggested by *Principia Mathematica*. But theoretically, $x \vee p(f)$ is a valid enunciation of *10, x being now a propositional variable, f an individual variable, p a functional variable. With this understanding, any 1-1 replacement of variables in an enunciation of *10 leaves it an enunciation of *10.

In order to make *10 of *Principia Mathematica* correspond as closely as possible to *9 thereof, we shall assume that in an enunciation of *10 all apparent variables will be distinct, and distinct from any real variables occurring therein.²⁸ No loss of generality is thus incurred, since *10 was built up on the assumption that the particular symbols used for apparent variables is irrelevant. On the other hand, as a result of this convention, the enunciations of *10 will consist of those of its unrestricted enunciations which could appear in *9 as definitions.²⁹

²⁶ Since, with the exception of the explanatory *11.07, the basis for this subsystem is given in *10, we shall henceforth refer to this subsystem as *10, it being understood that in its verbal title "one apparent variable" is replaced by "an arbitrary finite number of apparent variables." Actually, our version of this subsystem is narrower than that of *Principia Mathematica*, since we assume the variables x, y, z, \dots thereof to have a common range. This explains the concluding remark of the notes, "This completes the discussion (except for some questions on type)." At the start the notes state, "We shall now attempt to prove that the development of *10 is equivalent to the entire set A or B or C ," but at this stage only the reducibility of *10 is considered. Concerning the other half of the equivalence see the second paragraph of footnote 79.

²⁷ Free versus bound variables, in more recent terminology.

²⁸ This contrasts strongly with the treatment of apparent, i.e., bound variables in the critical treatments appearing in the literature, and is one of the features to be noted in the "parallel evolution" idea referred to in the introduction.

²⁹ Despite the theoretical advantages of an ideal *9 development over that of *10, the writer was forced to decide in favor of the latter because he has found the basis of *9 to be incapable of yielding the complete development desired. To

With this understood, we may then completely determine the enunciations of *10 as follows.³⁰ It is convenient to allow an unmodified variable, thus identified as propositional, to be an enunciation of *10. The enunciations of rank 0 of *10 are all unmodified variables p , and all simple functional expressions³¹ $f(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ where the individual variables x_{i_j} are arbitrarily distinct or repeated, but the functional variable f is distinct from all the x_{i_j} 's. The remaining enunciations of *10 are then inductively determined by the following three rules.

- (a). If P is an enunciation of *10, $\sim P$ is an enunciation of *10.
- (b). If P is an enunciation of *10 involving a real individual variable x , $(x)P$ is an enunciation of *10.
- (c). If P and Q are enunciations of *10, then $P \vee Q$ is an enunciation of *10 provided the following conditions are satisfied. (α) Any variable occurring in both P and Q either appears as a propositional variable in both P and Q , or as an individual variable in both, or as a functional variable in both. (β) In the third case of the preceding condition the corresponding simple functional expressions involve the same total number of arguments. (γ) Each apparent variable of P , and of Q , is distinct from all of the individual variables of Q , and of P respectively.

Note that $(\exists x)f(x)$ need not be referred to explicitly, since it is definable in *10. As a consequence of these rules of formation, consistency of context is preserved for each separate enunciation. Hence the need of (α) in (c); also of (β), for while we may have $f(x)$ and $f(x, y)$ occurring in different contexts, not so for one and the same enunciation. Condition (γ) of (c) embodies the above convention on apparent variables. It follows inductively that given any part of an enunciation P constituting a constituent of P of the form $(x)Q$ then x appears nowhere else in P than in that $(x)Q$.³²

We have already defined the enunciations of rank 0 of *10. A unique

be specific, the proofs to the effect that the primitive propositions of the (\sim, \vee) system are valid in *9 when p, q , and r are replaced by arbitrary enunciations of *9 do not universally go over when more than one of these enunciations involve an apparent variable, or when any of them involves more than one apparent variable. A lengthy communication to this effect was sent to one of the authors of *Principia Mathematica*. However, the heavy handed modification of *9 proposed by the writer to remove the inadequacy of the original formulation was quite out of keeping with the finesse of *Principia Mathematica*. Whether the new *8 of the revised edition (Appendix, vol. I) overcomes this difficulty, the writer cannot say. At any rate, that is not its expressed purpose.

³⁰ While the formulation given in this paragraph does not explicitly appear in the notes, it is clearly tacit in all of the development that is there given.

³¹ This phrase does not occur in the notes.

³² It would not then have occurred to the writer to define the occurrences of propositional variables etc. inductively as has since been done for free and bound variables, but such definitions clearly can be given.

rank can then be inductively assigned to each enunciation of *10 as follows. If P is of rank ρ , $\sim P$ and $(x)P$ are of rank $\rho + 1$; if P and Q are of ranks ρ_1 and ρ_2 , $P \vee Q$ is of rank $\rho + 1$, where $\rho = \max(\rho_1, \rho_2)$.

The enunciations of *10, while employing the parenthesis notation of canonical form B , cannot themselves be considered enunciations of the system in canonical form B to which *10 is to be reduced. Thus, in the latter but a finite number of primitive functions can be allowed, while in *10 we must allow for an infinite number of functional variables. We therefore first give a method of translating the enunciations of *10 into certain enunciations of the system to which it is to be reduced.

Variables of *10 are to be their own correspondents in the new system. To obtain the correspondents of the simple functional expressions of *10 we introduce two primitive functions, $a(F, X)$, $b(X, Y)$. The correspondent of $f(x)$ will then be $a(f, x)$, of $f(x_{i_1}, x_{i_2}, \dots, x_{i_n})$,

$$a(f, b(x_{i_1}, b(x_{i_2}, \dots, b(x_{i_{n-1}}, x_{i_n} \dots))).$$

To take care of $(x)P$ we introduce the primitive function $O(X, P)$, and inductively define the correspondent of $(x)P$ as $O(x, \bar{P})$, where \bar{P} is the correspondent of P . $\sim P$ and $P \vee Q$ may directly be introduced as primitive functions in the new system,³³ the correspondents of $\sim P$ and $P \vee Q$ of *10 being $\sim \bar{P}$ and $\bar{P} \vee \bar{Q}$, \bar{P} and \bar{Q} being the correspondents of P and Q . Given any enunciation of *10, a unique corresponding enunciation of the new system is thus determined, different enunciations of *10 always having different correspondents.

A major part of the present development is devoted to formally determining that subset of the set of all enunciations of the new system which consists of the correspondents of the enunciations of *10. For this purpose we introduce a primitive function $\alpha(P)$, and set up a canonical B basis as a result of which $\alpha(P)$ will be asserted when and only when P is the correspondent of an enunciation of *10. It will be convenient to set up this basis simultaneously with an inductive proof of its sufficiency based on the rank of an enunciation of *10.³⁴ This proof however will first require

³³ $P \vee Q$ should rather be written $\vee(P, Q)$, as is done in an illustration of the reduction of §4 appearing in our notes and, more extensively, in our work of 1924.

³⁴ Considerable liberty is here taken with the order of events as compared with the development in the notes. The notes first set up the basis for $\alpha(P)$ with P representing any matrix of *10, and carry the induction proof through to the point where it could "obviously" be completed. They then attempt to allow for the arbitrary introduction of apparent variables into P ; but what seemed to be a clever ending of this enterprise turns out to be quite inadequate. Perhaps because of the necessary duplication in proof caused by this order of events, the attempted proof by induction ends up with an "obviously" before this error could be discovered. Except for a relatively few changes indicated in succeeding footnotes, the actual basis as now presented is on the whole but a reshuffling of the basis developed in the notes.

the setting up of a partial basis for securing the assertion of $\alpha(P)$ in the following three cases: (a) P of rank 0, (b) P of the form $O(x, Q)$ where Q is of rank 0, (c) P of the form $Q \vee R$ where Q and R are of rank 0. By the rank of P we do not mean the rank of P as enunciation in the new system, but as translation of an enunciation of *10, i.e., we mean the rank of the latter enunciation. Likewise we shall refer to a simple functional expression P when we mean P is the correspondent of a simple functional expression.

Case (a) is easily covered by the following two primitive assertions and one operation.³⁵

$$(1). \vdash \alpha(p). \quad (2). \vdash \alpha[a(f, x)].$$

I. " $\vdash \alpha[a(F, X)]$ ", " $\vdash \alpha[a(X_1, F)]$ " produce " $\vdash \alpha[a(F, b(X_1, X))]$ ".³⁶ Of course, the operation of substitution of canonical form B is assumed already to have been postulated. With its help, (1) alone yields all " $\vdash \alpha(P)$ "'s where P is an unmodified, consequently a propositional, variable. As for P a simple functional expression, (2) alone, under substitution, yields all " $\vdash \alpha(P)$ "'s with P a simple functional expression of one argument. Note that f and x being distinct variables in (2), they remain distinct under substitution. In the inductive application of I, an assertion in the form of the first premise insures $a(F, X)$ being a simple functional expression with F a variable distinct from any variable in X , the second premise insures X_1 being a variable distinct from F so that in the conclusion $a(F, b(X_1, X))$ will be a simple functional expression satisfying the necessary condition, F distinct from any variable in $b(X_1, X)$. Whatever distinct variables F and X_1 are, that second premise can always be realized by a form of (2) under substitution. It follows that (2) and I suffice to yield all " $\vdash \alpha(P)$ "'s where P is a simple functional expression, as was desired. Our analysis, and the inapplicability of I to any other assertions, also shows that only " $\vdash \alpha(P)$ "'s with P a valid enunciation of rank 0 are thus obtained.

Case (b) will be covered by further adding

$$(3). \vdash \alpha O(x, a(f, x)).$$

³⁵ The notes have the added primitive assertion " $\vdash \alpha[a(f, b(x_1, x_2))]$ " necessitated by their using for the second premise of I, " $\vdash \alpha[a(f, b(X_1, X_2))]$ ". The trick employed in the present second premise, however, is used later in the notes. Where in (2) we now think of f and x as being the variables employed, the notes rather think of them as representing those variables, and so add the condition that those variables be distinct. Perhaps because of this difference, they invariably here use small letters for those operational variables which in fact could only represent variables. Yet, in a preliminary $\alpha(P)$ development immediately preceding, they specifically point out that all operational variables must be written as capitals, a practice we have uniformly followed in the present version of the notes development.

³⁶ While the notes invariably write productions with premises and conclusion in a vertical array, a luxury we allowed ourselves in §1, to save space we henceforth usually use the present horizontal display.

II. " $\vdash \alpha O(X, a(F, P))$ ", " $\vdash \alpha[a(Y, F)]$ " produce " $\vdash \alpha O(X, a(F, b(Y, P)))$ ".³⁷

III. " $\vdash \alpha O(X, a(F, X))$ ", " $\vdash \alpha[a(F, P)]$ " produce " $\vdash \alpha O(X, a(F, b(X, P)))$ ".

Thus, in $O(x, Q)$, with Q of rank 0, Q can only be a simple functional expression involving the individual variable x . Substitution and (3) then take care of all simple functional expressions of one argument. In II, the first premise insures X being a variable in P with $a(F, P)$ a valid simple functional expression, and hence F a variable, the second premise insures Y being a variable distinct from F , whence the validity of the conclusion. In III, the first premise makes F and X distinct variables, the second, $a(F, P)$ a valid simple functional expression, whence again the validity of the conclusion. If a simple functional expression be built up by successively inserting its arguments from right to left, (3) or III enables us to introduce the $O(x, R)$ operation the first time x is thus introduced, while II enables us to keep that operation until the desired $O(x, Q)$ is obtained.

Under case (c), three cases must be considered: (c₁) Q and R both propositional variables, (c₂) Q a propositional variable, R a simple functional expression, (c₃) Q and R simple functional expressions. The case where Q is a simple functional expression, R a propositional variable, follows from (c₂) by XII below.

For (c₁) we need merely have

$$(4). \vdash \alpha(p_1 \vee p_2). \quad (5). \vdash \alpha(p \vee p).$$

For (c₂) the following suffices.

IV. " $\vdash \alpha[a(P, X)]$ ", " $\vdash \alpha[a(F, b(P, X))]$ " produce " $\vdash \alpha[P \vee a(F, X)]$ ".

For the first premise insures P being a variable distinct from any variable in X , the second insures F being a variable distinct from P and any variable in X , and at the same time guarantees a form for X such that $a(F, X)$ is a valid simple functional expression. The conclusion is then valid, and may clearly be any valid " $\vdash \alpha(Q \vee R)$ " with Q a propositional variable, R a simple functional expression.

Case (c₃) again subdivides into (c_{3'}), the functional expressions have different functional variables, (c_{3''}) the functional expressions have the same functional variable. (c_{3'}) is taken care of by

V. " $\vdash \alpha[a(F_1, X)]$ ", " $\vdash \alpha[a(F_2, X)]$ ", " $\vdash \alpha[a(F_1, Y)]$ ", " $\vdash \alpha[a(F_2, Y)]$ ", " $\vdash \alpha[a(F_1, F_2)]$ " produce " $\vdash \alpha[a(F_1, X) \vee a(F_2, Y)]$ ".

For the premises insure $a(F_1, X)$, $a(F_2, Y)$ being simple functional expressions with the consequent variables F_1 and F_2 distinct by the last premise, and distinct from the variables in both X and Y by the first four premises. On the other hand, for case (c_{3''}) the number of arguments must be the

³⁷ The notes incorrectly write the second premise " $\vdash \alpha a(f, y)$ ". They uniformly omit the parenthesis in $\alpha(P)$ when P is of the form $O(x, Q)$, a convention we follow here. For the sake of uniformity we have supplied their only occasional omission of the parenthesis when P is of the form $a(x, Q)$.

same in two functional expressions. Unlike our treatment of the preceding case we do not now rely on the earlier " $\vdash \alpha(P)$ "'s with P a simple functional expression, but iteratively build up the desired $\vdash \alpha(P \vee Q)$'s by the following.

$$(6). \vdash \alpha[a(f, x_1) \vee a(f, x_2)]. \quad (7). \vdash \alpha[a(f, x) \vee a(f, x)].$$

VI. " $\vdash \alpha[a(F, Y_1) \vee a(F, Y_2)]$ ", " $\vdash \alpha[a(F, b(X_1, Y_1))]$ ", " $\vdash \alpha[a(F, b(X_2, Y_2))]$ " produce " $\vdash \alpha[a(F, b(X_1, Y_1)) \vee a(F, b(X_2, Y_2))]$ ".

We can now carry through the above suggested proof, and completion of the basis for determining all valid " $\vdash \alpha(P)$ "'s. Via case (a) we have already done this for all P 's of rank 0. Assume that it has been carried through for all P 's of rank less than or equal to ρ , and let P be of rank $\rho + 1$. Since P is thus of rank greater than 0, it will be in one of the three forms $\sim Q$, $O(x, Q)$, $Q \vee R$.

In the first case, since Q is of rank ρ , we shall have " $\vdash \alpha(Q)$ " by our induction. The desired " $\vdash \alpha(P)$ " will then be obtained via

$$\text{VII. } \vdash \alpha(P) \text{ produces } \vdash \alpha(\sim P).$$

In the second case, when Q is of rank 0, the desired assertion has already been taken care of in (b) above. Otherwise, Q will be in one of the three forms $\vdash R$, $O(y, R)$, $R \vee S$. These three possibilities are covered by the following four operations.

$$\text{VIII. } \vdash \alpha O(X, P) \text{ produces } \vdash \alpha(X, \sim P).$$

$$\text{IX. } \vdash \alpha O(X, P), \vdash \alpha O(Y, P), \vdash \alpha[a(X, Y)] \text{ produce } \vdash \alpha O(X, O(Y, P)).$$

$$\text{X. } \vdash \alpha O(X, P), \vdash \alpha(P \vee Q) \text{ produce } \vdash \alpha O(X, P \vee Q).$$

$$\text{XI. } \vdash \alpha O(X, Q), \vdash \alpha(P \vee Q) \text{ produce } \vdash \alpha O(X, P \vee Q).^{38}$$

Note that when $Q = \sim R$, $P = O(x, \sim R)$. With P of rank $\rho + 1$, R is then of rank $\rho - 1$, and hence $O(x, R)$ is of rank ρ . We will thus have " $\vdash \alpha O(X, R)$ ", whence " $\vdash \alpha O(x, \sim R)$ " by VIII. Similar considerations of rank render our induction effective in the other cases. In the case $Q = O(y, R)$, with $P = O(x, Q)$, apparent variables x and y must be distinct. This is secured in IX by the third premise, the second premise having already insured Y being a variable. Operations X and XI together clearly take care of $Q = R \vee S$. That they are universally valid follows from our entire development, and our conditions on valid enunciations. Thus in X the second premise insures $P \vee Q$ being a valid enunciation, the first that X is a real variable in P . X is therefore a real variable in $P \vee Q$ and $O(X, P \vee Q)$ is a valid enunciation.

³⁸ Where we now have X and XI, the notes insufficiently have the one operation " $\vdash \alpha O(x, P)$ ", " $\vdash \alpha O(x, Q)$ ", " $\vdash \alpha(P \vee Q)$ " produce " $\vdash \alpha O(x, P \vee Q)$ ". This is not invalid, but assume x to be present in both P and Q .

We come then to the third case, $P = Q \vee R$. If Q and R are of unequal ranks, the operation

XII. " $\vdash \alpha(P \vee Q)$ " produces " $\vdash \alpha(Q \vee P)$ "

enables us to assume that the rank of R is greater than the rank of Q . Since R is then of rank greater than 0, it will be in one of the three forms $\sim S$, $O(x, S)$, $S \vee T$. These cases will then be covered by the following three operations respectively.

XIII. " $\vdash \alpha(P \vee Q)$ " produces " $\vdash \alpha(P \vee \sim S)$ "

XIV. " $\vdash \alpha(P \vee R)$ ", " $\vdash \alpha O(x, S)$ ", " $\vdash \alpha(X \vee P)$ " produce
" $\vdash \alpha(P \vee O(x, S))$ ".³⁹

XV. " $\vdash \alpha(P \vee Q)$ ", " $\vdash \alpha(P \vee R)$ ", " $\vdash \alpha(Q \vee R)$ " produce
" $\vdash \alpha(P \vee (Q \vee R))$ ".

The applicability of XIII and XV is immediate. Thus, in the case of $P = Q \vee \sim S$, by our hypothesis on rank, $\sim S$ is of rank ρ , Q of rank less than ρ . S is then of rank $\rho - 1$, $Q \vee S$ therefore of rank ρ , whence " $\vdash \alpha(Q \vee S)$ " by our induction, and so " $\vdash \alpha(Q \vee \sim S)$ " by XIII. Their universal validity is also readily demonstrated. Thus in XV any occurrence⁴⁰ of the same variable in P and $Q \vee R$ will be such an occurrence in P and Q , or P and R , and hence will be an occurrence of the same kind of variable by the first premise or the second premise; similarly for the other conditions.⁴¹

In applying XIV to the case $P = Q \vee O(x, S)$, the premises would become " $\vdash \alpha(Q \vee S)$ ", " $\vdash \alpha O(x, S)$ " and " $\vdash \alpha(x \vee Q)$ ". Since x cannot be present in Q , the third premise as well as the first two will be valid assertions provided the ranks agree with our induction. But since in the present case Q is at most of rank $\rho - 1$, while S is of rank $\rho - 1$, $Q \vee S$ and $O(x, S)$ will be of rank ρ , $x \vee Q$ of rank at most ρ , so that those premises are asserted in accordance with our induction, and " $\vdash \alpha[Q \vee O(x, S)]$ " follows. As for the validity of XIV, the second premise shows X to be an individual variable present in R . Hence, were it present in P , the first premise would make it appear in P as an individual variable, the third as a propositional variable; for the variable X explicitly thus appears in the first term of the latter disjunction. The simultaneous assertion of the three premises therefore insures X not being present in P , whence the validity of the conclusion.⁴²

⁴⁰ This word in its present use is foreign to our notes, though hardly the idea.

⁴¹ In fact, the notes give as the criterion for the validity of " $\vdash \alpha f(P_1, P_2, \dots, P_n)$ ", where f is built up by \sim 's and \vee 's, that " $\vdash \alpha(P_i)$ ", " $\vdash \alpha(P_i \vee P_j)$ " be valid for all i 's and j 's. The first set of conditions is not needed unless $n = 1$ and $i \neq j$, the second set was written " $\vdash \alpha(P_i, P_j)$ " by an obvious slip.

⁴² In the notes the standard method of insuring x being a variable not present in P is to have as premise " $\vdash \alpha(O(x, a(f, x)) \vee P)$ ". The use of this method will be continued later. But in the present instance it would complicate the

Finally, then, let $P = Q \vee R$ with Q and R of equal ranks. When those ranks are zero, " $\vdash \alpha(P)$ " is obtained by the above special case (c). Otherwise, both Q and R are of ranks greater than 0. By XII we may therefore assume P to be in one of the following six forms, $(\alpha) : \sim S V \sim T$, $(\beta) : \sim S V O(x, T)$, $(\gamma) : \sim S V (X \vee Y)$, $(\delta) : O(x, S) V O(y, T)$, $(\epsilon) : O(x, S) V (X \vee Y)$, $(\zeta) : (S V T) V (X \vee Y)$. The first three forms may be rewritten $\sim S V R$ with S and R of unequal ranks. Hence $\alpha(\sim S V R)$ will be asserted by the previous discussion, and so $\alpha(\sim S V R)$ by XII followed by XIII followed by XII. For (δ) , we may first assert $\alpha[O(x, S) V T]$ by the unequal rank discussion, $\alpha O(y, T)$ by our induction, $\alpha[y V O(x, S)]$ by the unequal rank discussion, and then apply XIV. In (ϵ) and (ζ) , which may be rewritten together as $Q V (X \vee Y)$, $\alpha(Q V X)$, $\alpha(Q V Y)$ may be assumed asserted by the unequal rank case, $\alpha(X \vee Y)$ by our induction, whence the results follow by XV.

Our first object is thus achieved. That is, in the above system in canonical form B , $\alpha(P)$ will be asserted when and only when P is the correspondent of an enunciation in *10 *Principia Mathematica*. Indeed, thus far these are the only assertions of the system.

We have observed that in a system in canonical form A , and hence in the (\sim, V) system of *1-*5 *Principia Mathematica*, the operation of substitution II need merely be iteratively applied to the primitive assertions IV of the system after which only the productions III need be applied. We therefore interpret the postulational set up of *10 in the latter fashion, and assume that after the most general substitutions have been allowed for in the formal primitive propositions of *10, including those of *1-*5, all other assertions are to be obtained through the remaining rules of operation of the system.⁴³

Due to our convention on apparent variables, if a propositional variable p occurs several times in an assertion, we cannot merely substitute the same enunciation P of *10 for each p , but must in each instance at least

induction, and require separate treatment of additional special cases. Actually, by the addition of two added primitive assertions, and one new production, the notes correctly obtain all assertions of the above form with P a matrix. They then incorrectly conclude that with the addition of one more operation, all assertions of the above form are obtainable, and end the development with the operation having the above as a premise, " $\vdash \alpha(P)$ " as conclusion, claiming that all valid " $\vdash \alpha(P)$ "s are thus obtainable. Had the induction proof then been carried through in detail, it would have been seen that to obtain " $\vdash \alpha(O(x, a(f, x)) V (Q V R))$ " with the help of XV, " $\vdash \alpha(Q V R)$ " must *first* be obtained. We may finally observe that the notes have two additional productions omitted here, one a special case of VIII, the other a production whose application is covered by XII and XV.

⁴³ At least that is all that the notes allow us to say; for they do not bring up the question of an operation of substitution in *10 *Principia Mathematica*, but merely in fact allow for arbitrary substitutions in the primitive propositions. A quite cursory examination of the question suggests that here too this is equivalent to allowing arbitrary substitutions at any stage of a deductive process.

rewrite P on a different set of apparent variables before making the substitution. To allow for this, we introduce a new primitive function $\beta(P, Q)$ into the system to which *10 is being reduced, and a set of postulates therefor, so that $\beta(P, Q)$ will be asserted when and only when P and Q are correspondents of enunciations of *10 obtainable from each other by a mere interchange of apparent variables, and such that, in fact, $\alpha(P \vee Q)$ is asserted. We shall also have to introduce a primitive function $\sigma(x, y, P, Q)$ whose assertion is to mean that Q is obtained from P by replacing real individual variable x of P by a variable y not present in P , or present in P as a real individual variable distinct from x , and changing apparent variables so that $\alpha(P \vee Q)$ is asserted. Considerable economy is gained by giving a simultaneous development for the resulting common basis for these two primitives.⁴⁴

When P is of rank 0, it can have no apparent variables. We shall then have $\beta(P, Q)$ when and only when Q is identical with P . This case is then taken care of for $\beta(P, Q)$ by the following.

$$\vdash \beta(p, p); \quad \vdash \alpha(a(F, P)) \text{ produces } \vdash \beta(a, (F, P), a(F, P)).$$

In the case of $\sigma(x, y, P, Q)$, with P of rank 0, P can only be a simple functional expression. However an iterative build up is now needed to allow for an arbitrary number of arguments.

$$\begin{aligned} & \vdash \sigma(x_1, x_2, a(f, x_1), a(f, x_2)). \\ & \vdash \alpha(O(X, a(F_1, X)) \vee a(F, b(Y, P))) \\ & \quad \text{produces } \vdash \sigma(X, Y, a(F, b(X, P)), a(F, b(Y, P))). \\ & \vdash \sigma(X, Y, a(F, P), a(F, Q)) \\ & \quad \text{produces } \vdash \sigma(X, Y, a(F, b(X, P)), a(F, b(Y, Q))). \\ & \vdash \sigma(X, Y, a(F, P), a(F, Q)), \vdash \alpha(O(X, a(F, X)) \vee O(Z, a(F, Z))) \\ & \quad \text{produce } \vdash \sigma(X, Y, a(F, b(Z, P)), a(F, b(Z, Q))). \end{aligned}$$

Note that the primitive assertion, and first production, allow for the initial introduction of the x as the simple functional expression is built up right to left, the second production allows for any later x , the third for any later variable other than x , including y .⁴⁵

We may now assume $\beta(P, Q)$ and $\sigma(x, y, P, Q)$ to have been taken care of for P of rank less than or equal to ρ , and let P be of rank $\rho + 1$. The

⁴⁴ In the notes, on the other hand, the basis for $\beta(P, Q)$ is the first set up, and then, with its help, the basis for $\sigma(x, y, P, Q)$ is given. This is not done with complete accuracy, and further footnotes will point out the principal corrections thus necessitated in our present combined basis. The notes here cease numbering the primitive assertions and productions, and also omit quotation marks in the latter. We have therefore done the same.

⁴⁵ The notes do not give the first production, and in the second premise of the third production have the added term $O(Y, a(F, Z))$, here written with capitals, thus preventing Z from being Y .

following three productions then take care of $\beta(P, Q)$ according as $P = \sim R$, $P = O(z, R)$, $P = R \vee S$.⁴⁶

- $\vdash \beta(P, Q)$ produces $\vdash \beta(\sim P, \sim Q)$.
- $\vdash \sigma(X, Y, P, Q), \vdash \alpha(O(Y, a(F, Y)) \vee P)$
produce $\vdash \beta(O(X, P), O(Y, Q))$.
- $\vdash \beta(P, Q), \vdash \beta(R, S), \vdash \alpha(P \vee .R : Q \dots \vee .S)$
produce $\vdash \beta(P \vee R, Q \vee S)$.

Note that the second premise of the second production serves the purpose of excluding Y from P .

The production for $\sigma(x, y, P, Q)$ are more complicated. When $P = O(z, R)$, then $Q = O(w, S)$, so that S is obtained from R by replacing the two letters x and z by y and w respectively. This is then reduced to the σ operation by introducing a form intermediate between P and Q . On the other hand, when $P = R \vee S$, we need three productions, since we must explicitly allow for x being in R and S , in R only, in S only. The concluding productions are then the following.⁴⁷

- $\vdash \sigma(X, Y, P, Q)$ produces $\vdash \sigma(X, Y, \sim P, \sim Q)$.
- $\vdash \sigma(Z, Z', P, Q), \vdash \sigma(X, Y, Q, R), \vdash \alpha(O(Z, P) \vee O(Z', R))$,
 $\vdash \alpha(a(Z', Y))$ produce $\vdash \sigma(X, Y, O(Z, P), O(Z', R))$.
- $\vdash \sigma(X, Y, P, Q), \vdash \sigma(X, Y, R, S), \vdash \alpha(P \vee R. \vee .Q \vee S)$
produce $\vdash \sigma(X, Y, P \vee R, Q \vee S)$.
- $\vdash \sigma(X, Y, P, Q), \vdash \beta(R, S), \vdash \alpha(O(X, a(F, X) \vee R))$,
 $\vdash \alpha(P \vee R. \vee .Q \vee S)$ produce $\vdash \sigma(X, Y, P \vee R, Q \vee S)$.
- $\vdash \beta(P, Q), \vdash \sigma(X, Y, R, S), \vdash \alpha(O(X, a(F, X)) \vee P)$,
 $\vdash \alpha(P \vee R. \vee .Q \vee S)$ produce $\vdash \sigma(X, Y, P \vee R, Q \vee S)$.

The five formal primitive propositions of *1-*5 carried over to *10 under all possible substitutions of *10 are easily taken care of via the β function. Each will be replaced by a production which may be described as follows. In the given primitive proposition of *1-*5 replace all of the small letters occurring therein by different capital letters for all different posi-

⁴⁶ The second of these productions is new because of our fusion of the two bases; but the several productions replacing it in our notes find their counterparts in the later productions having σ in the conclusion. At this point the dot notation creeps into our notes intermingled with the parenthesis notation. Since we have retained the notation $p \vee q$, we also carry along the dots.

⁴⁷ The second of these five productions is incorrectly given in the notes, the need of an intermediary between P and R being overlooked; but the corresponding situation in the independently given β basis was correctly handled. The third production has its third premise inadequately written $\alpha(P \vee R)$. The need for the fourth and fifth productions was completely overlooked, both here, and in the corresponding situation for the β basis.

tions of the small letters. For each distinct small letter, if P_1, P_2, \dots, P_ν are the corresponding capital letters replacing it, introduce the premises $\vdash \beta(P_i, P_{i+1}), i = 1, \dots, \nu - 1$ in the production. If F is the primitive proposition in terms of capital letters thus all different, add the premise $\vdash \alpha(F)$. The production is then completed by adding "produce $\vdash F$ ". Note that the necessary $\vdash \alpha(F)$ itself imposes the conditions of distinctness of variables as required by the β -premises, while the latter otherwise merely insure "essentially the same" enunciations of *10, i.e., apart from the particular apparent variables used, always being substituted for the same small letter of the primitive proposition for all the occurrences of that small letter in the proposition.

The one operation of *1-*5 other than substitution, already allowed for, is directly taken care of by adding the production

$$\vdash P, \vdash \sim P \vee Q \text{ produce } \vdash Q$$

to our system.⁴⁸

The remaining formally significant primitive propositions of *10 *Principia Mathematica* as there given are the following.

*10.1. $\vdash: (x).\phi x. \supset .\phi y.$

*10.11. If ϕy is true whatever possible argument y may be, then $(x).\phi x$ is true.

*10.12. $\vdash: .(x).p \vee \phi x. \supset: p \vee .(x).\phi x.$ ⁴⁹

Of these, the first and third are primitive assertions of the system, the second a rule of operation. However, when subjected to all possible substitutions, the effect of all three will appear in our new system as operations. The latter are almost self-explanatory, and in order are the following.

$\vdash \sigma(X, Y, P, Q)$ produces $\vdash \sim O(X, P) \vee Q.$

$\vdash P, \vdash \alpha O(X, P)$ produce $\vdash O(X, P).$

$\vdash \beta(P, R), \vdash \sigma(X, Z, Q, S), \vdash \alpha[\sim O(X, P \vee Q) \vee: R \vee O(Z, S)]$
produce $\vdash \sim O(X, P \vee Q) \vee: R \vee O(Z, S).$

It is readily seen that in the first operation, the hypothesis insures the conclusion being a valid enunciation, so that no corresponding premise is needed.⁵⁰ In the second operation, the second premise insures X being a

⁴⁸ Strangely overlooked in the notes. See also the next footnote. It is readily verified that thanks to the operation of substitution of canonical form B , we may thus have the same P and the same Q occurring twice, instead of using different letters connected by the β relation.

⁴⁹ Primitive propositions *10.121, *10.122, and *11.07 are of a different nature, and are automatically taken care of by our use of canonical form B .

⁵⁰ The notes do have a second premise $\alpha O(x, P)$, itself changed from a first written α of the conclusion.

real individual variable of P , as desired. The complicated third operation is required by our convention on apparent variables. That is, P in its second occurrence must appear as an R with other apparent variables, Q as an S with not only other apparent variables than those of Q , but also with X changed to a Z , as X and Z are apparent variables of the conclusion. This time, unlike the situation for the first operation, the resulting two premises do not insure the conclusion being a valid enunciation, whence the third premise.

The desired reduction of *10 *Principia Mathematica* to a system in canonical form B is thus completed.

3. The problem of "tag"

The direction taken by the reductions of the next two sections will become clearer if we at least formulate a problem, christened "tag" by B. P. Gill, which has played a vital part in the present development. An early unpublished method of the writer completely solved the problem of determining for any two expressions in the (\sim, \vee) system of *Principia Mathematica*, or indeed in any system in canonical form A , what substitutions would make those expressions identical. Because of the form of the result, this method was termed the L. C. M. process. In passing from the (\sim, \vee) system to the whole of *Principia Mathematica*, attention was first centered on the "matrices" thus arising, and the problem arose of determining the substitutions on the variable propositional functions occurring therein which would make two such forms identical. The general problem proving intractable, successive simplifications thereof were considered, one of the last being this problem of "tag." Again, after the finiteness problem for systems in canonical form A involving primitive functions of only one argument was solved, an attempt to solve the problem for systems going, it seemed, but a little beyond this one argument case, led once more essentially to the selfsame problem of "tag." The solution of this problem thus appeared as a vital stepping stone in any further progress to be made.

In its first form the problem may be stated as follows. Given, a positive integer ν , and μ symbols which may be taken to be $0, 1, \dots, \mu - 1$. With each of these μ symbols a finite sequence of these symbols is associated as follows.

$$\begin{array}{lll} 0 & \rightarrow & a_{0,1}a_{0,2}\dots a_{0,\nu_0} \\ 1 & \rightarrow & a_{1,1}a_{1,2}\dots a_{1,\nu_1} \\ \dots & \dots & \dots \\ \mu - 1 & \rightarrow & a_{\mu-1,1}a_{\mu-1,2}\dots a_{\mu-1,\nu_{\mu-1}}. \end{array}$$

It is understood that in each sequence the same symbol may occur several times, and that a particular associated sequence may be null. In terms of this basis, we set up the following operation for obtaining from any given

non-null sequence

$$B = b_1 b_2 \dots b_\ell.$$

on the symbols $0, 1, \dots, \mu - 1$, a unique derived sequence B' on those symbols. To the right end of B adjoin the sequence associated with the symbol b_1 in the given basis, and from the left end of this augmented sequence remove the first ν elements — all if there be less than ν elements. As long as B' is not a null sequence, this operation can then be applied to B' to yield a sequence B'' , to B'' , if not null, to yield B''' , and so on. The problem of "tag" for the given basis is then to obtain a finite process for determining for any initial sequence B whether the resulting iterative process does or does not terminate.⁵¹

In the second form of the problem, the one that arose in connection with the finiteness problem, the initial sequence B may be considered part of the basis, and the problem would be to obtain a finite process for determining of any given sequence $c_1 c_2 \dots c_m$ on the μ symbols $0, 1, \dots, \mu - 1$ whether that sequence is or is not generated in the course of the above iteration of the given process, starting with B . Clearly, for this second form of the problem, the problem for a given basis is immediately solvable if the process is known to terminate. Where the process does not terminate, it is readily seen that according as the lengths of the resulting sequences are bounded, or unbounded, the resulting infinite sequence of the sequences will, from some point on, become periodic, or the length of the n -th sequence will increase indefinitely with n . In the first case the second form of the problem is again immediately solvable, while in the second case the solution would follow if a method were also found for determining of any given length of sequence a point in the process beyond which all derived sequences were of length greater than that given length.⁵²

The first form of the problem, emended to determine whether the iterative process was terminating, periodic, or divergent, thus seemed likely to cover both forms. In this emended form the problem of "tag" was made the major project of the writer's tenure of a Procter fellowship in mathematics at Princeton during the academic year 1920–21. Indeed, the reduction of the last section, effected early in that academic year, sealed this determination. And the major success of that project was the complete solution of the problem for all bases in which μ and ν were both 2.⁵³

⁵¹ In an early formulation of this problem, b_1 was first checked off, the corresponding associated sequence added, then the ν -th element after b_1 was checked off, corresponding associated sequences added, and so on. Whether the iterative process terminated or not then depended on whether the ever advancing check mark did or did not overtake the usually advancing right end of the sequence, whence the suggestive name proposed by Gill for the problem.

⁵² In this analysis we may have gone somewhat further than is justified by the notes.

⁵³ When either μ or ν is 1 the problem becomes trivial. By contrast, even this special case $\mu = \nu = 2$ involved considerable labor.

While considerable effort was expended on the case $\mu = 2, \nu > 2$, but little real progress resulted, such a simple basis as $0 \rightarrow 00, 1 \rightarrow 1101, \nu = 3$, proving intractable.⁵⁴ For a while the case $\nu = 2, \mu > 2$, seemed to be more promising, since it seemed to offer a greater chance of a finely graded series of problems. But when this possibility was explored in the early summer of 1921, it rather led to an overwhelming confusion of classes of cases, with the solution of the corresponding problem depending more and more on problems of ordinary number theory. Since it had been our hope that the known difficulties of number theory would, as it were, be dissolved in the particularities of this more primitive form of mathematics, the solution of the general problem of "tag" appeared hopeless, and with it our entire program of the solution of finiteness problems.

This frustration, however, was largely based on the assumption that "tag" was but a very minor, if essential, stepping stone in this wider program. In the late summer of 1921, however, the reductions carried through at Princeton in proving the equivalence of canonical forms A and B suggested a further transformation of canonical form B , and, indeed, led to a whole series of reductions with the final canonical form very close to the seemingly special form of the "tag." As these reductions are vital in the further evolution of our thought, we turn to them in the next two sections.

4. Reduction of canonical form B to a canonical form C

In our canonical form B , as well as A , on the one hand, we allow primitive functions of many variables, and thus rely on the parenthesis notation; on the other hand, we assume the availability of an infinite number of variables. We show in the present section that canonical form B , to be specific, can be reduced to a canonical form C where the boxes within a box symbolic form of the parenthesis notation is replaced merely by finite sequences of letters, and where, for a given system, the different letters so used constitute a once and for all given finite set.⁵⁵

⁵⁴ Note of course that an arbitrary initial sequence has to be allowed for. Numerous initial sequences actually tried led in each case to termination or periodicity, usually the latter. It might be noted that an easily derived "probability" prognostication suggested that in this case periodicity was to be expected.

⁵⁵ This of course is a characteristic of tag; but also, essentially, of systems in canonical form A involving only primitive functions of one argument. For if $f_1(p), \dots, f_\mu(P)$ are the primitive functions of such a system, an arbitrary enunciation thereof is in the form $f_{i_1}(f_{i_2}(\dots f_{i_m}(q) \dots))$, which may then as well be written $f_{i_1}f_{i_2} \dots f_{i_m}q$. Except for the one arbitrary variable q the enunciations are then just sequences of the primitive letters f_1, f_2, \dots, f_μ . Furthermore, each production is in the form, $g_1P_{j_1}, g_2P_{j_2}, \dots, g_\kappa P_{j_\kappa}$ produce gP_j , where the g 's represent fixed sequences of the primitive f 's. In the "homogeneous case," to which the more general case can be reduced, the sole operational variable of the

The basis of an arbitrary system in canonical form C is to be of the following form. There are a finite number of distinct primitive symbols a_1, a_2, \dots, a_μ . The "enunciations" of the system are simply all finite sequences of such symbols, repetitions of the same symbol being of course allowed. That is, an arbitrary enunciation of the system may be written

$$a_{i_1} a_{i_2} \dots a_{i_n}$$

with n arbitrary, the a_{i_j} 's arbitrarily a_1, a_2, \dots, a_μ . A specific finite set of such enunciations is set down to constitute the "primitive assertions" of the system.⁵⁶ Furthermore, a specific finite set of "productions" of the following form is set down to yield new assertions from old:

$$\begin{aligned} & g_{11} P_{i_1^1} g_{12} P_{i_2^1} \dots g_{1m_1} P_{i_{m_1}^1} g_{1(m_1+1)} \\ & g_{21} P_{i_1^2} g_{22} P_{i_2^2} \dots g_{2m_2} P_{i_{m_2}^2} g_{2(m_2+1)} \\ & \dots \dots \dots \\ & g_{k1} P_{i_1^k} g_{k2} P_{i_2^k} \dots g_{km_k} P_{i_{m_k}^k} g_{k(m_k+1)} \\ & \text{produce} \\ & g_1 P_{i_1} g_2 P_{i_2} \dots g_m P_{i_m} g_{m+1} \end{aligned}$$

where the g 's are specified sequences of the primitive a 's, including the null sequence, and each P of the conclusion is present in at least one premise. In the application of these productions the P 's may be identified with arbitrary sequences of the above type, it being understood however, that the conclusion may not be null.⁵⁷ The assertions of the system are then the

conclusion is also the operational variable of each premise. There then turns out to be no loss of generality in assuming all enunciations to be written with the same propositional variable. This may then be deleted, leaving only sequences of f 's.

⁵⁶ The notes here give up using an assertion sign, a practice we feel constrained to follow in keeping our account in the spirit of the notes. In connection with the about to be described productions, note that if A, B, \dots, E represent the sequences $a_1 a_2 \dots a_\ell, b_1 b_2 \dots b_m, \dots e_1 e_2 \dots e_p$ respectively, then $AB \dots E$ simply represents the sequence $a_1 a_2 \dots a_\ell b_1 b_2 \dots b_m \dots e_1 e_2 \dots e_p$.

⁵⁷ The application of these productions is not quite as automatic as in the case of the productions of canonical form B ; for in the latter a given assertion can be written in the form of a given premise in one and only one way, if at all. In the present case such uniqueness is achieved only under a particular hypothesis on the ranks of the operational variables occurring in the premise. While less would suffice, actually, since the sum of the ranks of the g 's and P 's in a given premise must be equal to the rank of the corresponding given assertion, rank now being the total number of letters in a sequence, but a finite number of such hypotheses are possible, and all can uniformly be tried out. Indeed, the successive reductions of this and the next section successively analyze away most of what is nonautomatic in the present as well as in the earlier canonical forms. In this connection see footnote 78.

enunciations obtainable by the repeated application of these operations to the primitive assertions, and all assertions so obtainable.

Given a system S_1 in canonical form B , we proceed to build up the basis of a system S_2 in canonical form C to which S_1 will be reducible. For precision, let the variables of S_1 be the infinite set $p_1, p_2, \dots, p_n, \dots$. Introduce a primitive letter a_0 in S_2 . Then let the above variables in order correspond to the enunciations, i.e., sequences, $a_0, a_0 a_0, \dots, a_0 a_0 \dots a_0, \dots$ of S_2 there being n a_0 's in the n -th sequence. To distinguish these enunciations of S_2 from others, we introduce a primitive letter α_0 , and suitable postulates therefor, to be part of the basis of S_2 , so that $\alpha_0 P$ will be asserted when and only when P is a finite sequence of a_0 's. The following postulates clearly suffice.

$$\alpha_0 a_0; \quad \alpha_0 P \text{ produces } \alpha_0 a_0 P.$$

Of course, it must be seen to that further postulates do not produce other assertions of the form $\alpha_0 P$ than the above.

Let the primitive functions of S_1 be $f_i(p_1, p_2, \dots, p_{m_i})$, $i = 1, 2, \dots, \mu$. Correspondingly introduce primitive letters a_i in S_2 . We shall assume that the parenthesis notation of S_1 has been replaced by an equivalent dot notation. Corresponding to an enunciation of the form $f_i(P_1, P_2, \dots, P_{m_i})$, the corresponding dot notation is to be $f_i \dots P_1 \dots P_2 \dots \dots \dots P_{m_i}$ where the same number of dots separate f_i and P_1 and each pair of consecutive P 's, and where that number is to be one more than the largest number of dots in any P . Introduce then a primitive letter b in S_2 to correspond to the dot of S_1 so rephrased. The enunciations of S_2 corresponding to the enunciations of S_1 will be certain sequences of a_i 's, $i = 0, 1, 2, \dots, \mu$, and b 's. Before these are singled out, we introduce a primitive letter β such that the assertion of $Q\beta P$ in S_2 is to mean that P is a sequence involving no other letters than $a_0, a_1, \dots, a_\mu, b$, that Q is a sequence of b 's, and that the number of b 's in Q is more than that of the largest uninterrupted sequence of b 's in P . This result is secured by adding the following postulates to S_2 , each postulate involving a_i being duplicated for $i = 0, 1, 2, \dots, \mu$.

$$\begin{aligned} & b\beta a_i, \quad bb\beta b; \\ & Q\beta P \text{ produces } Q\beta a_i P, \\ & Q_1\beta Q_2, Q_2\beta P \text{ produce } bQ_1\beta bQ_2, \\ & Q_1\beta Q_2 a_i P, bbQ_2\beta Q_1 \text{ produce } bQ_1\beta bQ_2 a_i P, \\ & Q_1\beta Q_2 a_i P, bbbQ_2 Q_3 \beta Q_1 \text{ produce } Q_1\beta bQ_2 a_i P. \end{aligned}$$

To prove this, we first show that from valid $Q\beta P$'s only valid $Q\beta P$'s result. Our first production merely annexes an a_i to the left of P in a valid $Q\beta P$, hence doesn't disturb the condition on the b 's, and so yields a valid $Q\beta P$. In the second production, the second premise guarantees that Q_2 consists of b 's only, P having the effect of an apparent variable. The two premises then

show Q_1 to be the sequence of b 's with one more b than Q_2 . But then the same is true of bQ_1 and bQ_2 . The second premise of the third production, along with the first, makes Q_2 a sequence of b 's one less than Q_1 , while the largest sequence of b 's in P is at least one less than the number of b 's in Q_1 . Hence in bQ_2a_iP the largest sequence of b 's is the sequence bQ_2 , whence the conclusion. On the other hand, in the fourth production, the second premise, with Q_3 playing the rôle of apparent variable, definitely insures Q_2 being a sequence of b 's at least two less than Q_1 , itself a sequence of b 's by the first premise. Hence, by the first premise, the largest sequence of b 's in P is exactly one less than Q_1 . The largest sequence of b 's in bQ_2a_iP is then again exactly one less than Q_1 , whence the conclusion. This analysis makes easy the verification that, actually, all valid $Q\beta P$'s result. Those where P consists of but one letter are given by the primitive assertions themselves. Assuming all those with P having n letters to have been found, those with $n+1$ letters can be found from them by annexing an a_i to the left of P with Q unchanged, or annexing b to the left of P , perhaps changing Q to bQ . The first effect is secured by the first operation. As for the second effect, when P has no a 's, it is secured by the second operation, where P has an a , and starts with a "maximal sequence" of b 's, the third operation secures the desired result, while if P has an a , but does not start with a maximal number of b 's, the fourth operation applies.⁵⁸

Now introduce a primitive letter α in system S_2 so that αP will be asserted when and only when P represents an enunciation in S_2 corresponding to one in S_1 . The following postulates, added to our growing basis for S_2 , are easily seen to produce this effect, if not later disturbed.⁵⁹

$$\alpha_0 P \text{ produces } \alpha P$$

$$Q\beta P_1 P_2 \dots P_{m_i}, \alpha P_1, \alpha P_2, \dots, \alpha P_{m_i} \\ \text{produce } \alpha a_i Q P_1 Q P_2 \dots Q P_{m_i}, \quad i = 1, 2, \dots, \mu.$$

Note that where P is a valid translation of an enunciation in S_1 , it can neither start nor end with b . Hence the largest sequence of b 's in $P_1 P_2 \dots P_{m_i}$ is the largest such sequence occurring in the several P 's. Our first production lays down the correspondents of variables as valid correspondents; and having the valid correspondents of all enunciations of rank ρ of S_1 , the μ cases of the second production yield the valid correspondents of all enunciations of S_1 of rank $\rho+1$.

As a result of the above, if the variables of S_1 are p_1, p_2, p_3, \dots , and they are set in 1-1 correspondence with the sequences $a_0, a_0 a_0, a_0 a_0 a_0, \dots$, a 1-1 correspondence is set up between the enunciations of S_1 , and those

⁵⁸ The notes do not give this proof, but merely point out the two places where an operational variable acts as a substitute for an apparent variable. However, the productions speak for themselves; we have merely pointed out what they say. Due to an earlier use of the letter c in what later became b , the third and fourth productions are written in the notes with c in place of b .

⁵⁹ We again allow a variable of S_1 to be an enunciation thereof.

enumerations P of S_2 for which αP is an assertion.

We turn now to the assertions of S_1 . As in the case of canonical form A it can be shown for canonical form B that every assertion in S_1 can be obtained by first obtaining a set of assertions from IV by the sole use of II' and then, starting with these assertions, merely employing III'. The set of all assertions obtainable from IV by the repeated use of II' are all assertions $h_i(p_{n_1}, p_{n_2}, \dots, p_{n_{\ell_i}})$, $i = 1, 2, \dots, \lambda$, with $n_1, n_2, \dots, n_{\ell_i}$ arbitrary distinct positive integers. We first then show how to get correspondents in S_2 of all such assertions in S_1 .⁶⁰

For this purpose introduce a new primitive letter δ such that $P\delta Q$ will be asserted when and only when P and Q are both sequences of a_0 's, but of unequal lengths. This result is clearly secured by the productions

$$\begin{aligned}\alpha_0 P &\text{ produces } a_0 \delta a_0 P, \\ \alpha_0 P &\text{ produces } a_0 P \delta a_0, \\ P \delta Q &\text{ produces } a_0 P \delta a_0 Q.\end{aligned}$$

Now our original primitive assertions in IV of S_1 are the λ particular enumerations $h_i(p_1, p_2, \dots, p_{\ell_i})$, $i = 1, 2, \dots, \lambda$, each with the assertion sign " \vdash " prefixed to it. By the method described above, for each of these enumerations there is a corresponding enunciation in S_2 . In these enunciations replace each sequence of j a_0 's, i.e., those arising indeed from p_j , by P_j , and symbolize the λ corresponding expressions by $\bar{h}_i(P_1, P_2, \dots, P_{\ell_i})$, $i = 1, 2, \dots, \lambda$. Actually, then, $\bar{h}_i(P_1, P_2, \dots, P_{\ell_i})$ is a specific sequence of letters $a_1, \dots, a_\mu, b, P_1, P_2, \dots, P_{\ell_i}$. Now introduce the symbol " \vdash " as a new primitive letter in S_2 so that, ultimately, " $\vdash P$ " will be asserted in S_2 when and only when P is the correspondent of an enunciations P' of S_1 with " $\vdash P'$ " an assertion in S_1 . For the above assertions in S_1 of the form " $\vdash h_i(p_{n_1}, p_{n_2}, \dots, p_{n_{\ell_i}})$ " this result is then achieved through the productions⁶¹

$$\begin{aligned}P_{j_1} \delta P_{j_2}, \quad j_1, j_2 = 1, 2, \dots, \ell_i, \quad j_1 < j_2, \\ \text{produce } \vdash \bar{h}_i(P_1, P_2, \dots, P_{\ell_i}), \quad i = 1, 2, \dots, \lambda.\end{aligned}$$

There remains then but the reproducing of the effect of III' of S_1 in S_2 .⁶² We shall merely describe the κ productions thus corresponding

⁶⁰ This part is new, but a more complicated process for achieving the same effect appears in the earlier portions of the notes, referred to in connection with §1, which prove the reducibility of canonical form B to A . We introduced the change since the notes show how to reduce canonical form A , not B , to C .

⁶¹ Where $m_i = 1$ the sole premise of the production would simply be $\alpha_0(P_1)$.

⁶² The notes here again are radically emended, but this time unnecessarily so. Except for minor slips, easily corrected, the notes method is correct, but that correctness eluded us until the final revision of the present account. Where we

to the κ productions of S_1 , and to be added to, and indeed, completing the basis of S_2 . If P represents any enunciation in S_1 , \tilde{P} the corresponding enunciation in S_2 , then the rank of P actually is the largest number of b 's in a sequence of consecutive b 's in \tilde{P} . If then an enunciation in S_1 is of the form $f_i(P_1, P_2, \dots, P_{m_i})$, f_i one of the primitive functions of S_1 , and Q_1, Q_2, \dots, Q_{m_i} are sequences of b 's in number equal to the rank of P_1, P_2, \dots, P_{m_i} respectively, a sequence of b 's Q , in number equal to the rank of $f_i(P_1, P_2, \dots, P_{m_i})$ will be the unique Q for which say $Q\beta Q_1 a_0 Q_2 a_0 \dots Q_{m_i}$ is asserted in S_2 . For we recall that the rank of $f_i(P_1, \dots, P_{m_i})$ is one more than the largest of the ranks of P_1, \dots, P_{m_i} . It follows that if $\tilde{P}_1, \tilde{P}_2, \dots, \tilde{P}_{m_i}$ represent the correspondents in S_2 of P_1, P_2, \dots, P_{m_i} respectively, then the correspondent of $f_i(P_1, P_2, \dots, P_{m_i})$ will be $a_i Q \tilde{P}_1 Q \tilde{P}_2 Q \dots \tilde{P}_{m_i}$. Now any expression $g(P_1, P_2, \dots, P_n)$ such that $g(p_1, p_2, \dots, p_n)$ is an enunciation in S_1 other than an unmodified variable can be written in one and only one way $f_i(R_1, R_2, \dots, R_{m_i})$ with f_i a primitive function of S_1 , the R_j 's being P 's, or similar expressions, involving some, if not all, of the P 's.⁶³ We may then inductively define the constituents of such an expression as $g(P_1, P_2, \dots, P_n)$ as itself and the constituents of R_1, R_2, \dots, R_{m_i} , any P being its own only constituent.⁶⁴ Consider any production in III' of S_1 and let its operational variables be P_1, P_2, \dots, P_k . Form the distinct constituents of the several premises and conclusion. They are clearly finite in number, and hence may be ordered in a finite sequence starting with P_1, P_2, \dots, P_k . For the j -th constituent in this sequence introduce the operational variable Q_j which is to represent a sequence of b 's equal in number to the unknown rank of that constituent. If then we introduce operational variables $\tilde{P}_1, \tilde{P}_2, \dots, \tilde{P}_k$ to represent the correspondents of P_1, P_2, \dots, P_k respectively, we can successively build up expressions representing the correspondents of all of the above constituents solely by means of the \tilde{P} 's and Q 's in the manner described above. The desired production in S_2 to take the place of the given one of S_1 may then be described as follows. Its conclusion will be the above built up correspondent of the given conclusion preceded by \vdash . Its premises will first include the correspondents of the given premises each preceded by \vdash . Sec-

now explicitly allow for the rank of each constituent of the symbolic expressions occurring in an operation, the notes merely consider the ranks of the operational variables thereof, and determine the ranks of the constituents by added hypotheses on the former ranks. That the finite number of sets of hypotheses given in the notes thus suffice was not seen by us while writing the account, but now can be said to have been obviously clear when the notes were written. Since the present method is considerably simpler, we leave it in the account.

⁶³ This is really a tacit postulate on symbolic expressions in a propositional calculus.

⁶⁴ Note that we here talk of the symbolic constituents of the symbolic expression $g(P_1, P_2, \dots, P_n)$ with P 's variable, whereas ranks mean the ranks of the functions of p_1, p_2, \dots thus represented.

ondly, they will include the k expressions $\alpha\tilde{P}_1, \alpha\tilde{P}_2, \dots, \alpha\tilde{P}_k$.⁶⁵ Finally, they will include a set of premises giving the conditions on the Q 's. These will first include the k premises $bQ_j\beta\tilde{P}_j, j = 1, 2, \dots, k$. Furthermore, for each constituent R_j other than a P_j , and corresponding unique expression $f_i(R_{j_1}, R_{j_2}, \dots, R_{j_{m_i}})$ therefor in terms of other constituents, we shall have the premise $Q_j\beta Q_{j_1}a_0Q_{j_2}a_0\dots Q_{j_{m_i}}$. It is then readily verified that if to each premise of the given production of S_1 is arbitrarily made to correspond an enunciation of S_1 , and to the \vdash prefixed corresponding premise of the production of S_2 is made to correspond the corresponding enunciation of S_2 , then being able to replace the P 's in the first production so that the premises become the corresponding enunciations of S_1 is equivalent to being able to replace the \tilde{P} 's and Q 's in the second so that the premises beginning with a \vdash become the corresponding enunciations of S_2 with \vdash prefixed, while the remaining premises become assertions in S_2 . Furthermore, that in the favorable case the P 's on the one hand, \tilde{P} 's and Q 's on the other, are thus uniquely determined, and with them the conclusions, which are then enunciations of S_1 , and of S_2 with \vdash prefixed, that correspond.

With the basis of S_2 , clearly in canonical form C , thus completed, it follows that an enunciation of S_1 is an assertion when and only when the corresponding enunciation of S_2 with \vdash prefixed is an assertion, so that S_1 has thus been reduced to S_2 . Note that we originally spoke of a system in canonical form C as having a finite number of distinct primitive symbols a_1, a_2, \dots, a_μ . For our above S_2 , with different μ , these symbols are $a_0, a_1, \dots, a_\mu, b, \alpha_0, \alpha, \vdash, \beta, \delta$. Furthermore, each assertion of S_2 involves one and only one of the last five symbols, and but one occurrence of that one symbol.⁶⁶

5. Successive reduction to normal form

Starting with canonical form C , we now introduce a series of reductions such that each formulation, while being included in the preceding, eliminates some formal complexity allowed in that preceding formulation. For a given system this simplification is achieved at the expense of an increase in the number of primitive letters employed, and in the number of productions constituting its basis.

Our first reduction of an arbitrary system in canonical form C is to one in which there is but one primitive assertion, and in which each production involves but a single premise, that one premise and corresponding

⁶⁵ These premises are probably unnecessary. That is, the succeeding conditions on the Q 's, coupled with the inductive result that if $\vdash P$ is asserted, αP is asserted, probably insure their satisfaction when the remaining premises are satisfied.

⁶⁶ Actually, for those enunciations of S_2 not in the form " $\vdash P$ ", the deducibility problem is immediately solvable.

conclusion, however, retaining all of the complexity allowed for above.⁶⁷ The general plan of the method involved is to formally allow for the logical product of arbitrary assertions in the given system, and operate within such products.

Let then S_1 be a system in canonical form C with primitive letters a_1, a_2, \dots, a_μ , S_2 the about to be described system to which S_1 is to be reduced. With a_1, a_2, \dots, a_μ also primitive letters of S_2 , introduce two new primitive letters u and a_0 in S_2 .⁶⁸ When the logical product of assertions, $P_1, P_2, P_3, \dots, P_n$ of S_1 is asserted in S_2 , it will appear in the form

$$ua_0P_1a_0uuua_0P_2a_0uuua_0P_3a_0\dots\underbrace{\dots u}_{n}a_0P_na_0\underbrace{\dots u}_{n+1},$$

each P being flanked on either side by a_0 . The separating u sequences are thus made to increase left to right by one each to enable us by the mere form of a premise to insure that certain operational variables therein must represent assertions of S_1 , if that premise is to be identified with an assertion in S_2 . The final basis for S_2 will reveal the necessary source of that insurance, i.e., that the only assertions of S_2 involving u are those of the above form. We shall call such an expression a product, the P 's therein the factors of the product.

We first introduce in the basis of S_2 certain productions whereby from the assertion of a product may be obtained the assertion of all products obtainable from the given product by a mere permutation of its factors. It suffices to allow for the interchange of any two consecutive factors. For the first two factors of a product this is achieved by

$$ua_0P_1a_0uuua_0P_2a_0uuua_0S \text{ produces } ua_0P_2a_0uuua_0P_1a_0uuua_0S,$$

our system being so devised that each product appearing therein has at least three factors. This allows that last a_0 to be assumed. The u, uu, uuu

⁶⁷ That one can do with one primitive assertion both here and in all latter reductions was not observed in the notes during the course of the corresponding development. However, in a later reference to the final formulation of this section in work corresponding to part II of this account specific reference is made to the one primitive assertion in a way that suggests that at some in between time this further simplification was noted. Likewise in the notes of 1924.

⁶⁸ The introduction of a_0 is new. That by its use the notes method is made neater is not our reason for its introduction. The notes consistently and incorrectly assume that an enunciation P_1 can always be written a_iP_2 where a_i is the first letter of the enunciation represented by P_1 . But this does not allow for P_1 being null; and since for P_1 consisting of one letter, P_2 would be null, the net effect of all the reductions of this section, if carried through exactly as in the notes, would be to impose on the P 's of a canonical form C the condition that their ranks exceed a certain fairly large number, thus at least vitiating the reduction of §4. This oversight is responsible for most of the changes introduced in the present section, changes which on the whole are minor.

of the premise are then "maximal" u sequences. As these u sequences differ by one each, P_1 and P_2 must be free from u 's, and hence, by our induction, be the two initial factors of the product. The interchange then results via the production. For two consecutive factors neither starting nor ending the product, the result is achieved by

$$Ra_0uQua_0P_1a_0uQuua_0P_2a_0uQuuuua_0S \\ \text{produces}$$

$$Ra_0uQua_0P_2a_0uQuua_0P_1a_0uQuuuua_0S.$$

Here Q must consist of u 's only. For otherwise a_0uQua_0 and a_0uQuua_0 would have their initial a_0 's followed by identical maximal u sequences. The u sequences uQu , $uQuu$ and $uQuuu$ are then maximal, and differ in length by one each. P_1 and P_2 again are consecutive factors of the product. Finally, for two factors ending a product the last production, rewritten with a_0S deleted, suffices.⁶⁹

The next production to be added to the bases of S_2 allows us to pass from the assertion of a product to the assertion of the first factor of a product, and hence, with the help of the previous three productions, to the assertion of an arbitrary factor of a product. The production is simply

$$ua_0Pa_0uua_0R \text{ produces } P.$$

In translating the operations of S_1 into operations within products of S_2 , we allow for passing from a product whose initial factors can be identified with the premises of an S_1 operation, to that product with the conclusion of the S_1 operation as additional factor. That additional factor must end the new product so as not to disturb the progression of the maximal u sequences. Let " G_1, G_2, \dots, G_k produce G " represent any one of the S_1 operations. Let H represent

$$ua_0G_1a_0uua_0G_2a_0 \dots \underbrace{u \dots u}_{k} a_0G_k a_0 \underbrace{u \dots u}_{k+1}.$$

Then the corresponding S_2 operation may be presented by⁷⁰

$$Ha_0Ra_0uQua_0Sa_0uQuu \text{ produces} \\ Ha_0Ra_0uQua_0Sa_0uQuua_0Ga_0uQuuu.$$

Note that the operational variables of this production are those of the S_1 production, and Q, R, S . Since each operational variable in G occurs in at least one of the G_i 's, our new production will indeed have the same

⁶⁹ We might observe that the notes go to considerable length in showing why the u method is thus effective.

⁷⁰ This production, as given in the notes, involves some slight errors.

operational variables in its conclusion as in its premise. The portion of the premise following H insures Q consisting of u 's only. This, with the form of H , insures G_1, G_2, \dots, G_k , being determined factors of the premise, G of the conclusion. Hence, the validity of our transformation of the S_1 production. The additional operational variables R and S require an assertion to which this production is applied to have at least $k+2$ factors, a requirement secured below. Of course, the basis of S_2 is to have the correspondent of each of the operations in the basis of S_1 .

With S_1 having κ productions, the above $\kappa+4$ productions constitute all of the productions in the basis of S_2 . Its sole primitive assertion is then formed as follows. Let L be the largest number of premises occurring in any production of S_1 . If S_1 has λ primitive assertions, let each be repeated L times to give λL sequences each involving no other letters than a_1, \dots, a_μ . If $\lambda L < L+2$, or $\lambda L < 3$, again duplicate one of these sequences the one or two times needed to avoid these inequalities. If then k_1, k_2, \dots, k_M are these duplicated primitive assertions of S_1 , the primitive assertion of S_1 will be their product⁷¹

$$ua_0k_1a_0uuua_0k_2a_0\dots u\underset{M}{\underbrace{\dots u}}a_0k_Ma_0\underset{M+1}{\underbrace{\dots u}}.$$

Now it is readily proved by induction that if at a certain point of the process for obtaining assertions in S_1 a certain finite set of assertions has been obtained, then there will be asserted in S_2 a product among whose factors are each of the above assertions repeated L times. For the primitive assertions of S_1 , this is insured by the primitive assertion of S_2 . Assume it to be true for the deductive process in S_1 at an arbitrary point, let P_2 be the corresponding assertion in S_2 , P_1 the next assertion obtained in S_1 , $P_{11}, P_{12}, \dots, P_{1k}$ the premises of the production of S_1 yielding conclusion P_1 . Then each P_{1j} appears as factor of P_2 indeed L times at least. Hence, from P_2 , by the first three productions of S_2 , an assertion P_2' can be obtained in which the first k factors are $P_{11}, P_{12}, \dots, P_{1k}$ respectively, whatever repetitions may occur among those P 's. The production of S_2 corresponding to the one of S_1 in question will then add P_1 as a factor to P_2' . Mere repetition of the application of this production will then yield P_2'' , which will be P_2' with L additional factors equal to P_1 . The induction is thus established. It follows that for each assertion P_1 in S_1 there will be an assertion P_2 in S_2 having P_1 as a factor. By the first three productions of S_2 this factor can be made the first factor of an assertion in S_2 , and hence, by the fourth production of S_2 , P_1 itself will be an assertion of S_2 . That is, every assertion of S_1 is an assertion of S_2 . Our basis for S_2 shows that the only other assertions of S_2 are products of assertions of S_1 , and

⁷¹ Instead of this single primitive assertion, the notes require the assertion of the products of $L+1$ of the h 's, repeated or not for all such choices of h 's, thus, for the moment, overlooking the obvious simplification to one primitive assertion.

so not wholly written on the letters of S_1 . Hence, an enunciation of S_1 is an assertion of S_1 when and only when it is an assertion of S_2 , whence the reduction of S_1 to S_2 .

In our second reduction of canonical form C the productions, all with single premises by the previous reduction, now take the more special form

$$g_1 P_1 g_2 P_2 \dots g_m P_m g_{m+1}$$

produces

$$\bar{g}_1 P_1 \bar{g}_2 P_2 \dots \bar{g}_m P_m \bar{g}_{m+1}$$

where, however, m , and of course the g 's, may vary from operation to operation. By contrast, in the previous productions P 's could be repeated, have different arrangements in premise and conclusion, and in part be missing from the conclusion while present in the premise.

Again, let the primitive letters of the given system be symbolized a_1, a_1, \dots, a_μ . Let its i -th production be

$$g_1 P_1 g_2 P_2 \dots g_m P_m g_{m+1}$$

produces

$$\tilde{g}_1 P_{j_1} \tilde{g}_2 P_{j_2} \dots \tilde{g}_{\tilde{m}} P_{j_{\tilde{m}}} \tilde{g}_{\tilde{m}+1}$$

where it is understood that each letter except P has i for additional subscript. The subscripts of the P 's need not be distinct in premise or conclusion, while the different subscripts of the P 's in the conclusion all appear in the premise. However, the letter P occurs exactly $m + \tilde{m}$ times in the production.

We introduce a new primitive letter u , and for each such production two new primitive letters v_i, w_i . In obtaining the effect of the i -th production we will, as above, leave this subscript i understood. v_i will be used in passing from an assertion involving a 's only that could be the premise of the i -th production to one which has both that premise and corresponding conclusion recognizable within it; w_i in passing from such a composite assertion to the desired conclusion only. The efficacy of our method will depend on each assertion in the new system which involves v or w having that letter only at the beginning of the assertion, and in the first case always involving exactly $2m + \tilde{m}$ u 's, in the second, \tilde{m} u 's. Our new productions will in every case explicitly exhibit this v and $2m + \tilde{m}$ u 's, or w and \tilde{m} u 's, so that we can be sure that in their application the operational variables can represent sequences of a 's only. Except for a minor preliminary type, all of our "v-assertions" will be in the form

$$v u g_1 P_1 u Q_1 u g_2 P_2 u Q_2 \dots u g_m P_m u Q_m g_{m+1} u \tilde{g}_1 Q_{m+1} u \tilde{g}_2 Q_{m+2} \dots u \tilde{g}_{\tilde{m}} Q_{m+\tilde{m}} \tilde{g}_{\tilde{m}+1},$$

and when so asserted will have the following properties. The sequence of a 's $g_1 P_1 Q_1 g_2 P_2 Q_2 \dots g_m P_m Q_m g_{m+1}$ is an assertion of the given, and indeed

new system, while the sequences of a 's $Q_1, Q_2, \dots, Q_m, Q_{m+1}, \dots, Q_{m+\tilde{m}}$ can, in order, be identified with $P_{j_1}, P_{j_2}, \dots, P_{j_m}, P_{\tilde{j}_1}, \dots, P_{\tilde{j}_{\tilde{m}}}$, that is, any two Q 's corresponding to P 's with identical subscripts are equal. Note that with all $2m + \tilde{m}$ u 's exhibited, the g 's being given, the P 's and Q 's of such an assertion are uniquely identifiable in the assertion. Our method depends on the fact that when such an assertion is obtained in which the P 's are null, then, due to the equalities forced on the Q 's, $g_1 Q_1 g_2 Q_2 \dots g_m Q_m g_{m+1}$ becomes an assertion on a 's only that can be identified with the premise of the i -th production of the given system, and hence $\tilde{g}_1 Q_{m+1} \tilde{g}_2 Q_{m+2} \dots \tilde{g}_{\tilde{m}} Q_{m+\tilde{m}} \tilde{g}_{\tilde{m}+1}$ an expression on a 's only that will be the corresponding conclusion. Of course, each production about to be described is directly seen to be in the desired newly simplified form.

Since a null assertion has been excluded from our systems, each assertion of the given system is of the form $a_j P$, $j = 1, 2, \dots, \mu$. The productions

$$a_j P \text{ produces } va_j Pu \dots u$$

with $2m + \tilde{m}$ u 's in $u \dots u$ changes each "a-assertion," i.e., assertion involving a 's only, into what we shall call the intermediate v form. As all other assertions of our new system will begin with v or w , these productions will be inapplicable to them. If now an a-assertion can be the premise of the i -th production, its intermediate v form will be put into primary v form, or just v form, by the production

$$vg_1 P_1 g_2 P_2 \dots g_m P_m g_{m+1} u \dots u$$

produces

$$vug_1 P_1 uug_2 P_2 uu \dots g_m P_m ug_{m+1} u \tilde{g}_1 u \tilde{g}_2 u \dots u \tilde{g}_{\tilde{m}} \tilde{g}_{\tilde{m}+1}.$$

Of course this production may be applicable without the P 's being identifiable with those of the premise of the i -th production. But, comparing this conclusion with our general v form, we see that it satisfies the requirement thereof with all Q 's null. Now any set of a -sequences that could be identified with the $P_{j_1}, P_{j_2}, \dots, P_{j_m}, P_{\tilde{j}_1}, \dots, P_{\tilde{j}_{\tilde{m}}}$, of the i -th production can be built up as follows. Start with the set of null sequences. Let $Q_1, Q_2, \dots, Q_m, Q_{m+1}, \dots, Q_{m+\tilde{m}}$ be any such derived set of a -sequences. Let $Q_{j_1}, Q_{j_2}, \dots, Q_{j_\nu}$, j 's increasing, be any subset thereof corresponding to all P 's with subscripts equal to a given subscript, a_j any one of the primitive a 's. Then $\dots, a_j Q_{j_1}, \dots, a_j Q_{j_2}, \dots, a_j Q_{j_\nu}, \dots$, all other Q 's unchanged, will also be such a set of a -sequences. Rewrite the subscript sequence j_1, j_2, \dots, j_ν in the form $j_1, \dots, j_\lambda, j_{\lambda+1}, \dots, j_\nu$ so that $j_\lambda m, j_{\lambda+1} > m$, and let $j_{\lambda+1} - m = \tilde{j}_1, \dots, j_\nu - m = \tilde{j}_{\lambda}$. Of course we may have $\lambda = \nu$. Now for each such choice of original P subscript, and each a_j , introduce the production

$$v \dots ug_{j_1} P_{j_1} a_j u Q_{j_1} \dots ug_\lambda P_{j_\lambda} a_j u Q_{j_\lambda} \dots u \tilde{g}_{j_1} Q_{j_{\lambda+1}} \dots u \tilde{g}_{j_\lambda} Q_{j_\nu} \dots$$

produces

$$v \dots ug_{j_1} P_{j_1} ua_j Q_{j_1} \dots ug_\lambda P_{j_\lambda} ua_j Q_{j_\lambda} \dots u \tilde{g}_{j_1} a_j Q_{j_{\lambda+1}} \dots u \tilde{g}_{j_\lambda} a_j Q_{j_\nu} \dots$$

all of the rest of both premise and conclusion being as in the type v form above. Such a production will then change a valid v form into a valid v form, the effect being however to "drain" the P 's of such a form and "swell" the Q 's. If then an assertion of the given system can be put in the form of the premise of the i -th production, the corresponding intermediate v form will pass into a v form such that successive application of the above productions will completely drain the P 's thereof; and, indeed, conversely. This marks the end of the first half of the passage from a -assertion to a -assertion in the new system. While the second half could be set up by means of similar w productions in reverse, with interchange of emphasis on premise and conclusion of the i -th production,⁷² the following method is simpler. With P 's all null, the v form determines the desired a -conclusion as described above. The about to be introduced w forms each have exactly \tilde{m} u 's all explicitly appearing in the productions. From such a v form with P 's all null the first w form is obtained via

$$\begin{aligned} vug_1uQ_1ug_2uQ_2 \dots ug_m uQ_m g_{m+1} u\tilde{g}_1 Q_{m+1} u\tilde{g}_2 Q_{m+2} \dots u\tilde{g}_{\tilde{m}} Q_{m+\tilde{m}} \tilde{g}_{\tilde{m}+1} \\ \text{produces} \\ wg_1Q_1g_2Q_2 \dots g_m Q_m g_{m+1} u\tilde{g}_1 Q_{m+1} u\tilde{g}_2 Q_{m+2} \dots u\tilde{g}_{\tilde{m}} Q_{m+\tilde{m}} \tilde{g}_{\tilde{m}+1}. \end{aligned}$$

We can now get rid of the no longer interesting part of this w form, i.e., the part between w and the first u thereof, by the μ productions

$$\begin{aligned} wa_j P u\tilde{g}_1 P_1 u\tilde{g}_2 P_2 \dots u\tilde{g}_{\tilde{m}} P_{\tilde{m}} \tilde{g}_{\tilde{m}+1} \\ \text{produces} \\ wP u\tilde{g}_1 P_1 u\tilde{g}_2 P_2 \dots u\tilde{g}_{\tilde{m}} P_{\tilde{m}} \tilde{g}_{\tilde{m}+1} \end{aligned}$$

iteratively applied till letter by letter what was the original a -assertion disappears. The desired a -conclusion then would be obtained via

$$\begin{aligned} wu\tilde{g}_1 P_1 u\tilde{g}_2 P_2 \dots u\tilde{g}_{\tilde{m}} P_{\tilde{m}} \tilde{g}_{\tilde{m}+1} \\ \text{produces} \\ \tilde{g}_1 P_1 \tilde{g}_2 P_2 \dots \tilde{g}_{\tilde{m}} P_{\tilde{m}} \tilde{g}_{\tilde{m}+1}. \end{aligned}$$

Our final system will then be on the primitive letters $a_1, \dots, a_\mu, u, v_1, w_1, v_2, w_2, \dots, v_\kappa, w_\kappa$, κ being the number of productions of the given system. The one primitive assertion of the new system will be the one primitive assertion of the given system, the productions of the new system, all of the above productions for each of the κ productions of the given system. Our above analysis then easily shows that the assertions of the new system involving no other letters than a_1, \dots, a_μ are exactly the assertions of the given system, and the desired reduction has been effected.

⁷² This is the method of the notes. In the previous work one obvious error, discovered at the time, is corrected as directed by the notes, and a few minor changes are introduced.

Our third and penultimate simplifying reduction of canonical form C is to one where the operations are of the form

$$\begin{aligned} g_1 P g_2 \\ \text{produces} \\ \bar{g}_1 P \bar{g}_2, \end{aligned}$$

i.e., involve but a single operational variable. Again let a system in the previous simplified form have primitive letters a_1, a_2, \dots, a_μ , and κ operations, the number of P 's in the premise, and hence conclusion, of the i -th operation being m_i . For the i -th operation, with $i = 1, 2, \dots, \kappa$, and each primitive letter a_j we introduce $2m_i + 1$ new primitive letters $a_{ji}^1, a_{ji}^2, \dots, a_{ji}^{2m_i}, a_{ji}^{2m_i+1}$.⁷³ We also introduce the primitive letter a_{0i} along with $a_{0i}^1, a_{0i}^2, \dots, a_{0i}^{2m_i}, a_{0i}^{2m_i+1}$.⁷⁴ With one such operation in mind at a time we will as above omit the extra subscript i . Apart from the use of a_0 and a_j^j 's, needed to take care of g 's or P 's that are null, the essence of our method is to pass from an a -assertion in the form $g_1 P_1 g_2 P_2 \dots g_m P_m g_{m+1}$ to an assertion $g_1^1 g_2^3 \dots g_{m+1}^{2m+1} P_1^2 P_2^4 \dots P_m^{2m}$ where the superscript k say indicates that each a_j in the corresponding expression is here written a_j^k . As a result our premise will now have the form gP with

$$g = g_1^1 g_2^3 \dots g_{m+1}^{2m+1}, \quad P = P_1^2 P_2^4 \dots P_m^{2m}.$$

In detail, we first introduce μ productions

$$a_j P \quad \text{produces} \quad a_0 a_j P, \quad j = 1, 2, \dots, \mu,$$

which will be applicable in fact only to assertions on a_1, a_2, \dots, a_μ , and changes any such assertion Q into $a_0 Q$. We then introduce a finite series of finite sets of productions depending in number on m and μ . The first set has the one production

$$a_0 g_1 P \quad \text{produces} \quad a_0^1 g_1^1 a_0 P a_0^2.$$

Inductively let the conclusion of the sole production in the $(2k - 1)$ -st set be in the form $G_k a_0 P a_0^{2k}$. Then the $(2k)$ -th set has the μ productions

$$G_k a_0 a_j P \quad \text{produces} \quad G_k a_0 P a_j^{2k}, \quad j = 1, 2, \dots, \mu,$$

the $(2k + 1)$ -st set the sole production

$$G_k a_0 g_{k+1} P \quad \text{produces} \quad G_k a_0^{2k+1} g_{k+1}^{2k+1} a_0 P a_0^{2k+2}.$$

⁷³ The notes observe that we could be more economical if desired.

⁷⁴ This is new, and is necessary for the reason discussed in footnote 68. In fact, here the notes rely on the g 's as well as P 's not being null.

This is to hold for $1k < m$, while for $k = m$ the sole production of the $(2m + 1)$ -st set is to be

$$G_m a_0 g_{m+1} a_0^2 P \text{ produces } G_m a_0^{2m+1} g_{m+1}^{2m+1} a_0^2 P.$$

We then readily see that starting with an assertion on a_1, \dots, a_μ in the form $g_1 P_1 g_2 P_2 \dots g_m P_m g_{m+1}$, one can, with the aid of these productions, obtain as an assertion

$$a_0^1 g_1^1 a_0^3 g_2^3 \dots a_0^{2m-1} g_m^{2m-1} a_0^{2m+1} g_{m+1}^{2m+1} a_0^2 P_1^2 a_0^4 P_2^4 \dots a_0^{2m} P_m^{2m}.$$

Furthermore, note that starting with an assertion on a_1, a_2, \dots, a_μ , flanked on the left by a_0 as above, one can apply the above operations only in the following order, if at all. First, the sole operation of the first set; and inductively, if the operation in the $(2k - 1)$ -st set has last been applied, the next applicable operation can only be an operation in the $2k$ -th set or the operation in the $(2k + 1)$ -st set, if an operation in the $(2k)$ -th set has last been applied, the next applicable operation can only be an operation in the same set, or the operation in the next set. Furthermore, the last operation in its premise explicitly indicates the a_0^2 , first introduced into an assertion only as a result of the first operation. It readily follows that if the last operation does enter into a possible sequence of operations, the conclusion thereof can have no letter a_j in it without a superscript. The entire given assertion has thus been translated; and it is readily seen that the last assertion, and hence given assertion, are and can be put in the forms above given.

The actual correspondent of the original i -th operation in translated form may then be written simply

$$\begin{aligned} a_0^1 g_1^1 a_0^3 g_2^3 \dots a_0^{2m-1} g_m^{2m-1} a_0^{2m+1} g_{m+1}^{2m+1} P \\ \text{produces} \\ a_0^1 \bar{g}_1^1 a_0^3 \bar{g}_2^3 \dots a_0^{2m-1} \bar{g}_m^{2m-1} a_0^{2m+1} \bar{g}_{m+1}^{2m+1} P; \end{aligned}$$

and the passage from this translated conclusion to the actual conclusion can be effected by a set of productions the reverse of those above given. That is, in each of the above productions prior to the actual correspondent of the i -th production replace all g_j 's by \bar{g}_j 's and *interchange hypothesis and conclusion*. The resulting productions then clearly suffice to yield the conclusion yielded in the original i -th production. True, the complete set of productions thus set up to take the place of the original i -th production may now allow other paths than from assertion on a_1, \dots, a_μ , down the first group of productions, through the intermediate production, and up the second group of productions to new assertion on a_1, \dots, a_μ .⁷⁵ But it

⁷⁵ This could have been avoided by the v, w method used earlier.

is readily seen that any departures from this progression merely constitute unravelings of parts of such progression, or, apart from such unravelings, constitute shortcuts of valid full progressions of this type. Since, furthermore, one can change the set of productions one is working with only when an assertion on a_1, \dots, a_μ alone is obtained, the validity of our reduction follows.

Our final reduction is to a system whose operations are in the form

$$\begin{aligned} gP \\ \text{produces} \\ Pg' \end{aligned}$$

The present method assumes that in the productions of the previous system, all in the form

$$\begin{aligned} g_1Pg_2 \\ \text{produces} \\ g_1'Pg_2' \end{aligned}$$

g_1 and g_2 are never null. We therefore actually first need the following preliminary reduction.⁷⁶ Introduce a new primitive letter a_0 , and if h is the sole primitive assertion of the given system let a_0ha_0 be the sole primitive assertion of the new system. Replace each of the above operations of the given system by

$$a_0g_1Pg_2a_0 \text{ produces } a_0g_1'Pg_2'a_0$$

and finally add the production

$$a_0Pa_0 \text{ produces } P.$$

Except for the last production the new system may be said to be simply isomorphic⁷⁷ with the old, P being an assertion in the given system when and only when a_0Pa_0 is an assertion in the new system. The last operation then merely recovers the assertions of the given system. Note that even that last operation is in the desired form with neither g_1 nor g_2 null.

Assume then that such is our given system with primitive letters again a_1, a_2, \dots, a_μ . We introduce new primitive letters $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_\mu$, and "translating productions"

$$a_jP \text{ produces } P\bar{a}_j, \quad \bar{a}_jP \text{ produces } Pa_j, \quad j = 1, 2, \dots, \mu.$$

⁷⁶ This is new. The notes erroneously state that as a consequence of the previous reductions P must in fact represent sequences having at least two letters, and on that basis easily show how to replace such productions with g_1, g_2 null by an equivalent set with neither g_1 nor g_2 null.

⁷⁷ This concept was introduced in the notes in connection with the equivalence proof referred to in §1.

Starting with an assertion of the form $a_{i_1} \dots a_{i_j} a_{i_{j+1}} \dots a_{i_n}$ these productions will yield only assertions of the form $a_{i_{j+1}} \dots a_{i_n} \bar{a}_{i_1} \dots \bar{a}_{i_j}, \bar{a}_{i_1} \dots \bar{a}_{i_j} \bar{a}_{i_{j+1}} \dots \bar{a}_{i_n}, \bar{a}_{i_{j+1}} \dots \bar{a}_{i_n} a_{i_1} \dots a_{i_j}$, in addition to the original assertion. Only one of these $2n$ distinct forms consists wholly of unbarred letters, i.e., the original form, while continued application of the above operations merely keeps deriving these $2n$ "equivalent forms" cyclically, so that anyone can thus be obtained from any other.

Our reduction will then be effected if for each operation " $g_1 P g_2$ produces $g_1' P g_2'$ " of the given system we introduce in the new system the operation

$$\begin{aligned} & \bar{g}_2 g_1 P \\ & \text{produces} \\ & P g_2' \bar{g}_1', \end{aligned}$$

where \bar{g}_2 , for example, is g_2 with each letter replaced by the corresponding barred letter. Of course, the one primitive assertion of the given system is also the one primitive assertion of the new system. Note that if at any point an assertion without barred letters appears, then if it can be written $g_1 P g_2$, the first given translating operations derive from it $\bar{g}_2 g_1 P$, hence the above yields $P g_2' \bar{g}_1'$, and so finally $g_1' P g_2'$ is obtained as desired. That is, the new system contains all of the assertions of the given system. It further follows that the assertions of the new system consist only of the assertions of the given system and their equivalents. For, proceeding inductively, this clearly remains true under the translating operations. Now suppose it is true of an assertion in the form $\bar{g}_2 g_1 P$. Since \bar{g}_2 and g_1 are not null, $\bar{g}_2 g_1$ alone exhibits a change from barred to unbarred letters. P therefore must consist of unbarred letters only. $\bar{g}_2 g_1 P$ is therefore a translation of the assertion $g_1 P g_2$ of the original system, and hence the conclusion $P g_2' \bar{g}_1'$ is a translation of $g_1' P g_2'$, also an assertion of the original system. The desired reduction has thus been effected.

A system of the last given type will be said to be in *normal form*, any system reducible to such a system, *reducible to normal form*, whence the heading of the section just concluded.⁷⁸

⁷⁸ Note that in the case of systems in normal form, as also for those in the form immediately preceding, that unique identifiability of operational variables which was possessed by systems in canonical form B , and lost in C , is regained. This, of course, is due to the presence of but one operational variable in each production. Indeed, the formal mechanism for obtaining new assertions from old is far simpler for a system in normal form than for canonical forms A or B . And if it be desired to change its set of assertions from, as it were, a growing population to a uniquely determined infinite sequence of assertions, but the following procedure need be instituted. Order the productions of the system. Start with the primitive assertion as 'given assertion,' the first operation as 'given operation,' the primitive assertion as sole member of the 'given sequence of assertions,' and iterate as follows. Apply, if possible, the given operation to the given assertion as premise and add the resulting conclusion to the given sequence of assertions, thus forming

6. Closing the circle

Apart from the reduction of *10 *Principia Mathematica* to canonical form *B*, our work has consisted of a series of reductions starting with canonical form *A* and ending with systems in normal form. Using $S_1 \rightarrow S_2$ to symbolize S_1 is reducible to S_2 , our definition of reducibility clearly yields the result, if $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_3$ then $S_1 \rightarrow S_3$. If then we can show that systems in normal form are reducible to canonical form *A*, it will follow that all of our canonical forms from first to last are equivalent, that is, for any two of them any system of either is reducible to some system of the other. Stated otherwise, the solution of the finiteness problem for all systems in one canonical form would yield the solution of the finiteness problem for all systems in the other.⁷⁹

We shall perform our reduction in two stages, first to a certain type of system in canonical form *B*, and then just this type of system to canonical form *A*. Corresponding to the primitive letters a_1, a_2, \dots, a_μ of the given system in normal form, we introduce primitive first order functions $a_1(p), a_2(p), \dots, a_\mu(p)$. We cannot quite let an arbitrary enunciation $a_{i_1} a_{i_2} \dots a_{i_n}$ of the given system correspond to $a_{i_1}(p) a_{i_2}(p) \dots a_{i_n}(p)$ in the sense of a formal product of *n* factors. But corresponding to the way in which an ordinary product of *n* elements is secured by a dyadic operation and associative law thereon, we get the desired effect by introducing a primitive second order function $b(p, q)$. The correspondent of $f = a_{i_1} a_{i_2} \dots a_{i_n}$

the new given sequence of assertions. If the given operation is not the last, retain the given assertion as given assertion, and pass to the next operation as given operation. If the given operation is the last, pass from the given assertion to the next assertion, and to the first operation as given operation.

⁷⁹ This idea of closing the circle of reductions by reducing systems in normal form to canonical form *A* does not explicitly appear in the notes. It, however, was probably considered to be one of those things that could obviously be done. While, as a result, the first of the two stages in which this reduction is carried out is "apocryphal," the second of those two stages merely uses a method which in more general form was used in the notes to reduce the canonical form intermediate between *A* and *B* to *A*. In fact, to somewhere preserve this method is our excuse for introducing this section.

As to *10 being merely attached to this circle, the unpublished note referred to in footnote 18, categorically states that a proof of the reducibility of canonical form *A* to *10 "is nearly completed," and as a result even suggests that the solution of the finiteness problem for *10 would yield the solution of the finiteness problem for all of *Principia Mathematica*. An examination of the notes reveals the reduction to have been carried through in three stages: first the reduction to *10 of mathematical systems using *10 as logic, second the reduction to the latter of what were termed "algebraic systems," third the reduction of systems in canonical form *A* to algebraic systems. It is for the second reduction that one half of the necessary two-way proof was postponed. Not having checked the parts of the proof that were given, we cannot guarantee their correctness. In this connection, see footnote 90.

of the given system will then be

$$b(a_{i_1}(p), b(a_{i_2}(p), \dots, b(a_{i_{n-1}}(p), a_{i_n}(p)) \dots)),$$

which we will symbolize by $f(p)$ for purposes of discussion. Of course, when $n = 1$ we simply have $f(p) = a_{i_1}(p)$.

Corresponding to the above mentioned associative law we introduce the operations

$$b(P, b(Q, R)) \text{ produces } b(b(P, Q), R),$$

$$b(b(P, Q), R) \text{ produces } b(P, b(Q, R)),$$

$$b(b(P, Q), b(R, S)) \text{ produces } b(P, b(b(Q, R), S)),$$

which have the effect of enabling us to pass from any mode of inserting b -parentheses in the sequence $a_{i_1}(p), a_{i_2}(p), \dots, a_{i_n}(p)$ to any other mode. All of the resulting forms may then be considered correspondents of f , with $f(p)$ the principal correspondent.

If h symbolizes the one primitive assertion of the given system, our new system will have the primitive assertion $h(p)$. The operations " $g_i P$ produces $P g_i'$ " will be suitably taken care of if we allow for the passage from some correspondent of the hypothesis to some correspondent of the conclusion. Note that for g_i, g_i', P not null, a correspondent of $g_i P$ will be in the form $b(g_i(p), Q)$, of $P g_i'$, $b(Q, g_i'(p))$. We then correspondingly introduce the operation

$$b(g_i(P), Q) \text{ produces } b(Q, g_i'(P)).$$

In fact, since the primitive assertion of the new system involves but the one variable p , all assertions of the system may be considered to be written on the one variable p , since the operation of substitution of canonical form B will merely reproduce these assertions on other variables. In the application of the last given operation, P , in fact, will only be identifiable with that variable p .

To allow for P null in the operation of the given system, the separate operation

$$g_i(P) \text{ produces } g_i'(P)$$

must be added. If either g_i or g_i' is null, we may clearly assume the other not null, while P then certainly can't be null, since the null assertion has been specifically excluded from our systems. When g_i' is null we then need but the one operation

$$b(g_i(P), Q) \text{ produces } Q.$$

When g_i is null we must explicitly insure our conclusion being written on but a single variable. For P of more than one letter, we may write $P = a_j P'$, $j = 1, 2, \dots, \mu$ and correspondingly set up the μ operations

$$b(a_j(P), Q) \text{ produces } b(b(a_j(P), Q), g_i'(P)).$$

For P of one letter we likewise have

$$a_j(P) \text{ produces } b(a_j(P), g_i'(P)), \quad j = 1, 2, \dots, \mu.$$

This completely takes care of the first reduction.

For the second reduction introduce a new primitive first order function $k(p)$ and alter the preceding system as follows. Replace the primitive assertion $h(p)$ by $h(k(p))$, retain the preceding productions, and change the operation of substitution to that of canonical form A . Now it is readily proved by induction that every assertion of the resulting system will be of the form $F(k(P))$ where $F(p)$ does not involve the primitive function k . Furthermore, if $F(k(P)) = F'(k(P'))$, with $F(p)$ and $F'(p)$ not involving k , we see by successively stripping away necessarily identical outmost primitive functions that $F = F'$, $P = P'$, i.e., that such a form is unique.⁸⁰ Again by induction, if $F_1(k(P_1)), F_2(k(P_2)), \dots, F_n(k(P_n))$ are the successive assertions in an arbitrary proof of our system, we find that P_i is either identical with P_{i-1} , or obtainable from it by a substitution. It follows that the only assertions of our system of the form $F(k(p))$ are obtained by deductive processes in which no other substitutions are employed than that of variable for variable. But with substitution thus limited, the new system is simply isomorphic with the old under the replacement $p \leftrightarrow k(p)$, p an arbitrary variable for completeness. It follows that an enunciation $F(p)$ of the previous system is an assertion thereof when and only when $F(k(p))$ is an assertion of the new system, whence our second reduction.

For our original system we then have that $a_{i_1} a_{i_2} \dots a_{i_n}$ is an assertion thereof when and only when

$$b(a_{i_1}(k(p)), b(a_{i_2}(k(p)), \dots, b(a_{i_{n-1}}(k(p)), a_{i_n}(k(p))) \dots))$$

is an assertion in the system in canonical form A , as desired.

We have observed in §3 how the seemingly simple problem of "tag" in fact proved intractable for $\mu = 2, \nu > 2$, of bewildering complexity for $\mu > 2, \nu = 2$. In view of our reduction of canonical form A to a form as close to that of "tag" as the normal form, the difficulty of "tag" is no longer surprising.⁸¹ While this suggested that special cases of "tag" might well be worth consideration as major problems in themselves, the following further reduction of the normal form seemed more promising.

We merely state the result. Given any system in normal form on letters a_1, a_2, \dots, a_μ , its assertions will be the assertions, on those letters,

⁸⁰ This is a special aspect of the L.C.M. process referred to early in §3.

⁸¹ In fact, at one point late in the work on "tag," it seemed that the regularity induced by always removing μ elements from the beginning of a sequence was responsible for the intrusion of number theory in the development, so that it was tentatively suggested that "tag" be generalized to a form which, indeed, is exactly that of the later derived normal form.

in a system on letters $a_1, a_2, \dots, a_\mu, \tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_{\tilde{\mu}}$, having a finite number of primitive assertions, and a finite number of operations of the following form

$$\begin{array}{cccc}
 gP & Pg & g_1 P & Pg_1 \\
 \text{produces} & \text{produces} & g_2 P & Pg_2 \\
 g'P & Pg' & \text{produce} & \text{produce} \\
 & & g'P & Pg'.
 \end{array}$$

While this reduction seems to undo much of the simplification that was achieved by our previous reduction, especially in that it allows productions with more than one premise, the fact that in each production the g 's all occur on the same side of the operational variable makes the formulation analogous to canonical form *A* with primitive functions of one argument only, and the finiteness problem for that case was solved! In fact, when all the productions have the g 's on the same side, the resulting system is essentially in this special canonical *A* form, and is actually reducible thereto by the "*k* method" last given. By further study, the resulting solution of the finiteness problem was extended to those of the present systems in which first order operations only occur, i.e., operations with but one premise, and then indeed to those systems in which only the second order operations were restricted to having the g 's all on one side, e.g., to systems having only operations of the first three of the above four types.⁸²

The resulting methods held out the possibility of an attack on the finiteness problem for systems having all four of the above types of operations, though certain of the difficulties of "tag" even then seemed glimmering in the distance. And just when hope was thus renewed for a solution of the general finiteness problem, a fuller realization of the significance of the previous reductions led to a reversal of our entire program.

II. The Anticipation

7. Generated sets of sequences

The power of canonical form *B* was demonstrated in §2 by the reduction of *10 *Principia Mathematica* to a single system in that canonical form. From this experience, and the knowledge of the kind of forms and the kind of operations appearing in the whole of *Principia Mathematica*, or could be made to appear if a complete symbolic development thereof were given, it becomes reasonably certain that all of *Principia Mathematica* can in similar fashion be reduced to a system in canonical form *B*. In the absence of the forbidding amount of work needed to actually carry out this

⁸² While the solution for the first order cases is completely written up in our notes (we have not, however, checked this solution) our authority for the more general result is but a statement to that effect in the writer's diary for that date.

reduction, added strength is lent to the above conclusion by the further reductions carried through in §4 and §5; for if the meager formal apparatus of our final normal systems can wipe out all of the additional vastly greater complexities of canonical form B , the more complicated machinery of the latter should clearly be able to handle formulations correspondingly more complicated than itself.

Granting the reducibility of the system of *Principia Mathematica* to a system in canonical form B , the reductions of §4 and §5 show that it is therefore further reducible to a system in normal form. We shall not linger, however, over the fact that the finiteness problem for the whole of *Principia Mathematica* would therefore be solved if the finiteness problem for the formally simple normal systems were solved.⁸³

Our present interest centers in the fact that in each reduction carried through in §5 the assertions of the given system are exactly those assertions of the new system which involve only letters of the given system. It follows that the assertions of any system in the wide canonical form C with primitive letters a_1, a_2, \dots, a_μ are those assertions of a system in normal form with primitive letters $a_1, a_2, \dots, a_\mu, \tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_{\tilde{\mu}}$ which involve only the letters a_1, a_2, \dots, a_μ . In fact, more generally, the same conclusion holds for the assertions involving only the letters a_1, a_2, \dots, a_μ of any system in canonical form C with primitive letters $a_1, a_2, \dots, a_\mu, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_{\hat{\mu}}$.⁸⁴ If then we think of canonical form C as a method of generating a set of (finite) sequences on letters a_1, a_2, \dots, a_μ , i.e., the set of assertions involving only those letters, we see that the generated sets of sequences yielded by all systems in canonical form C are the same as those yielded by the formally simpler normal systems.

Now for any system in canonical form C the premises and conclusion of any production thereof could be completely described in logical terms and the primitive relation of precedence in a sequence. This suggests the possibility of describing more complicated operations for the purpose of generating sets of sequences. Suppose each operation is of the form a certain number of premises, described in logical terms, gives rise to a certain conclusion, likewise described. Such an operation may be written P_1, P_2, \dots, P_k produces P where P_1, P_2, \dots, P_k, P have certain properties $f_1(P_1), f_2(P_2), \dots, f_k(P_k), f(P)$. Note that the P 's are not the operational variables of the operation, but that the latter are allowed for in the properties mentioned.⁸⁵ Now suppose that a set of postulates is set up for

⁸³ "Formally simple" refers to the bases of the normal systems.

⁸⁴ This observation does not seem to have been made explicitly in the notes. In the more general discussion begun in the next paragraph we therefore assume that no other letters than a_1, \dots, a_μ appear in canonical form C .

⁸⁵ This account is taken from a notes summary of previous developments written in February 1922. Mention is made of there being a little difficulty in connecting up the variables in P_1, \dots, P_k and P . We here think of these as free variables in terms of which the properties in question are expressed.

sequences on letters a_1, a_2, \dots, a_μ , and *Principia Mathematica* is used as the logic of the resulting mathematical system. Granting the generality of *Principia Mathematica*, sequences $P_1, P_2, \dots, P_\kappa, P$ will have the properties $f_1(P_1), f_2(P_2), \dots, f_\kappa(P_\kappa), f(P)$ when and only when the latter are assertions in the above *sequence-Principia Mathematica* system. If then we add the postulates on sequences and the system of *Principia Mathematica* to our assumed system for generating sets of sequences on a_1, a_2, \dots, a_μ , the above type operation of the latter can be written in the form

$$f_1(P_1), f_2(P_2), \dots, f_\kappa(P_\kappa), f(P), P_1, P_2, \dots, P_\kappa \text{ produce } P.$$

As a result, our system for generating a set of sequences on a_1, a_2, \dots, a_μ becomes the formal system of *Principia Mathematica* supplemented by certain postulates and operations of the same general type. Granting that *Principia Mathematica* can be reduced to a system in normal form, the same can be expected of the present system in its new form. If then, in translating the enunciations of our complicated system into enunciations of the normal system the letters a_1, a_2, \dots, a_μ are left unchanged, enunciations which are mere sequences of such letters being their own correspondents in the normal system, it will follow that the set of sequences on a_1, a_2, \dots, a_μ generated by our given system is again but the set of assertions of the resulting normal system which involve only the letters a_1, a_2, \dots, a_μ .

In view of the generality of the system of *Principia Mathematica*, and its seeming inability to lead to any other generated sets of sequences on a given set of letters than those given by our normal systems, we are led to the following generalization.

Every generated set of sequences on a given set of letters a_1, a_2, \dots, a_μ is a subset of the set of assertions of a system in normal form with primitive letters $a_1, a_2, \dots, a_\mu, a_1', a_2', \dots, a_\mu'$, i.e., the subset consisting of those assertions of the normal system involving only the letters a_1, a_2, \dots, a_μ .

8. Unsolvability of the finiteness problem for normal systems

In trying to test the correctness of the above generalization by our experience with sets of sequences the following counter-example seems to present itself. By that generalization, the only sets of sequences involving a single letter a would be the corresponding subsets of all normal systems on letters $a, a_1, a_2, \dots, a_\mu, \mu = 0, 1, 2, 3, \dots$. Now the class of all such normal systems is clearly enumerable (as will be seen in more detail in the next section). Suppose then that we define a class of a -sequences, i.e., sequences involving only the letter a ,⁸⁶ as follows. Enumerate the above normal systems, and

⁸⁶ Not to be confused therefore with the a -assertions of Part I which could involve any or all of the letters a_1, a_2, \dots, a_μ .

have $aa\dots a$ with m a 's in or not in our class according as it is not in or in the set of assertions of the m -th normal system. The resulting class of a -sequences will then differ from the corresponding subset of each of our normal systems, in seeming contradiction with our generalization.

Actually that generalization is not thus contradicted. For in our example we have merely *defined* a set of a -sequences, whereas to yield a true counter-example we must show how to *generate* that set, i.e., set up a system of "combinatory iteration"⁸⁷ whose operations would at some time yield each and every a -sequence in that set, but would never yield an a -sequence not in the set. On the other hand, suppose that the finiteness problem were solved for the class of all normal systems. Then for each of the above normal systems that solution would in a finite number of steps tell whether $aa\dots a$ with m a 's is or is not in that m -th normal system. An operation could then be set up which in order would pass down the above normal systems, for the m -th apply the test for $aa\dots a$ with m a 's being or not being in the system,⁸⁸ have a production which in the latter case produces $aa\dots a$ with m a 's for the desired system, and then in any case pass on to the next system. This operation iterated would then actually generate the above defined set of a -sequences. That is, a solution of the finiteness problem for all normal systems would yield a counter-example disproving the correctness of our proposed generalization.

Now we mentioned in §3 how an extended attempt to solve the simplified form of this finiteness problem "Tag" led to ever increasing difficulties, with all the complexities of number theory in the offing. On the other hand, nothing in the above argument directly weakens the reasoning that led us to our generalization. We therefore hold on to that generalization⁸⁹ and conclude that *the finiteness problem for the class of all normal systems is unsolvable*, that is, that *there is no finite method which would uniformly enable us to tell of an arbitrary normal system and arbitrary sequence on the*

⁸⁷ In the abstract referred to in footnote 19, mention is made of the method of combinatory iteration as an alternative to the truth-table method. Where the latter method involves an analysis of the logical situations a given deductive system may formalize, the former eschews all interpretation, and studies the system merely as a formal system. The operations of the system are then described as "combinatory" since they largely involve but a reshuffling of symbols; and it is through the "iteration," i.e., continued reapplication, of these combinatory operations that the entire system is obtained. We may note that the present development was entirely unaffected by the writer's published or unpublished work on the truth-table method.

⁸⁸ The notes at first required this finite test to give an upper bound to the number of steps required to perform the test as a function, presumably, of m . This unnecessary requirement was later deleted.

⁸⁹ In thus resolving this dilemma, the writer was greatly influenced by having heard, not long before, of Brouwer's rejecting the law of the excluded middle. This revolution in the writer's thought was largely energized by the immediately prior reading of Poincaré's *Foundations of Science*.

*letters thereof whether that sequence is or is not generated by the operations of the system from the primitive sequence of the system.*⁹⁰

The correctness of this result is clearly entirely dependent on the trustworthiness of the analysis leading to the above generalization. Apart from the details of that analysis having been given only in the special work of the first part of this paper, it is fundamentally weak in its reliance on the logic of *Principia Mathematica*, how weak will be seen in §10. This weakness could in part be overcome by replacing *Principia Mathematica* as the logic to be used in connection with postulates on sequences by an operational logic based on mathematical induction in the spirit of our detailed formal reductions, though keeping the primitives of *Principia Mathematica*. Thus, if we have different symbols a_1, a_2, \dots, a_μ , and consider an arbitrary sequence $a_{i_1}, a_{i_2} \dots a_{i_n}$ thereof, then we can introduce the propositional function "There is an a_i in the sequence" as follows. Introduce a new letter b such that the assertion of $ba_i bP$ is to mean that P is a sequence on a_1, a_2, \dots, a_μ involving the letter a_i . This will be accomplished by adding the one primitive assertion $ba_i ba_i$, and the two sets of productions, with $j = 1, 2, \dots, \mu$,

$$ba_i bP \text{ produces } ba_i ba_j P, \quad ba_i bP \text{ produces } ba_j bPa_j.$$

Similarly, for "There is no a_i in the sequence," etc. etc.⁹¹

But for full generality a complete analysis would have to be made of all the possible ways in which the human mind could set up finite processes for generating sequences. The beginning of such an attempt will be found in the Appendix. In the meantime, however, assuming the correctness of our characterization of generated sets of sequences, a mathematical derivation of the unsolvability of the finiteness problem for normal systems as a consequent theorem should be feasible. This is done at least in outline in the next section, and leads us in §10 to far-reaching conclusions on the nature of logical activity, and hence of mathematics.

⁹⁰ With slightly different wording, this was stated as a "Theorem" in the notes. Despite the closing of the circle of §6 not appearing explicitly in the notes, it was undoubtedly realized at the time that this result carried along with it the unsolvability of the finiteness problem for each of the canonical forms A on. Less certain, however, is our having paused at the time to realize that the completion of the proof of the reducibility of canonical form A to *10 *Principia Mathematica*, referred to in the second paragraph of footnote 79, would yield the unsolvability of the latter's finiteness problem. It remains uncertain, therefore, to what extent the writer anticipated Church's result on the unsolvability of the deducibility problem for the restricted functional calculus. (See Alonzo Church, *A note on the Entscheidungsproblem*, Journal of Symbolic Logic, vol. 1 (1936), pp. 40-41; also the reference in footnote 18.)

⁹¹ See footnote 8.

9. Outline of a minimum mathematical development

We recall the definition of a normal system. Let a_1, a_2, \dots, a_μ be μ distinct symbols given in order. We have given an initial sequence

$$A = a_{i_1} a_{i_2} \dots a_{i_\lambda}$$

made up of these symbols, repeated at pleasure, and a certain finite number of operations of the form

$$g_i P \text{ produces } Pg'_i$$

where g_i and g'_i are also such finite sequences. For ease in demonstrations we shall assume these operations to be given in order, that they need not be distinct, that g_i and g'_i may be either or both of them null, and that P may represent a null sequence provided g'_i is not also null.

We shall call the set of sequences resulting from the iterated application of these operations starting with the initial sequence a *normal system*. We shall not distinguish between normal systems which differ from each other only in the specific primitive symbols employed. We may therefore imagine these symbols to be the first μ symbols in any infinite sequence of symbols, say the first μ positive integers.

With this understanding, we first show that *the set of all normal systems can itself be ordered in an infinite sequence*. Actually it is the set of all possible bases of normal systems that is so ordered, so that one and the same normal system, as set of sequences, may appear several times in the ordering. By the *complexity* of a basis of a normal system we shall mean the total number of symbols appearing

- (a) in the set of symbols a_1, \dots, a_μ ,
 - (b) in the initial sequence,
 - (c) in the various operations, where we count each P as a separate symbol.
- We first then imagine the possible bases of normal systems divided into classes according to their complexities, and, as also below, order these classes in order of increasing complexity. Now in each class separate the bases into subclasses according to the number of primitive symbols, and correspondingly order these subclasses. Likewise in each subclass separate and order according to the number of operations. In each of the last found subclasses separate the bases according to the ranks of

$$A, g_1, g'_1, g_2, g'_2, \dots, g_k, g'_k,$$

rank of a sequence being the number of letters therein, i.e., its length, and order the resulting classes according to the rank of the first of the sequences which differ in rank for two classes. In each of the resulting classes two bases will be identical when and only when the single combined sequences

$C = Ag_1g_1'g_2g_2' \dots g_kg_k'$ are identical. As the number of primitive symbols μ is the same for all bases within a single class, if we interpret C as a number written in arabic notation with base $\mu+1$, i.e., $a_1 = 1, a_2 = 2, \dots, a_\mu = \mu$,⁹² the bases can finally be ordered within each class according to the number C represents. As a result the set of all bases for normal systems is ordered; and since the number of bases with given complexity is seen from the ensuing orderings to be finite, the entire ordering is that of a single infinite sequence.

We shall refer to the above ordering as the σ -ordering and use it in all subsequent work.

With our understanding about the primitive symbols of normal systems, all normal systems have the same primitive symbol a_1 which we shall, for simplicity, replace by a . Now consider the following set of sequences involving only the letter a : $a \dots a$ with n a 's is or is not in the set according as it is not or is in the n -th normal system in the σ -ordering. We shall refer to this set as the N -set. The mathematical (as opposed to logical) basis for the no finite method theorem lies in the following almost trivial theorem.

Theorem. *There is no normal system with the property that if its first primitive symbol be replaced by a then the set of resulting sequences involving only the letter a is the N -set.*

For such a normal system would appear at some point in the σ -ordering, say it would be the m -th. But then the set of a -sequences present in the normal system, and the N -set, would differ with respect to the presence of at least the sequence $a \dots a$ with m a 's; for by the definition of the N -set, if this sequence is present in the normal system it is absent from the N -set, if absent in the normal system it is present in the N -set.

As stated this theorem would be trivial were it not for the all embraciveness of normal systems.

Stated positively the theorem amounts to the following. Given a normal system S , then there exists a normal system S' such that if S' is the m -th system in σ , then it is false that $a \dots a$ with m a 's is in S and not in S' , or in S' and not in S . The proof of this existence consists in pointing out the object, to wit, S is such a system.

Our remaining "theorems" deserve that name only in the sense that a complete mathematical proof thereof clearly can be given – as contrasted with our generalization of §7. In the absence of the details of the proof we enclose the word "theorem" in a parenthesis.⁹³

⁹² The notes incorrectly say "radix μ with $a_1 = 0, a_1 = 1, \dots, a_\mu = \mu - 1$," a_2 written a_1 by a slip. Since the combined sequence may start with a_1 's, we can not let $a_1 = 0$.

⁹³ The notes use a "*" next to "Theorem" for the same purpose. Just prior to

We first have the important intermediate

(Theorem). *There exists a normal system K and a correspondence C such that for each normal system and enunciation thereof there is one and only one enunciation in K by correspondence C , and such that such an enunciation in K is asserted when and only when the corresponding normal system versus enunciation is such that the enunciation is an assertion in that normal system.*

The proof would have to refer back to the reduction proofs of §4 and §5. As these are not entirely determinate, they would have to be made so before the description of system K becomes as complete as that of the σ -ordering.⁹⁴

We shall refer to the system K as the *complete normal system* because, in a way, it contains all normal systems.⁹⁵

Now given any normal system M , we shall say that there exists a *finite-normal-test* for system M if there exists a normal system M' such that among the primitive letters of M' are all the primitive letters of M , and in addition, among possibly others, a primitive letter b , and such that if P is an enunciation of M we shall have P an assertion in M' when and only when it is an assertion in M , while bP is an assertion in M' when and only when P is not an assertion in M . Thus, for each P consisting wholly of primitive symbols in M one and only one of the two sequences P , bP is an assertion in M' , and that according as P is or is not an assertion in M .⁹⁶ We then have the fundamental

this development, the notes had taken up the "Probably Fallacious Suggestion for a Non-provable Theorem" quoted in the Appendix. This led to some misgivings concerning the no finite method theorem, and the present development was undertaken for the sole purpose of clearing up those misgivings. When, then, it was obvious that a "theorem" could be proved by the same sort of methods as were used in the reductions of §4 and §5, the detailed proof was "postponed" so that the continuing work on rendering the basic generalization of §7 unimpeachable would not be unduly interrupted.

⁹⁴ Reading between the lines, we may suggest that the proof would first set up a system K' in canonical form B serving the purpose of K , this being relatively easy since the infinite number of variables used in canonical form B would allow for the infinite number of primitive symbols occurring in the totality of all normal systems. K' would then be reduced to the desired normal system K by the reductions of §4 and §5. In the statement of the theorem "correspondence" must be understood as "effective correspondence," the theorem otherwise being without significance.

⁹⁵ The "complete normal system" would thus correspond to Turing's "universal computing machine." See *computable numbers*, pp. 241-246.

⁹⁶ The letter b thus serves as a symbol for negation. The full generality of this definition may be seen from the following equivalence. By the negative of a set of sequences on letters a_1, a_2, \dots, a_μ we shall mean the set of all sequences on those letters not in the given set. It is then readily proved that the existence of

(Theorem.) Then there exists no finite-normal-test for the complete normal system K .

The reductio ad absurdum proof would run as follows. Suppose there were such a system. Then out of it another system could be constructed which would have for its a -set the N -set. But this has been proved to be impossible. The method of proof would have to reduce the σ -ordering to a normal system operation.

We now examine the positive content of such a proof. Let L be any normal system on the letters of K and at least one additional letter b , and let L have the property that for each enunciation P of K one and only one of the two enunciations P, bP is in L . Let us write

$(m\text{-th system in } \sigma, a \dots a \text{ with } m \text{ } a's)$

for that enunciation of K which corresponds to the m -th system in σ and enunciation $a \dots a$ thereof by the preceding theorem. Then from L , by the argument in our supposed proof, we would get a system L' such that $a \dots a$ with m a 's would be in L' when and only when

$b(m\text{-th system in } \sigma, a \dots a \text{ with } m \text{ } a's)$

is in L .⁹⁷ Now let L' be the m' -th system in σ . Then if $a \dots a$ with m' a 's is in L' ,

$b(m'\text{-th system in } \sigma, a \dots a \text{ with } m' \text{ } a's)$

is in L , if $a \dots a$ with m' a 's is not in L' ,

$b(m'\text{-th system in } \sigma, a \dots a \text{ with } m' \text{ } a's)$

is not in L , whence, by our hypothesis on L ,

$(m'\text{-th system in } \sigma, a \dots a \text{ with } m' \text{ } a's)$

is in L . Hence L as a possible finite-normal-test for K gives the wrong answer for $a \dots a$ with m' a 's being in, or not being in, the m' -th system in σ , L' .⁹⁸

That is, for each normal system L of the above kind, a normal system L' can be found such that if L' is the m' -th system in σ , then L considered as a finite-normal-test for K gives the wrong answer for $a \dots a$ with m' a 's being in L' . While the proof thus gives the case for which the answer is

a finite-normal-test for a normal system M is equivalent to the negative of M being a generated set, generated set here being defined by the generalization of §7.

⁹⁷ Note the positive nature of the assumed process. That is, from the presence of

$b(m\text{-th system in } \sigma, a \dots a \text{ with } m \text{ } a's)$

in L would be obtained the presence of $a \dots a$ with m a 's in L' .

⁹⁸ That is, L as finite-normal-text for K would say that according as $a \dots a$ with \tilde{m} a 's is or is not in L' ,

$(\tilde{m}\text{-th system in } \sigma, a \dots a \text{ with } \tilde{m} \text{ } a's)$

is not or is in K . But, by the construction of K , the reverse is true.

wrong, it does not of itself tell what answer is given. This will contrast strongly with our conclusion of the next section where a different hypothesis is imposed on L .⁹⁹

10. Incompleteness of Symbolic Logics

In our last discussion we forced L to give a unique answer to each question, as it were, put by the complete normal system K , and found that in at least one instance it then had to give the wrong answer. We now, rather, require L to give the right answer when it answers at all, and see what this weakening of the requirements for a finite-normal-test for K leads to.

Specifically, let L be a normal system whose primitive letters include the primitive letters of K and another primitive letter b , and which has the property that for any normal system S and enunciation P thereof (S, P) appears in L when and only when P is in S , while if $b(S, P)$ is in L , then P is not in S . This property is equivalent to the following. (S, P) is in L when and only when it is in K , while if $b(S, P)$ appears in L , (S, P) is not present in K . Any such L we shall call a *normal deductive-system* adjoined to K . K itself, with merely the letter b added to its primitive letters, is such a system. While the first half of the property for L could have been made as weak as the second, there is no reason for doing so since by suitably adjoining K to such a weak L the stronger L would result.

If then L is a normal-deductive-system adjoined to K , as at the end of the last section, we can obtain from it a normal system L' such that L' has $a \dots a$ with m a 's in it when and only when

$b(m\text{-th system in } \sigma, a \dots a \text{ with } m \text{ }a\text{'s})$

is in L . But now it follows that if L' is the m' -th system in σ , then $a \dots a$ with m' a 's is not in L' . For $a \dots a$ with m' a 's could only appear in L' through

$b(m'\text{-th system in } \sigma, a \dots a \text{ with } m' \text{ }a\text{'s})$

being in L . But by our definition of a normal-deductive-system if

$b(m'\text{-th system in } \sigma, a \dots a \text{ with } m' \text{ }a\text{'s})$

is in L , $a \dots a$ with m' a 's is not in that m' -th system in σ , i.e., not in L' . Hence also $b(m'\text{-th system in } \sigma, a \dots a \text{ with } m' \text{ }a\text{'s})$

⁹⁹ The theorem just discussed, when combined with the generalization of §7, therefore shows not only that the finiteness problem for the class of all normal systems is unsolvable, but that the same is true for a particular one of them, namely K . This, while not explicitly stated at the time was then fully realized, and receives explicit mention in our notes for Feb. 28, 1924. We might here also mention two theorems stated, and at least proved in outline, under the date of March 7, 1924. "There is no finite method for testing whether an arbitrary system does or does not admit of a finite test." "There is no finite method for testing whether an arbitrary system has a finite or infinite number of assertions." Some doubts were expressed, however, about one stage of the contemplated proof of the second result.

is not in L_j for it were, then, by the construction of L' , $a \dots a$ with m a 's would be in L' . We therefore have the very important

(Theorem). *No normal-deductive-system is complete, there always existing a normal system S and enunciation P thereof such that P is not in S while $b(S, P)$ is not in the normal-deductive-system.*

We now have the still more important

(Theorem). *No normal-deductive-system is equivalent to the complete logical system (if such there be); better, given any normal-deductive-system there exists another which second proves more theorems (to put it roughly) than the first.*

The proof would run as follows. Let L be the given deductive system. The proof envisaged above would be an informal proof to the effect that a system $L' = S_{m'}$, the m' -th normal system in σ , could be constructed such that $a \dots a$ with m' a 's is not in $S_{m'}$ while $b(S_{m'}, a \dots a$ with m' a 's) is not in L . On formalizing the proof there seems to be no doubt that it could be reduced to normal form, and on being adjoined to L would give a system where a theorem is proved which is not provable in L .¹⁰⁰

When these results are expanded to the dimensions of the no finite method theorem we are led to the following

*A complete symbolic logic is impossible.*¹⁰¹

¹⁰⁰ I.e., the correspondent of

$$b(S_{m'}, a \dots a \text{ with } m' \text{ } a\text{'s}) \quad (*)$$

in the wider system. Except for the non-essentials, we have stated the theorem, and given the outline of proof, as they appear in the notes. Actually, to obtain this result, a far less drastic procedure would suffice. We need merely add $(*)$ as a primitive assertion to L , and reduce the resulting system to normal form — the normal form of L having been spoiled by having two primitive assertions instead of one. However, the proof first contemplated is probably far more satisfactory. For the entire development should lead away from the purely formal as the final ideal of the mathematical science, with a consequent return to postulates that are to be "self-evident" properties of the now meaningful mathematical science under consideration. Merely adding $(*)$ as a new postulate would then be inadmissible. And if it be said that the previous informal development constitutes a heuristic justification of such an addition, the reply would be, incorporate that development in the system itself. And so the formalization suggested in the original outlined proof; new postulates introduced in that formalization to be "self-evident" properties of the modes of symbolization etc. of the old logic.

¹⁰¹ This has a double content. Mere incompleteness, as in the first of the two "Theorems" preceding, might not rule out the logic being as complete as it ever could be made. Fundamental, then, is the added effect of the second theorem, which rules out the possibility of a completed symbolic logic. That is, any symbolic logic can be made more complete. It is doubtful if the writer ever paused to

This is an iconoclastic result from the formal logician's point of view since it means that logic must not only in some parts of its description (as in the operations), but in its very operation be informal. Better still, we may write

The Logical Process is Essentially Creative

This conclusion, so in line with Bergson's "Creative Evolution," is not so much contrary to Russell's viewpoint (since he does not fully express himself) but to that of C. I. Lewis as given in Chapter VI, Section III, of his "Survey of Symbolic Logic." It makes of the mathematician much more than a kind of clever being who can do quickly what a *machine* could do ultimately. We see that a *machine* would never give a complete logic; for once the machine is made *we* could prove a theorem it does not prove.

Appendix

While the formal reductions of Part I should make it a relatively simple matter to supply the details of the development outlined in §9 and the beginning of §10, that development owes its significance entirely to the universal character of our characterization of an arbitrary generated set of sequences as given in §7. Establishing this universality is not a matter for mathematical proof, but of psychological analysis of the mental processes involved in combinatory mathematical processes. Because these seemed to be sufficiently simple to be exhaustively described, the writer gave up a direct use of *Principia Mathematica* as a partial verification of the characterization in question, planning rather that the incompleteness of the logic of *Principia Mathematica* would be a corollary of the more general result.¹⁰² Actually, we can present but fragments of the proposed analysis

note the mere incompleteness of a symbolic logic in the sense of the existence of some undecidable proposition therein, for experience with Zermelo's axiom, the axiom of infinity, and the theory of types clearly leads one to expect incompleteness in the upper reaches of a symbolic logic. Rather was the emphasis placed on the stronger concept of incompleteness with respect to a fixed subject matter, in the present instance the propositions stating whether a given sequence is or is not generated by the productions of a given normal system from its initial sequence. Likewise, Gödel would stress, for example, the incompleteness of any symbolic logic with respect to the class of arithmetical propositions. Where we say "symbolic logic" the tendency now is to say "finitary symbolic logic." However, it seems to the writer that logic should be considered essentially a human enterprise, and that when this is departed from, it is *then* incumbent on such a writer to add a qualifying "non-finitary."

¹⁰² On the other hand, as late as September 3, 1929, among a few observations on our work of 1924, appears the following in our notes. "Plan of first proving *Principia* inadequate through special analysis [as decided December, 1925] still

of finite processes. Being but fragments, these contributions are given by direct quotation from our notes and diary.¹⁰³ Quotations from our notes will be headed by a longer double line, from the diary by a shorter.

The unsolvability of the finiteness problem for all normal systems, and the essential incompleteness of all symbolic logics, are evidences of limitations in man's mathematical powers, creative though these be. They suggest that in the realms of proof, as in the realms of process, a problem may be posed whose difficulties we can never overcome; that is that we may be able to find a definite proposition which can never be proved or disproved.¹⁰⁴ This theme will protrude itself ever so often in our immediate task of obtaining an analysis of finite processes.

... nature as infinite intelligence ...

... Fermat's theorem perhaps unprovable due to weakness of our logical apparatus. Have a good vision of the infinitude of integers and stateable properties of them which cannot be unravelled by our logical process of syllogism etc.

... Think logic for finite operations ... Really only get here what would correspond to a non-growing machine operating in a discrete symbol-space.

We begin here a *derivation* of the logic of *finite operations* and ultimately of all the *logic of mathematics* from *first principles*. These principles are supposed to be a digest of our experience of the logico-mathematical activity ...

We first note that we have to do with a certain activity of the human mind as situated in the universe. As activity this logico-mathematical process has certain

temporal properties

as situated or performed in the universe it has certain

spatial properties.

seems good. Will not be work wasted but help clarify above." The plan however included prior calisthenics at other mathematical and logical work, and did not count on the appearance of a Gödel!

¹⁰³ See footnote 5.

¹⁰⁴ See the end of footnote 1.

Now the objects of this activity may be anything in the universe. The method seems to be essentially that of *symbolization*. It may be noted that language, the essential means of human communication is just symbolization. In so far as our analysis of this activity is concerned, its most important feature is its ability for being *self-conscious*. Since our present study is this activity we shall here *not* consider the *original* objects which are symbolized in the process, but only the relations and operations upon these resulting *symbols* and the effect of this *self-consciousness*. . . .

We are then to consider this activity of the human mind and watch it in this process of *creating symbols*. We may then note as before that the process itself is *temporal* but we are to think of the result of the process as *spatial*. We are then to think of these symbols as created in the flux of the universe and *preserved* there through time. This gives our first principle, i.e.,

A. The Principle of the Preservation of Symbols. . . .

We are then to think of the result of logical thought as certain *spatial configurations of symbols*; and our study will then consist in studying the *further effects* brought about by the processes of *symbolization* and *self-reflection*. Now in our subject we are to regard our symbols as without properties except that of permanence, distinguishability and that of being part of certain symbol-complexes. But this latter is essential, i.e. that these symbols enter into certain spatial (not Euclidean or continuous etc., but spatial as opposed to temporal to be described later) relations. These relations themselves can then be symbolized and the new symbols are *again* in space and have certain spatial relations, etc. So much for the further effect of symbolization of the spatial properties. But in addition we have this self-reflectiveness. This is a reflection of the process. This process is then itself *symbolized* and symbolized by a *spatial symbol*.

We thus have a continued activity which produces symbols which are spatial. This activity turns on itself and symbolizes its temporal character by a spatial symbol. These spatial symbols have certain spatial relations which are in turn symbolized by a spatial symbol.

Better we have a three fold order of things.

- (a). Activity in time which is creative. This is the source of the process.
- (b). By reflection this activity itself is frozen into spatial properties.
- (c). These spatial relations are symbolized by spatial symbols.
- (d). These symbols have no further symbolizable properties *internally* as it were and so end the descent. (This is essentially Bergsonian).

So much for the introductory discussion before coming to an analysis of these spatial relations etc. We may note here however, that it is in the specific bringing in of the self-consciousness and its symbolic representation

that this differs from other schemes. In the old mathematics the *formal* and *informal* were confused, in symbolic logic the informal tended to be neglected. Right here we may see that the greatest ordinal etc. fallacies come just from this neglect; we may thus expect the above to give us a rational theory of types. We may note as a fundamental property of the above that everything ends up in spatial relations and these are transformed into themselves. This is at the basis of the no finite method theorem.

...but really only the productiveness and not creativeness of mind
here recorded.

*A Probably Fallacious Suggestion for a
Non-Provable Theorem*

Above we considered the question of a finite method for testing all theorems of a certain class. The no finite method theorem did not depend on how the test is shown to be a true test.

In trying to obtain a theorem which can not be proved either true or false the valid methods of proof have to be stated. Now as in finite methods the set of finite proofs probably forms an illegitimate totality. In the first case however, any one could be reduced to an equivalent one which is 'predicative.' The state of affairs for the second is not quite clear.

In any case consider the following enunciation: *For each deductive system of the normal form and enunciation in it there exists a finite method of proving or disproving the derivability of the enunciation in the system.* It will be noticed that the enunciation wants a finite method of proof for each case not a test for all. Now this may not be a mathematical theorem in that the notion of general finite method of proof is used whereas the set of them may be an illegitimate totality. If it is a definite enunciation it would seem to be neither provable nor disprovable. For if it were proved we would really have a method of proof described for all cases at once and so have a finite test.¹⁰⁵ On the other hand to disprove it we would have to show a case where the deductibility could neither be proved nor disproved. But if the deductibility exists then it can be proved. Hence showing it can't be proved amounts to disproving it, and so we can't prove that it can't be proved nor disproved both. Hence the above theorem can't be proven false.

¹⁰⁵ Tarski recently pointed out to the writer that this argument is invalid since the supposed proof might be non-constructive.

.....

This seems to lead to a distinction between finite operations and finite methods of proof. All finite operations can be lumped into one system while all proofs can not. ... Though both the set of all finite operations and all finite proofs are illegitimate yet there is an equivalent legitimate set in the former case but not in the latter — at least that can be described. ... may not this prevent us from ever *proving* a theorem *non-provable*? The only hope seems to be a method of invariants and illegitimate totality induction.

....

... we do not completely determine the process of proof; we simply *watch it* in its activity, from that note some of its properties which may enable a non-provability theorem to be possible.

... creative totality versus Cantor's transfinite?¹⁰⁶

Think relation between the creative process¹⁰⁷ and transfinite ordinal numbers; see where the creative process is really *Principia* superimposed upon types which follow each other as the transfinite ordinals. This explains the relation but brings up question of Ω , etc.

Think real numbers etc. as creative totalities. ...¹⁰⁸

Write up this beginning of an operational theory of Mathematics.¹⁰⁹

What we must do is further analyse the process of proof. It was shown that it is creative, but the creative and non-creative parts were intermingled. What we must now do is isolate the creative germ in the

¹⁰⁶ This and succeeding entries follow all of the work corresponding to Part II. The entries are in chronological order.

¹⁰⁷ See the end of §10.

¹⁰⁸ There follows a description of mental states during mathematical discovery which is probably too "cloudy" for inclusion here.

¹⁰⁹ The surface is here but barely scratched. We may, however, note the conclusion, "The Cantor Theorem is not true in the operational theory of infinity," i.e., the Cantor-Bernstein-Schröder theorem. The notes add, "In any case this whole thing is *premature at this stage*."

*thinking process.*¹¹⁰

The following suggestions came up.

- (a). The conclusion that man is not a machine is invalid. All we can say is that man cannot construct a machine which can do all the thinking he can. To illustrate this point we may note that a kind of machine-man could be constructed who would prove a similar theorem for his mental acts.
- (b). The creative germ seems not to be capable of being purely presented but can be stated as consisting in constructing ever higher types. These are as transfinite ordinals and the creative process consists in continually transcending them by seeing previously unseen laws which give a sequence of such numbers. Now it seems that this complete seeing is a complicated process mostly subconscious. But it is not given till it is made completely conscious. But then it ought to be constructable purely mechanically.

This last is an assumption which is to be fundamental in the whole discussion. It is to be the Axiom of Irreducibility¹¹¹ for Finite Operations.

- (c). We prove a sequence not capable of being derived in a normal system by showing it does not satisfy a property which is shown by Mathematical Induction to be possessed by all sequences which are generated. This is old. The new thing is that this hereditary property may be of various types, and it is thus that the creative element appears in proof. (All of above before reading Brouwer.)
-

The following suggestions occur.

- (a). The following would be the beginning of a definition of a general non-growing machine for (b) of yesterday. It will have a finite number of parts a_i , i.e., $a_1, a_2 \dots a_n$ and a finite number of relations of an arbitrary number of parts i.e.,

$$b_j(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{m_j}).$$

At each instant certain of these parts are related and not others. Specify this. Then we would have say a definite rule for a certain configuration of relations producing another. This would however be a machine closed in itself. Instead we want it to be capable of creating say im-

¹¹⁰ The diary, at the corresponding point, takes a religious turn.

¹¹¹ "Reducibility" probably meant.

pressions in an outside medium and of having such impressions affect it. This I have not completely described.¹¹²

- (b). Instead of having the bald assumption of (b) of yesterday we may have the milder assumption that a new sequence may use ideas in defining it, but these ideas can only be those previously defined. This assumption looks less general. If it can be put clearly then we can prove that every such new idea will ultimately be defined mechanically since each idea used has been so defined, i.e., we use an operational-induction.
-

Think Creative process. . . See work on Theory of Deduction in three parts (a). No finite method theorem. Almost have all of this. (b). The nature of proof. Only have certain starting points but no connected whole as yet. (c). On the question of proving theorems unprovable both ways. Have almost nothing here. . .

. . . think creative process. See where a symbol for 'produce' would come in in each generalization of operations; and as this symbol became 'external' it would as it were change into activity. Thus like \supset but not $\sim p \vee q$ as $\sim p$ defined in terms of it.

. . . Again see a *class* as an *operation*.

Think above. Again get picture I had before of transfinite ordinals as machines creating machines etc. (machine for operation).¹¹³

¹¹² Apart from the incompleteness referred to at the end of this note, this fails to be an anticipation of the concept of the Turing machine (see *computable numbers*) in its very attempt to allow for the structure of the machine. However, in the case of a growing machine, the number of states of the machine would no longer be finite, and reference would have to be made to the structure of the machine. The writer did not attempt to use this partial concept of a non-growing machine as an easy path towards the justification of the generalization of §7 since the clear possibility of having machines producing machines, and, more generally, of a machine directing its own growth, seemed to make any inferences drawn from the nature of a non-growing machine inconclusive.

¹¹³ A certain amount of detailed work occurs in the notes on the generation of ordinals considered as operations. However, as this was largely undertaken for the experience to be thus gained, we do not quote any of it. Concerning it we may quote a criticism of Sept. 3, 1929 of the work of 1924. "In trying to see beyond this I seemed to get lost as in the former work on generating transfinite ordinals. All I could see was that every now and then with increasing difficulty a new idea comes." Also might be mentioned the appearance in the notes of a partial check of Galois theory as a finite process, and a "logical description" of

... proof consists in raising the unseeing combinatory iteration which is of the same type to ever higher types where see through it all. ... power of seeing due to things being mutually exclusive ... difficulty of combinatory iteration is that you have destruction ...¹¹⁴

... 'transcendental definitions.' Ought to be able to give such for 'finite proof' - i.e., like definition of well-ordered series. ... order of creativeness of a proof.

... Also get idea of constructing a theorem for each type of creativeness provable by proof of that type but not by lower type.

... This leads to a new hierarchy of proofs, proofs of proofs, etc.

Finite operations illuminated as generated by three principles (1) Symbolic 'manipulation' (2) Symbolization (3) Iteration.

We return here to a more complete discussion and analysis of the very first part of the present research i.e., in connection with finite methods. We shall here generalize to finite methods for obtaining any results not just *test* for truth or falsity ...

We shall here first give what is at least a first approximation to a definitive solution of the difficulty of finding a *natural normal* form for symbolic representation.

There are three stages in the analysis we give. In the first stage we have the things symbolized.

.... This then gives us our second stage in our analysis, namely a system of symbolizations for corresponding mathematical states.

.... First these symbolizations will be conceived of as being spatial. ... But we shall also assume them to be finite and we might say discrete. ... First each symbolization can be considered to consist of a finite number of unanalyzable parts (unanalyzable from the standpoint of the symbolization)

the symbol forms of canonical form A.

¹¹⁴ An attempt to "dig deeper" led to a vision of the "Birth of Consciousness." In our quotations we have omitted these more extravagant features of notes, and, chiefly, diary.

these parts having certain properties and certain relations with each other. If we allow unary relations we need merely talk of a finite number of parts related in certain ways. The ways in which these parts can be related will be assumed to be specified for the whole system of symbolizations.

We thus have what may be called a symbol-space in which these symbolizations or symbol complexes as we may now call them are ...

... It will be assumed that these spatial properties and spatial relations are themselves not further analyzable in the given discussion, in other words that a single undivided act of judgment is involved in finding out whether a certain part has or has not a certain property or that certain parts are or are not related in a certain way. Now the system of symbolizations in question is essentially to be a human product and each symbolization a human way of describing the original mathematical state. The need for the following assumption is then readily seen: that the number of these elementary properties and relations used is finite and that there is a certain specific finite number of elements in each relation.

We are now ready to come to the third and last stage in this analysis. We can do this by assuming what seems to be the evident intent of the above analysis and that is that the symbol-complexes are completely determined by specifying all the properties and relations of its parts as above mentioned. Hence for the given system of symbol-complexes we have a certain number of relations which can be made to include properties by allowing but one variable and may be written

$$a_1(x_1, x_2, \dots x_{\nu_1}), a_2(x_1, x_2, \dots x_{\nu_2}), \dots, a_n(x_1, x_2, \dots x_{\nu_n}).$$

Then if \prod is to mean logical product of all elements, then each complex of the system can be completely described in the following form

$$\prod_i a_{n_i}(x_{j_{1i}}, x_{j_{2i}}, \dots, x_{j_{\nu_{n_i} i}})$$

which gives all the relations enjoyed by the elements or parts $x_1, x_2, \dots x_m$ of the symbol complex.

Now these descriptions completely describe the symbol-complexes which in turn represent the original mathematical states. Hence these descriptions can be considered to represent or symbolize those mathematical states.

We thus have it that every system of symbolic representations which satisfies the assumptions we have given is equivalent to one of these systems. These latter (gotten by varying $\nu_1, \nu_2, \dots, \nu_n$ and n) constitute then a normal set of systems of symbolic representations.¹¹⁵

¹¹⁵ Under the date of Sept. 4, 1929 appears the following. "The neighborhood idea applied to the general symbol space will lead to difficulties due to inter-

I. Finite Methods

II. Theory of Deduction

III. Theory of Conception or Concepts

...I study Mathematics as a product of the human mind and not as absolute. ...

...Notion of meaning bothers me. Put it as subconscious perception of things associated with symbol.

...get idea of Psychic Ether,...see 'language' as a difficulty...

...in all finite methods or just methods only the following principles are used: symbol-manipulation, symbolization and iteration. ...although iteration is the process, it is merely machine like, and the real thought was put into that which is iterated.

.....

Corresponding to the three stages in the analysis of symbolism there were three stages in the analysis of method.¹¹⁶ In the first stage were any and all methods. To allow for the most wonderful creations my image of such methods involved dark clouds pierced by flashes of lightning accompanied by rolling thunder. These methods then being regarded as existing in time were symbolized at each time (second stage).¹¹⁷ Due to discreteness and finiteness we would thus have a finite sequence of symbol-complexes representing the various stages in the method. The third stage would then consist in reducing the method of passing from symbol-complex to an op-

relations. In fact, this difficulty was mentioned elsewhere. The general symbol space should be reconsidered." We rely on memory in saying that it was by then realized that even though each symbol complex was determined by the logical product of relations satisfied, the converse was complicated by the fact that the symbol space itself might force the elements to have certain relations when certain others were satisfied, so that, in fact, a finite process would have to be set up which would generate exactly those logical products of relations corresponding to symbol complexes.

¹¹⁶ The past tense is used because the notes here describe previous ruminations.

¹¹⁷ That is, were imagined to be so symbolized. There is confusion here between the method or rule, and the carrying out thereof.

eration of normal type on symbol-complexes of the above normal type.

... it was seen that even in the first stage we really haven't got Mathematics but a more or less vague symbolization of it. In other words that in all cases what we are *conscious of* is not mathematics but a symbolization of it. That these symbols are more or less vague, that in making them more precise we really resymbolize the symbols etc. It was thus immediately evident why the *process* of symbolization must inevitably come in.

... the difficulties in connection with the notion of *meaning* came in. It was then recognized that a symbol has meaning because it is *connected* with all the things which give it meaning though these things are in the subconscious regions. Meaning is then the result of subconscious perception. The symbol is like an island which rises above the sea of the unconscious and is itself connected with the vaster regions below.

The question was how iteration was brought in through mere spatial symbolization. How activity at all. It was then seen that activity is symbolized by a spatial symbol. Hence in the symbol-complex representing the finite method certain parts represent temporal things and so give rise to activity. How iteration? Well, the symbol-complex which is spatial has temporal *permanency*, hence merely by its persisting through time, it keeps on giving directions and so keeps on stimulating the activities which are then said to be iterated.

This brought on a difficulty that all the symbol-complex can do is suggest, thus requiring a mind to interpret and direct. ... the relation of mind and one might say matter. It was given in a more general form, i.e., not just for a definite method but in all cases. The distinction was that in a definite method we have a deterministic system whereas otherwise an indeterministic one where ideas appear without cause (i.e., cause within the system considered).

... a notion which should be very fertile indeed, i.e., that of the *Psychic Ether*. In so far as this contains the things which give meaning to symbols it is just the unconscious. ... it is the seat of the birth of new ideas; it is also the region where all the vaguer processes operate especially intuition, 'hunches' etc. When these become precise they crystalize from the psychic ether. Clear symbolizations are then to be regarded like atoms in this ether. In fact one may think of them as Kelvin thought of his vortex atoms in connection with the lumeniferous ether.

... a new difficulty came up, i.e., how to place language of the ordinary kind in all this. It loomed up as a very queer thing in all this: its fluency etc., etc. The following suggestions came up. Its linear order can be easily

associated with time. But it is rather difficult to see just what its essence is. On the one hand . . . it seems to be used to call up images as when we read a novel. It then merely is the suggestion while we are conscious of the thing it suggests. On the other hand we often are conscious of only the words, but as having meaning as when we assent to the statement: Every continuous function is integrable. Whether these two are the only aspects of language, and how they are related to each other and how they connect up with the psychic ether is yet to be determined. The relation between ordinary language and its predecessor hieroglyphics should help to bring out the connection between the two aspects above mentioned.

... on the next attack try to put everything into the psychic ether and then . . . proceed to the next real difficulty. This . . . has to do with just the way in which the connections between vague ideas turn into connections between precise ideas. When this is done we should not be far from the normal form for finite methods.

It is to be noted that the analysis of methods has become of much greater importance than formerly. In fact it almost looks as if analysis of proof and concepts will almost be

Corollaries

of the present analysis.

... see very prettily how can avoid physical symbols etc. by noting that effect in our reasoning just as if only imagined them; i.e., Psychic Ether complete in itself.¹¹⁸ Like Bergson's theory of memory. See my work as a complete and scientific psychical system – the first. (Mechanics is such a physical system).

Proceed to get the mechanism or rather mechanics of the Psychic Ether. . . . we can only *handle* one thing at a time . . .

... in our mental activities the following occurs repeatedly: often a

¹¹⁸ This serious error was corrected in our work of 1924 where the dualism of the physical world versus the mental world was stressed in contrast with the above monism. Fundamental is the distinction between the static outer symbol-space with its assumed capacity for bearing symbol-complexes of unbounded complexity, and the dynamic mental world with, however, its obvious limitations. This has been fully emphasized by Turing in his finite number of mental states hypothesis (referred to in footnote 9). Perhaps we should quote the following item from the diary, entered some two weeks later. "...the beauty of the self-sufficiency of the psychical system. See clearly the symbols we imagine as floating in the psychic ether etc. Raises the question may not matter similarly be the visions of God? . . . by saying above we can think of physical things as though just visions in our psychic ether we have a *Psychic Principle of Equivalence . . .*"

given symbol has full meaning for us and we simply observe the symbol. When it has lost some of its 'meaning' we replace it by the thing it symbolizes. This is the converse of symbolization ...¹¹⁹

... Really little time was spent above on the way in which operations were symbolized. What was seen was the atomic operations. ... But the way in which these combine recombine etc. was not noticed. ...

... mathematical operations are had when the symbols become clear-cut enough to be handled as individuals and in combinatory fashion. ...

Another difficulty has not been brought out and that is the place of logic as relating to truth and falsehood. It seems that we will have to bring in some logic, that of direct verification of the existence of an elementary relation between symbols.

... recognizing the truth of a certain logical situation is a process since in general it can't be done at a glance. It must then be reduced to elementary recognition only.

Another example of this difficulty is that of substituting a particular case for a variable. This seems to involve

- (1) Recognition of the special case as a value: Elementary Recognition.
- (2) Saying this in symbols – symbolization.
- (3) Substitution – symbolic manipulation.

In any case the reduction seems possible so that only elementary recognition seems to be needed for the logical aspect of the *operational description* of Mathematics.¹²⁰

.....

A summary of the *method* used above is as follows. Try to give a complete *description* of what goes on. In this description we symbolize everything. That symbolizes away most things. But a few things cannot be symbolized away, because in the *transformation* produced by the symbolization they of necessity *reappear*. These things are meaning, symbolization, symbol manipulation, iteration, sense perception or direct verification, and perhaps a few other things. These will then constitute the elements out of which the description is built up in addition to mere symbols.

The City College
The College of the City of New York

¹¹⁹ This is part of an "outline of complete solution" which, however, largely duplicated earlier entries, followed by a discussion of difficulties yet to be conquered.

¹²⁰ While many other such evaluations of work done has been omitted in our quotations, the following should be mentioned. "...the time for complete reorganization is over. The main outline of the work is completed and we really have a case of Filling In." Actually, but the surface of the problem was thus, perhaps, barely scratched; the problem, that is, of describing "all the finite processes of the human mind," at least in so far as they might concern the generalization of §7.

FORMAL REDUCTIONS OF THE GENERAL COMBINATORIAL DECISION PROBLEM.*

By EMIL L. POST.

1. **Introduction.** It is not new to the literature that the usual form of a symbolic logic with its parenthesis notation and infinite set of variables can be transformed into one in which the *enunciations*, i. e., formulas of the system, are finite sequences of letters.¹ the different letters constituting a once-and-for-all given finite set. If the primitive letters of such a system are represented by a_1, a_2, \dots, a_μ , an arbitrary enunciation of the system will take the form $a_{i_1} a_{i_2} \dots a_{i_n}$, $n = 1, 2, 3, \dots, i_j = 1, 2, \dots, \mu$. In describing the basis of such a system it is convenient to use new letters to represent finite sequences of the above primitive letters. If then A, B, \dots, E represent the sequences $a_{i_1} a_{i_2} \dots a_{i_\rho}$, $a_{j_1} a_{j_2} \dots a_{j_\sigma}, \dots, a_{m_1} a_{m_2} \dots a_{m_\phi}$ respectively, $AB \dots E$ will represent the sequence $a_{i_1} a_{i_2} \dots a_{i_\rho} a_{j_1} a_{j_2} \dots a_{j_\sigma} \dots a_{m_1} a_{m_2} \dots a_{m_\phi}$.

We shall say that such a system is in *canonical form* if its basis has the following structure.² The *primitive assertions* of the system are a specified finite set of enunciations of the above form. The operations of the system are a specified finite set of *productions*, each of the following form:

$$\begin{aligned} & g_{11} P_{i'1} g_{12} P_{i'2} \dots g_{1m_1} P_{i'm_1} g_{1(m_1+1)} \\ & g_{21} P_{i''1} g_{22} P_{i''2} \dots g_{2m_2} P_{i''m_2} g_{2(m_2+1)} \\ & \dots \dots \dots \dots \dots \\ & g_{k_1} P_{i_1^{(k)}} g_{k_2} P_{i_2^{(k)}} \dots g_{km_k} P_{i_{m_k}^{(k)}} g_{k(m_k+1)} \\ & \text{produce} \\ & g_1 P_{i_1} g_2 P_{i_2} \dots g_m P_{i_m} g_{m+1}. \end{aligned}$$

* Received November 14, 1941; Revised April 11, 1942.

¹ More exactly, "strings" of "marks," to use terms of C. I. Lewis (*A Survey of Symbolic Logic*, Berkeley, 1918: chapter VI, sec. III).

² This formulation stems from the "Generalization by Postulation" of the writer's "Introduction to a general theory of elementary propositions," *American Journal of Mathematics*, vol. 43 (1921), pp. 163-185 (see p. 176). We take this opportunity to make the following *Emendation*: Lemma 1 thereof (pp. 177-178) requires the added condition that the expressions replacing the r 's do not involve any letter upon which a substitution is made in the given deductive process. This necessitates several minor changes in the proof of the theorem there following. Actually, both Lemma 1 and its companion Lemma 2 admit of further simplification, with the proof of the theorem then being valid as it stands.

In this display the g 's represent specified sequences of the primitive a 's, including the null sequence, while the P 's represent the operational variables of the production, and, in the application of the production, may be identified with arbitrary sequences of this type. In this notation, the distinct operational variables of a given production are to constitute the finite set of symbols P_1, P_2, \dots, P_M for some positive integral M . We then add the restriction that each operational variable in the conclusion of a production is present in at least one premise of that production, it having been understood that each premise and conclusion has at least one operational variable. We further assume that no identification of the operational variables is permitted which would lead to the conclusion being null.³ The *assertions* of the system are then the primitive assertions, and all enunciations obtainable by the repeated application of the given productions starting with the primitive assertions.⁴ More precisely, the class of assertions is the smallest subclass of the class of enunciations which contains the primitive assertions, and which, for each admissible assignment of values to the operational variables of each of the given productions, contains the enunciation represented by the conclusion of

³ For the proof of Section 2 to be universally valid it is necessary that the operations themselves exclude the possibility of a null assertion. Since a null conclusion could arise only from an operation whose conclusion consists of operational variables only, while under our restriction at least one of these operational variables is not to be null, we can achieve the desired automatic exclusion of the null assertion by replacing each such operation by the following equivalent finite set of operations. For each operational variable P_i in the conclusion of such an operation, and each primitive letter a_j , form the operation obtained by replacing P_i by $a_j P_i$ throughout the given operation. This modification need only be made on the given system in canonical form; for the productions introduced in Section 2 all have their single premises consist of more than just operational variables, so that the nonexclusion of the null assertion would have no further effect on the system.

⁴ That this leads to the constructive generation of the class of assertions is readily verified. In particular, in trying to identify the premises of a production with corresponding previously obtained assertions, an explicit hypothesis on the "rank" of the operational variables involved, rank of a sequence being the total number of letters therein, would immediately lead to their unique determination or impossibility of realization, and hence correspondingly to a unique conclusion or impossibility of derivation. Since the sum of the ranks of the fixed g 's and the fixed number of operational variables of a premise must equal the fixed rank of the assertion that premise is to be identified with, only a finite number of such hypotheses are admissible, and all can be uniformly tried out. In practice, a system in canonical form will usually be so constructed that a given assertion can be written in the form of a given premise in one and only one way, if at all. This uniqueness is automatically achieved by systems using the parenthesis notation, and is, of course, obviously attained in the systems in normal form about to be mentioned.

the production whenever it contains the enunciations represented by the several premises of the production.

A very special case of the canonical form is what we term the normal form. A system in canonical form will be said to be in *normal form* if it has but one primitive assertion, and, each of its productions is in the form

$$\begin{array}{c} g P \\ \text{produces} \\ Pg'. \end{array}$$

The main purpose of the present paper is to demonstrate that every system in canonical form can formally be reduced to a system in normal form. The two forms may therefore in fact be said to be *equipotent*. More precisely, we prove the following

THEOREM. *Given a system in canonical form with primitive letters a_1, a_2, \dots, a_μ , a system in normal form with primitive letters $a_1, a_2, \dots, a_\mu, a'_1, a'_2, \dots, a'_{\mu'}$ can be set up such that the assertions of the system in canonical form are exactly those assertions of the system in normal form which involve no other letters than a_1, a_2, \dots, a_μ .*

As a result of this theorem the decision problem for a system in canonical form is reduced to the decision problem for the corresponding system in normal form. For an enunciation of the former system is an assertion when and only when it is an assertion of the latter system. Hence any procedure which could effectively determine for an arbitrary enunciation of the system in normal form whether it is or is not an assertion thereof would automatically do the same for the system in canonical form. Now by methods such as those referred to in the opening sentence of this introduction, it can be shown that the problem of determining for an arbitrary well-formed formula in the λ -calculus of Church whether it has or has not a normal form (Church)⁵ can be reduced to the decision problem for a particular system in our canonical form. While Church has proved the above problem unsolvable in a certain technical sense, in the interest of economy we invoke his identification of λ -definability with effective calculability to conclude that as a result the decision problem for that particular system in canonical form, and hence for the class of systems in canonical form, is unsolvable. We are thus led to the more surprising result

⁵ Alonzo Church, "An unsolvable problem of elementary number theory," *American Journal of Mathematics*, vol. 58 (1936), pp. 345-363.

that there can be no effective procedure for determining for an arbitrary system in normal form and arbitrary enunciation thereof whether that enunciation is or is not an assertion of the system. That is, the decision problem is unsolvable for the class of normal systems, and indeed, by the previous argument, for a certain particular one of them.⁶

The present paper is not the place to review the reasons why the equivalent mathematical definitions of combinatory solvability based on the technical concepts of λ -definability, general recursive function, and computability⁷ can confidently be accepted as being the complete equivalent of combinatory solvability in the intuitive sense. Granting the initial establishment of the unsolvability of a particular decision problem by virtue of its being directly coextensive with the technical definition of solvability adopted, the chief method of establishing the unsolvability of further removed decision problems is by reducing the known unsolvable problem, by more or less ingenious formal devices, to those other problems.⁸ Our reduction of the decision problem for the complicated canonical form to that of the simple normal form illustrates this in some measure. And it may be that because of its formal simplicity, the normal form may lend itself more readily to representation in specialized mathematical developments, and the unsolvability of its decision problem thus lead to the unsolvability of various hitherto unsolved decision problems of classical mathematics.

Of more immediate promise is the fact that the concepts of the present paper, with the help of its basic theorem, easily lead to an independent approach to unsolvable problems which may be far simpler than, say, the λ -calculus of Church. In this connection we may note that if we define a *normal set* of

⁶ Absolutely unsolvable, that is, to use a phrase due to Church. By contrast, the undecidable propositions of Gödel's epoch making paper of 1931 (see footnote 7) are but relatively undecidable, the very proof of their undecidability in the given logic leading to an extension of that logic in which they are, indeed, proved to be true. A fundamental problem is the question of the existence of absolutely undecidable propositions, that is, propositions which in some *a priori* fashion can be said to have a determined truth-value, and yet cannot be proved or disproved by any valid logic.

⁷ For the first two see the paper referred to in footnote 5, for the third see A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* (2), vol. 42 (1937), pp. 230-265. We might also add the writer's "Finite combinatory processes-formulation I," *Journal of Symbolic Logic*, vol. 1 (1936), pp. 103-105. The basic paper is, of course, that of Kurt Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I," *Monatshefte für Mathematik und Physik*, vol. 38 (1931), pp. 173-198.

⁸ A very important instance of such a reduction is Gödel's transformation of the iterative recursive proposition into the non-iterative arithmetical proposition.

sequences on a_1, a_2, \dots, a_μ as the set of assertions on those letters only of any system in normal form with primitive letters a_1, a_2, \dots, a_μ and a finite number of additional letters, and a *canonical set* similarly via a system in canonical form, then the above theorem has as an immediate consequence the

COROLLARY. *The class of canonical sets is identical with the class of normal sets.*

For every normal set is *ipso facto* a canonical set; while if a canonical set is given by a certain system in canonical form, the theorem shows that it is also given by the corresponding system in normal form, and hence is a normal set. Now the canonical form naturally lends itself to the generating of sets by the method of definition by induction, while redefining the resulting canonical sets as normal sets makes it easy to use them as building blocks in further constructions. As a result of this alternating use of the canonical form as method, normal set as object, the Church development is easily paralleled.⁹ And since at each step only normal sets of sequences are obtained, we are led to identify the intuitive concept of *generated set* with normal set for much the same reasons that led Church to identify effective calculability with λ -definability. Under this identification, the intuitive concept of a *solvable set* of sequences on a_1, a_2, \dots, a_μ , i. e., one for which there is an effective procedure for determining whether a given sequence on those letters is or is not in the set, becomes precisely the *binormal set*, i. e., a set such that both it, and its complement with respect to the set of all finite sequences on a_1, a_2, \dots, a_μ , are normal. The resulting definition of solvability then easily leads to the unsolvability of the decision problem for the class of normal systems, as well as for a particular one of them. We may note this interchange of primary and secondary concept as compared with the Church development; for normal set corresponds to recursively enumerable set, binormal set to (general) recursive set.¹⁰

⁹ More completely, the Gödel, Church, Kleene, Rosser development.

¹⁰ While this equivalence undoubtedly follows from the reduction of the λ -calculus to a system in normal form, it would probably be more easily established by way of Turing's concept of computability. A few initial properties of normal and binormal sets may here be noted. With the ordinary Boolean operations on classes in question, the class of all normal sets constitutes a (distributive) lattice, of all binormal sets, a Boolean ring, of all binormal sets on a given finite set of letters, a Boolean algebra. Every infinite normal set contains an infinite binormal set. Query: Does there exist an infinite set which is the complement of a normal set, relative to the given set of letters, and does not contain an infinite binormal set? [Added in proof: yes]. There is no theoretical loss of generality in restricting ourselves to normal sets on a single letter

Before turning to the proof of our basic theorem given in the next section, we wish to mention a further transformation of the normal form which is of interest for its juxtaposition of the solvable and unsolvable, and state a problem which largely determined the direction taken by the reductions of the next section, and may offer further opportunities for unsolvability proofs. By making the question of whether a given sequence has a certain succession of primitive letters at one end depend on a related sequence having a corresponding succession of letters at the other end, the following result can be proved. Given any system in normal form on primitive letters a_1, a_2, \dots, a_μ , an enunciation P thereof will be an assertion when and only when $\alpha P \alpha$ is an assertion in a corresponding effectively derivable system in canonical form on letters $a_1, a_2, \dots, a_\mu, \alpha, a'_1, a'_2, \dots, a'_\mu$, having a finite number of primitive assertions, and a finite number of operations of the following forms,

$$\begin{array}{llll}
 gP & Pg & g_1P & Pg_1 \\
 \text{produces} & \text{produces} & g_2P & Pg_2 \\
 g'P , & Pg' , & \text{produce} & \text{produce} \\
 & & g'P , & Pg'.
 \end{array}$$

A solution of the decision problem for the derived system then immediately yields a solution of the decision problem for the given system. Much of the simplicity of the normal form is thus given up in order to have the g' of the conclusion of each production on the same side of the operational variable as the g , or now g' s, of the premise, or premises. But it is this very fact that leads to a solution of the decision problem for certain classes of these systems given *ab initio*. Indeed, for those systems in which all the productions have the g 's on the same side a solution of the decision problem follows almost immediately from the solution of a decision problem given by the writer in a former paper.¹¹ This solution has been extended by the writer to those of the above systems having only first order productions, i. e., productions with but one premise, and it has at least been seen by the writer how to extend these

a —normal sets of natural numbers essentially. The following is then probably the analogue for normal systems, i. e., systems in normal form, of Turing's universal computing machine. A fixed finite set of normal operations involving a and a single additional letter b can be set up such that by varying a single primitive assertion on a and b all normal sets on a are obtained.

¹¹ "On a simple class of deductive systems," abstract, *Bulletin of the American Mathematical Society*, vol. 27 (1921), pp. 396-7. The systems in question are those systems of the formulation referred to in footnote 2 whose primitive functions are all functions of one argument.

solutions to those of the above systems in which only the second order productions are restricted to having their g 's all on the same side, e. g., to systems having only productions of the first three of the above four types. For the class of all of the above systems, as indeed for a certain particular one of them, the decision problem is of course unsolvable as a consequence of the corresponding result for systems in normal form.

The problem referred to above takes two related forms. Both forms employ the following "tag" operations as we shall call them. Given a positive integer v , and μ symbols which may be taken to be $0, 1, \dots, \mu - 1$, we associate with each of the μ symbols a finite sequence of these symbols as follows.

$$\begin{aligned} 0 &\rightarrow a_{0,1} a_{0,2} \cdots a_{0,v_0} \\ 1 &\rightarrow a_{1,1} a_{1,2} \cdots a_{1,v_1} \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \mu - 1 &\rightarrow a_{\mu-1,1} a_{\mu-1,2} \cdots a_{\mu-1,v_{\mu-1}} \end{aligned}$$

It is understood that in each sequence the same symbol may occur several times, and that a particular associated sequence may be null. Given any non-null sequence

$$B = b_1 b_2 \cdots b_l$$

on the symbols $0, 1, \dots, \mu - 1$, a unique derived sequence B' on those symbols is determined as follows. To the right end of B adjoin the sequence associated with b_1 , the first symbol of B , and from the left end of the resulting augmented sequence remove the first v symbols—all if there be less than v symbols. Starting with a given tag operation, and a given sequence A on its primitive symbols, we can then iterate the tag operation to yield $A_1 = A, A_2 = A'_1, A_3 = A'_2, \dots$, the process terminating when and only when the null sequence is thus obtained. The first form of the problem of "tag" for a given tag operation is then to find an effective procedure for determining of an arbitrary initial sequence whether the above iterative process does or does not terminate.¹² In the second form of the problem we assume both the tag operation and the

¹² In an earlier formulation of this problem we merely checked off each successive v -th letter of the sequence starting with b_1 , at the same time adding the corresponding associated sequences to the right end of the sequence. Whether the iterative process terminated or not then depended on whether the constantly advancing check mark did or did not overtake the monotonically advancing right end of the sequence, whence the suggestive name of "tag" given the problem by B. P. Gill.

initial sequence to be given, and ask for an effective procedure for determining of an arbitrary sequence whether it is or is not one of the sequences obtained from the given sequence by the iteration of the given tag operation.

The first form of the problem of tag was intensely studied by the writer.¹³ Extended by the further dichotomy of the non-terminating cases to the periodic (sequences bounded), and divergent (sequences unbounded), the problem was completely solved for all cases in which both μ and ν are 2. But little real progress can be reported for μ or ν greater than 2, the problem for such a simple basis as $0 \rightarrow 00, 1 \rightarrow 1101, \nu = 3$ having proved intractable.¹⁴ In its second form the problem is almost the decision problem for a special type of normal system. In fact, we may define a *monogenic normal system* as one in which the g 's of the premises form a *complete set*, i. e., a set g_1, g_2, \dots, g_k such that each of the sequences of length equal to the maximum length ν of the g 's can be written in the form $g_i P$ for one and only one i . Except for the tag operation being applicable to sequences of length less than ν , a system of tag in its second form is then a monogenic normal system in which the g 's constitute all of the μ^ν sequences in question, while the corresponding g 's are identical for all g 's having the same initial symbol.

For a given tag operation the solution of the first form of the problem of tag probably leads to the solution of the second form of the problem. This is immediately so for those initial sequences which lead to termination or periodicity; and, while the mere hypothesis of divergence seems insufficient to guarantee a corresponding solution, the actual proof of divergence would probably make the definition of divergence effective, in which case the solution of the second form of the problem would again follow. For the writer, the little progress made in the solution of the first form of the problem make both forms, in their full generality, candidates for unsolvability proofs. Even more so, therefore, the decision problem for the class of monogenic normal systems. Among normal systems there is a "complete normal system" to which every normal system can be reduced, and whose decision problem is consequently unsolvable. A most interesting situation would obtain should it be shown that the complete normal system cannot be reduced to a monogenic normal system,

¹³ During the writer's tenure of a Procter fellowship at Princeton University, 1920-1921.

¹⁴ Numerous initial sequences actually tried led in each case to termination or periodicity, usually the latter.

while the decision problem for the class of monogenic normal systems is otherwise shown to be unsolvable.¹⁵

2. Reduction of the canonical form to the normal form. Our reduction of the canonical form to the normal form is the result of four successive reductions.¹⁶ Each of these reductions yields a formulation which is included in the preceding formulation, but eliminates some formal complexities allowed in that preceding formulation. For a given system this simplification is achieved at the expense of an increase in the number of primitive letters employed, and in the number of productions appearing in its bases.

Our first reduction of an arbitrary system in canonical form is to one in which there is but one primitive assertion, and in which each production involves but a single premise. That one premise, and corresponding conclusion, however, may have all the complexity allowed for in the general canonical form. The general plan of the method involved is to formally introduce the logical products of arbitrary assertions of the given system, and operate within such products.

Let then S_1 be a system in canonical form with primitive letters a_1, a_2, \dots, a_μ , S_2 the system, about to be described, to which S_1 is to be reduced. With a_1, a_2, \dots, a_μ also primitive letters of S_2 , introduce two new primitive letters u and a_0 in S_2 . When the logical product of assertions, $P_1, P_2, P_3, \dots, P_n$ of S_1 is asserted in S_2 , it will appear in the form

$$ua_0P_1a_0uua_0P_2a_0uuua_0P_3a_0\dots u \underbrace{\dots u}_{n} a_0P_na_0u \underbrace{\dots u}_{n+1},$$

each P being flanked on either side by a_0 . The separating u sequences are thus made to increase left to right by one each to enable us by the mere form of a premise to insure that certain operational variables therein must represent assertions of S_1 , if that premise is to be identified with an assertion in S_2 . The final basis for S_2 will reveal the necessary source of that insurance, i. e., that the only assertions of S_2 involving u are those of the above form. We shall call such an expression a product, the P 's therein the factors of the product.

¹⁵ It is easy to talk of obtaining a property of all normal solutions which could not be satisfied by a solution of a given decision problem; but this is probably equivalent to finding one of those not immediately obvious effectively calculable invariants of conversion which Church reports as still unfound in 1936. (See p. 358 of the paper referred to in footnote 5).

¹⁶ Not counting a minor reduction needed to validate the last of the four.

We first introduce in the basis of S_2 certain productions whereby from the assertion of a product may be obtained the assertion of all products obtainable from the given product by a mere permutation of its factors. It suffices to allow for the interchange of any two consecutive factors. For the first two factors of a product this is achieved by

$$ua_0P_1a_0uua_0P_2a_0uuua_0S \text{ produces } ua_0P_2a_0uua_0P_1a_0uuua_0S,$$

our system being so devised that each product appearing therein has at least three factors. This allows that last a_0 to be assumed. The u, uu, uuu of the premise are then "maximal" u sequences. As these u sequences differ by one each, P_1 and P_2 must be free from u 's, and hence, by our induction, be the two initial factors of the product. The interchange then results via the production. For two consecutive factors neither starting nor ending the product the result is achieved by

$$Ra_0uQua_0P_1a_0uQuua_0P_2a_0uQuuuua_0S$$

produces

$$Ra_0uQua_0P_2a_0uQuua_0P_1a_0uQuuuua_0S.$$

Here Q must consist of u 's only. For otherwise a_0uQua_0 and a_0uQuua_0 would have their initial a_0 's followed by identical maximal u sequences. The u sequences $uQu, uQuu$ and $uQuuu$ are then maximal, and differ in length by one each. P_1 and P_2 again then are consecutive factors of the product. Finally, for two factors ending a product, the last production, rewritten with a_0S deleted, suffices.

The next production to be added to the bases of S_2 allows us to pass from the assertion of a product to the assertion of the first factor of a product, and hence, with the help of the previous three productions, to the assertion of an arbitrary factor of a product. The production is simply

$$ua_0Pa_0uua_0R \text{ produces } P.$$

In translating the operations of S_1 into operations within products of S_2 , we allow for passing from a product whose initial factors can be identified with the premises of an S_1 operation, to that product with the conclusion of the S_1 operation as additional factor. That additional factor must end the new product so as not to disturb the progression of the maximal u sequences. Let " G_1, G_2, \dots, G_k produce G " represent any one of the S_1 operations. Let H represent

$$ua_0G_1a_0uua_0G_2a_0\dots\underbrace{u\dots u}_{k}a_0G_k\overbrace{a_0u\dots u}^{k+1}$$

Then the corresponding S_2 operation may be represented by

$$Ha_0Ra_0uQua_0Sa_0uQuu \text{ produces}$$

$$Ha_0Ra_0uQua_0Sa_0uQuua_0Ga_0uQuuu.$$

Note that the operational variables of this production are those of the S_1 production, and Q, R, S . Since each operational variable in G occurs in at least one of the G_i 's, our new production will indeed have the same operational variables in its conclusion as in its premise. The portion of the premise following H insures that Q consists of u 's only. This, with the form of H , insures that G_1, G_2, \dots, G_k are determined factors of the premise and G of the conclusion. Hence our transformation of the S_1 production is valid. The additional operational variables R and S require an assertion to which this production is applied to have at least $k+2$ factors, a requirement secured below. Of course, the basis of S_2 is to have the correspondent of each of the operations in the basis of S_1 .

With S_1 having κ productions, the above $\kappa+4$ productions constitute all of the productions in the basis of S_2 . Its sole primitive assertion is then formed as follows. Let L be the largest number of premises occurring in any production of S_1 . If S_1 has λ primitive assertions, let each be repeated L times to give λL sequences each involving no other letters than a_1, \dots, a_μ . If $\lambda L < L+2$, or $\lambda L < 3$, again duplicate one of these sequences the one or two times needed to avoid these inequalities. If then k_1, k_2, \dots, k_M are these duplicated primitive assertions of S_1 , the primitive assertion of S_2 will be their product

$$ua_0k_1a_0uua_0k_2a_0\dots u\dots u \underbrace{a_0k_Ma_0u\dots u}_{M+1}$$

Now it is readily proved by induction that if at a certain point of the process for obtaining assertions in S_1 a certain finite set of assertions has been obtained, then there will be asserted in S_2 a product among whose factors are each of the above assertions repeated L times. For the primitive assertions of S_1 , this is insured by the primitive assertion of S_2 . Assume it to be true for the deductive process in S_1 at an arbitrary point, let P_2 be the corresponding assertion in S_2 , P_1 the next assertion obtained in $S_1, P_{11}, P_{12}, \dots, P_{1k}$ the premises of the production of S_1 yielding conclusion P_1 . Then each P_{1j} appears as a factor of P_2 indeed L times at least. Hence from P_2 , by the first three productions of S_2 , an assertion P'_2 can be obtained in which the first k factors are $P_{11}, P_{12}, \dots, P_{1k}$ respectively, whatever repetitions may occur among those P 's. The production of S_2 corresponding to the one of S_1 in

question will then add P_1 as factor to P'_2 . Mere repetition of the application of this production will then yield P''_2 , which will be P'_2 with L additional factors equal to P_1 . The induction is thus established. It follows that for each assertion P_1 in S_1 there will be an assertion P_2 in S_2 having P_1 as factor. By the first three productions of S_2 this factor can be made the first factor of an assertion in S_2 , and hence, by the fourth production of S_2 , P_1 itself will be an assertion of S_2 . That is, every assertion of S_1 is an assertion of S_2 . Our basis for S_2 shows that the only other assertions of S_2 are products of assertions of S_1 , and so not wholly written on the letters of S_1 . Hence, an enunciation of S_1 is an assertion of S_1 when and only when it is an assertion of S_2 , whence the reduction of S_1 to S_2 .

In our second reduction of the canonical form the productions, all with single premises by the previous reduction, now take the more special form

$$g_1 P_1 g_2 P_2 \cdots g_m P_m g_{m+1}$$

produces

$$\bar{g}_1 P_1 \bar{g}_2 P_2 \cdots \bar{g}_m P_m \bar{g}_{m+1}$$

where, however, m , and of course the g 's, may vary from operation to operation. By contrast, in the previous productions P 's could be repeated, have different arrangements in premise and conclusion, and in part be missing from the conclusion while present in the premise.

Again let the primitive letters of the given system be symbolized a_1, a_2, \dots, a_μ . Let its i -th production be

$$g_1 P_{j_1} g_2 P_{j_2} \cdots g_m P_{j_m} g_{m+1}$$

produces

$$g'_1 P_{j'_1} g'_2 P_{j'_2} \cdots g'_{m'} P_{j'_{m'}} g'_{m'+1}$$

where it is understood that each letter except P has i for additional subscript. The subscripts of the P 's need not be distinct in premise or conclusion, while the different subscripts of the P 's in the conclusion all appear in the premise. However, the letter P occurs exactly $m + m'$ times in the production.

We introduce a new primitive letter u , and for each such production two new primitive letters v_i, w_i . In obtaining the effect of the i -th production we shall, as above, leave this subscript i understood. v_i will be used in passing from an assertion involving a 's only that could be the premise of the i -th production to one which has both that premise and corresponding conclusion recognizable within it; w_i in passing from such a composite assertion to the

desired conclusion only. The efficacy of our method will depend on each assertion in the new system which involves v or w having that letter only at the beginning of the assertion, and in the first case always involving exactly $2m + m'$ u 's, in the second, m' u 's. Our new productions will in every case explicitly exhibit this v and $2m + m'$ u 's, or w and m' u 's, so that we can be sure that in their application the operational variables can represent sequences of a 's only. Except for a minor preliminary type, all of our "v-assertions" will be in the form

$$vug_1P_1uQ_1ug_2P_2uQ_2 \cdots ug_mP_muQ_mg_{m+1}ug'_1Q_{m+1}ug'_2Q_{m+2} \cdots ug'_{m'}Q_{m+m'}g'_{m'+1},$$

and when so asserted will have the following properties. The sequence of a 's $g_1P_1Q_1g_2P_2Q_2 \cdots g_mP_muQ_mg_{m+1}$ is an assertion of the given, and indeed new system, while the sequences of a 's $Q_1, Q_2, \dots, Q_m, Q_{m+1}, \dots, Q_{m+m'}$ can, in order, be identified with $P_{j_1}, P_{j_2}, \dots, P_{j_m}, P_{j_{m+1}}, \dots, P_{j_{m'+1}}$, that is, any two Q 's corresponding to P 's with identical subscripts are equal. Note that with all $2m + m'$ u 's exhibited, the g 's being given, the P 's and Q 's of such an assertion are uniquely identifiable in the assertion. Our method depends on the fact that when such an assertion is obtained in which the P 's are null, then, due to the equalities forced on the Q 's, $g_1Q_1g_2Q_2 \cdots g_mQ_mg_{m+1}$ becomes an assertion on a 's only that can be identified with the premise of the i -th production of the given system, and hence $g'_1Q_{m+1}g'_2Q_{m+2} \cdots g'_{m'}Q_{m+m'}g'_{m'+1}$ an expression on a 's only that will be the corresponding conclusion. Of course, each production about to be described is directly seen to be in the desired newly simplified form.

Since a null assertion has been excluded from our systems, each assertion of the given system is of the form a_jP , $j = 1, 2, \dots, \mu$. The productions

$$a_jP \text{ produces } va_jPu \dots u$$

with $2m + m'$ u 's in $u \dots u$ changes each "a-assertion," i. e., assertion involving a 's only, into what we shall call the intermediate v form. As all other assertions of our new system will begin with v or w , these productions will be inapplicable to them. If now an a-assertion can be the premise of the i -th production, its intermediate v form will be put into primary v form, or just v form, by the production

$$vg_1P_1g_2P_2 \cdots g_mP_muQ_mg_{m+1}u \dots u$$

produces

$$vug_1P_1uug_2P_2uu \dots g_mP_muQ_mg_{m+1}ug'_1ug'_2u \dots ug'_{m'}g'_{m'+1}.$$

Of course this production may be applicable without the P 's being identifiable with those of the premise of the i -th production. But, comparing this conclusion with our general v form, we see that it satisfies the requirement thereof with all Q 's null. Now any set of a -sequences that could be identified with the $P_{j_1}, P_{j_2}, \dots, P_{j_m}, P_{j'_1}, \dots, P_{j'_{m'}}$ of the i -th production can be built up as follows. Start with the set of null sequences. Let $Q_1, Q_2, \dots, Q_m, Q_{m+1}, \dots, Q_{m+m'}$ be any such derived set of a -sequences. Let $Q_{j_1}, Q_{j_2}, \dots, Q_{j_v}$, j 's increasing, be any subset thereof corresponding to all P 's with subscripts equal to a given subscript, a_j any one of the primitive a 's. Then $\dots, a_j Q_{j_1}, \dots, a_j Q_{j_2}, \dots, a_j Q_{j_v}, \dots$, all other Q 's unchanged, will also be such a set of a -sequences. Rewrite the subscript sequence j_1, j_2, \dots, j_v in the form $j_1, \dots, j_\lambda, j_{\lambda+1}, \dots, j_v$ so that $j_\lambda \leq m$, $j_{\lambda+1} > m$, and let $j_{\lambda+1} - m = j'_1, \dots, j_v - m = j'_{v'}$. Of course we may have $\lambda = v$. Now for each such choice of original P subscript, and each a_j , introduce the production

$$v \dots ug_{j_1} P_{j_1} a_j u Q_{j_1} \dots ug_\lambda P_{j_\lambda} a_j u Q_{j_\lambda} \dots ug'_{j'_1} Q_{j_{\lambda+1}} \dots ug'_{j'_{v'}} Q_{j_v} \dots$$

produces

$$v \dots ug_{j_1} P_{j_1} ua_j Q_{j_1} \dots ug_\lambda P_{j_\lambda} ua_j Q_{j_\lambda} \dots ug'_{j'_1} a_j Q_{j_{\lambda+1}} \dots ug'_{j'_{v'}} a_j Q_{j_v} \dots$$

all of the rest of both premise and conclusion being as in the type v form above. Such a production will then change a valid v form into a valid v form, the effect being however to "drain" the P 's of such a form and "swell" the Q 's. If then an assertion of the given system can be put in the form of the premise of the i -th production, the corresponding intermediate v form will pass into a v form such that successive application of the above productions will completely drain the P 's thereof; and, indeed, conversely. This marks the end of the first half of the passage from a -assertion to a -assertion in the new system. While the second half could be set up by means of similar w productions in reverse, with interchange of emphasis on premise and conclusion of the i -th production, the following method is simpler. With P 's all null, the v form determines the desired a -conclusion as described above. The w forms, about to be introduced, each have exactly m' u 's all explicitly appearing in the productions. From such a v form with P 's all null the first w form is obtained via

$$vug_1 u Q_1 ug_2 u Q_2 \dots ug_m u Q_m g_{m+1} ug'_{j_1} Q_{m+1} ug'_{j_2} Q_{m+2} \dots ug'_{j_{m'}} Q_{m+m'} g'_{j_{m'+1}}$$

produces

$$wg_1 Q_1 g_2 Q_2 \dots g_m Q_m g_{m+1} ug'_{j_1} Q_{m+1} ug'_{j_2} Q_{m+2} \dots ug'_{j_{m'}} Q_{m+m'} g'_{j_{m'+1}}.$$

We can now get rid of the no longer interesting part of this w form, i. e., the part between w and the first u thereof, by the μ productions

$$wa_j P u g'_{j1} P_1 u g'_{j2} P_2 \cdots u g'_{jm'} P_{m'} g'_{m'+1}$$

produces

$$w P u g'_{j1} P_1 u g'_{j2} P_2 \cdots u g'_{jm'} P_{m'} g'_{m'+1}$$

iteratively applied till letter by letter what was the original a -assertion disappears. The desired a -conclusion then would be obtained via

$$w u g'_{j1} P_1 u g'_{j2} P_2 \cdots u g'_{jm'} P_{m'} g'_{m'+1}$$

produces

$$g'_{j1} P_1 g'_{j2} P_2 \cdots g'_{jm'} P_{m'} g'_{m'+1}.$$

Our final system will then be on the primitive letters $a_1, \dots, a_\mu, u, v_1, w_1, v_2, w_2, \dots, v_\kappa, w_\kappa$, κ being the number of productions of the given system. The one primitive assertion of the new system will be the one primitive assertion of the given system, the productions of the new system, all of the above productions for each of the κ productions of the given system. Our above analysis then easily shows that the assertions of the new system involving no other letters than a_1, \dots, a_μ are exactly the assertions of the given system, and the desired reduction has been effected.

Our third and penultimate simplifying reduction of the canonical form is to one where the operations are of the form

$$g_1 P g_2$$

produces

$$\bar{g}_1 P \bar{g}_2,$$

i. e., involve but a single operational variable. Again let a system in the previous simplified form have primitive letters a_1, a_2, \dots, a_μ , and κ operations, the number of P 's in the premise, and hence conclusion, of the i -th operation being m_i . For the i -th operation, with $i = 1, 2, \dots, \kappa$, and each primitive letter a_j we introduce $2m_i + 1$ new primitive letters $a'_{ji}, a''_{ji}, \dots, a^{(2m_i)}_{ji}, a^{(2m_i+1)}_{ji}$. We also introduce the primitive letter a_{0i} and its $2m_i + 1$ primed equivalents. With one such operation in mind at a time we shall, as above, omit the extra subscript i . Apart from the use of a_0 and $a^{(j)}_0$'s, needed to take care of g 's or P 's that are null, the essence of our method is to pass from an a -assertion in the form $g_1 P_1 g_2 P_2 \cdots g_m P_m g_{m+1}$ to an assertion $g'_1 g''_2 \cdots g^{(2m+1)}_{m+1} P''_1 P^{IV}_2 \cdots P^{(2m)}_m$ where the superscript k say indicates that each a_j in the corresponding expression is here written $a_j^{(k)}$. As a result our premise will now have the form gP with $g = g'_1 g''_2 \cdots g^{(2m+1)}_{m+1}$, $P = P''_1 P^{IV}_2 \cdots P^{(2m)}_m$.

In detail, we first introduce μ productions

$$a_j P \text{ produces } a_0 a_j P, \quad j = 1, 2, \dots, \mu,$$

which will be applicable in fact only to assertions on a_1, a_2, \dots, a_μ , and changes any such assertion Q into $a_0 Q$. We then introduce a finite series of finite sets of production depending in number on m and μ . The first set has the one production

$$a_0 g_1 P \text{ produces } a'_0 g'_1 a_0 P a''_0.$$

Inductively let the conclusion of the sole production in the $(2k - 1)$ -st set be in the form $G_k a_0 P a^{(2k)}_0$. Then the $(2k)$ -th set has the μ productions

$$G_k a_0 a_j P \text{ produces } G_k a_0 P a^{(2k)}_j, \quad j = 1, 2, \dots, \mu,$$

the $(2k + 1)$ -st set the sole production

$$G_k a_0 g_{k+1} P \text{ produces } G_k a^{(2k+1)}_0 g^{(2k+1)}_{k+1} a_0 P a^{(2k+1)}_0.$$

This is to hold for $1 \leq k < m$, while for $k = m$ the sole production of the $(2m + 1)$ -st set is to be

$$G_m a_0 g_{m+1} a''_0 P \text{ produces } G_m a^{(2m+1)}_0 g^{(2m+1)}_{m+1} a''_0 P.$$

We then readily see that starting with an assertion on a_1, \dots, a_μ in the form $g_1 P_1 g_2 P_2 \dots g_m P_m g_{m+1}$, one can, with the aid of these productions, obtain as an assertion

$$a'_0 g'_1 a''_0 g'''_2 \dots a^{(2m-1)}_0 g^{(2m-1)}_m a^{(2m+1)}_0 g^{(2m+1)}_{m+1} a''_0 P''_1 a^{IV}_0 P^{IV}_2 \dots a^{(2m)}_0 P^{(2m)}_m.$$

Furthermore, note that starting with an assertion on a_1, a_2, \dots, a_μ , flanked on the left by a_0 as above, one can apply the above operations only in the following order, if at all. First, the sole operation of the first set; and inductively, if the operation in the $(2k - 1)$ -st set has last been applied, the next applicable operation can only be an operation in the $2k$ -th set or the operation in the $(2k + 1)$ -st set, if an operation in the $(2k)$ -th set has last been applied, the next applicable operation can only be an operation in the same set, or the operation in the next set. Furthermore, the last operation in its premise explicitly indicates the a''_0 , first introduced into an assertion only as a result of the first operation. It readily follows that if the last operation does enter into a possible sequence of operations, the conclusion thereof can have no letter a_j in it without a superscript. The entire given assertion has thus been translated; and it is readily seen that that last assertion, and hence given assertion, are and can be put in the forms above given.

The actual correspondent of the original i -th operation in translated form may then be written simply

$$a'_0 g'_1 a'''_0 g'''_2 \cdots a^{(2m-1)}_0 g^{(2m-1)}_m a^{(2m+1)}_0 g^{(2m+1)}_{m+1} P$$

produces

$$a'_0 \bar{g}'_1 a'''_0 \bar{g}'''_2 \cdots a^{(2m-1)}_0 \bar{g}^{(2m-1)}_m a^{(2m+1)}_0 \bar{g}^{(2m+1)}_{m+1} P;$$

and the passage from this translated conclusion to the actual conclusion can be effected by a set of productions the reverse of those above given. That is, in each of the above productions prior to the actual correspondent of the i -th production replace all g_j 's by \bar{g}_j 's, and *interchange hypothesis and conclusion*. The resulting productions then clearly suffice to yield the conclusion yielded by the original i -th production. True, the complete set of productions thus set up to take the place of the original i -th production may now allow other paths than from assertion on a_1, \dots, a_μ , down the first group of productions, through the intermediate production, and up the second group of productions to new assertion on a_1, \dots, a_μ .¹⁷ But it is readily seen that any departures from this progression merely constitute unravelings of parts of such a progression, or, apart from such unravelings, constitute shortcuts of valid full progressions of this type. Since, furthermore, one can change the set of productions one is working with only when an assertion on a_1, \dots, a_μ alone is obtained, the validity of our reduction follows.

Our final reduction is to a system whose operations are in the form

$$gP$$

produces

$$Pg'.$$

The present method assumes that in the productions of the previous system, all in the form

$$g_1 Pg_2$$

produces

$$g'_1 Pg'_2,$$

g_1 and g_2 are never null. We therefore actually first need the following preliminary reduction. Introduce a new primitive letter a_0 , and if h is the sole primitive assertion of the given system let $a_0 h a_0$ be the sole primitive assertion of the new system. Replace each of the above operations of the given system by

$$a_0 g_1 Pg_2 a_0 \quad \text{produces} \quad a_0 g'_1 Pg'_2 a_0$$

and finally add the production

¹⁷ This could have been avoided say by the v, w method used earlier.

a_0Pa_0 produces P .

Except for the last production the new system may be said to be simply isomorphic with the old, P being an assertion in the given system when and only when a_0Pa_0 is an assertion in the new system. The last operation then merely recovers the assertions of the given system. Note that even that last operation is in the desired form with neither g_1 nor g_2 null.

Assume then that such is our given system with primitive letters again a_1, a_2, \dots, a_μ . We introduce new primitive letters $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_\mu$, and "translating productions"

a_jP produces $P\bar{a}_j$, \bar{a}_jP produces Pa_j , $j = 1, 2, \dots, \mu$.

Starting with an assertion of the form $a_{i_1} \dots a_{i_j} a_{i_{j+1}} \dots a_{i_n}$, these productions will yield only assertions of the form $a_{i_{j+1}} \dots a_{i_n} \bar{a}_{i_1} \dots \bar{a}_{i_j}$, $\bar{a}_{i_1} \dots \bar{a}_{i_j} \bar{a}_{i_{j+1}} \dots \bar{a}_{i_n} a_{i_{j+1}} \dots a_{i_n}$, in addition to the original assertion. Only one of these $2n$ distinct forms consists wholly of unbarred letters, i.e., the original form, while continued application of the above operations merely keeps deriving these $2n$ "equivalent forms" cyclically, so that anyone can thus be obtained from any other.

Our reduction will then be effected if for each operation " g_1Pg_2 produces $g'_1Pg'_2$ " of the given system we introduce in the new system the operation

\tilde{g}_2g_1P
produces
 $Pg'_2\tilde{g}'_1$,

where \tilde{g}_2 , for example, is g_2 with each letter replaced by the corresponding barred letter. Of course, the one primitive assertion of the given system is also the one primitive assertion of the new system. Note that if at any point an assertion without barred letters appears, then if it can be written g_1Pg_2 , the first given translating operations derive from it \tilde{g}_2g_1P , hence the above yields $Pg'_2\tilde{g}'_1$, and so finally $g'_1Pg'_2$ is obtained as desired. That is, the new system contains all of the assertions of the given system. It further follows that the assertions of the new system consist only of the assertions of the given system and their equivalents. For, proceeding inductively, this clearly remains true under the translating operations. Now suppose it is true of an assertion in the form \tilde{g}_2g_1P . Since \tilde{g}_2 and g_1 are not null, \tilde{g}_2g_1 alone exhibits a change from barred to unbarred letters. P therefore must consist of unbarred letters only. \tilde{g}_2g_1P is therefore a translation of the assertion g_1Pg_2 of the original system, and hence the conclusion $Pg'_2\tilde{g}'_1$ is a translation of $g'_1Pg'_2$, also an assertion of the original system. The desired reduction has thus been effected.

The original system in canonical form has thus been reduced to a system in normal form. At each stage in that reduction the primitive letters of the new system are the primitive letters of the preceding system and a finite number of additional letters, while the assertions of that preceding system are exactly those assertions of the new system which involve only primitive letters of the preceding system. The same is then true of the original system in canonical form and the final system in normal form, whence the theorem of the introduction.¹⁸

THE CITY COLLEGE,
COLLEGE OF THE CITY OF NEW YORK.

¹⁸ While the present paper is presented as a contribution to the literature as it now exists, its main development, that of Section 2, as well as the further transformation mentioned in the introduction, was obtained by the writer in substantially the form here given in the summer of 1921. The larger development of which this was a part is no longer essentially new, but may be worth a brief résumé. In the fall of 1920, starting with the formulation referred to in footnote 2 as canonical form *A*, the operation of substitution of that formulation was weakened in two successive stages to yield canonical forms *B* and *C*, the last only allowing any 1-1 replacement of variables by variables. It was first shown in detail that each of these forms was reducible to the other two, and then the project that led to the above changes of canonical form *A* was carried through, namely the reduction of what would now be termed the restricted functional calculus of *Principia Mathematica* to a particular system in canonical form *C*. Much of this work has since been found to be seriously in error, but easily corrected by the methods then employed. As a result of the last reduction it appeared obvious to the writer that all of *Principia Mathematica* could likewise be reduced to a system in canonical form *C*. In the summer of 1921, the intervening work on the problem of tag suggested the reduction of canonical form *C* to the canonical form of the present paper, and this reduction was followed by the successive reductions to normal form essentially as given in Section 2. The added methods thus revealed led us to conclude that not only *Principia Mathematica*, but any symbolic logic whose operations could effectively be reproduced in *Principia Mathematica*, and hence probably any (finitary) symbolic logic could be reduced to a system in canonical form, and consequently to a system in normal form. But now the entire direction of our thought, that of solving the decision problem for arbitrary systems, was reversed. Having noted the identity of canonical sets and normal sets referred to in the introduction, our last conclusion was transformed into the generalization that every generated set of sequences on a finite sets of letters was a normal set. The seeming counter example furnished by the diagonal method then led to an informal proof that the decision problem for the class of systems in normal form was unsolvable. In the early fall of 1921, the formal proof of this unsolvability, referred to in the introduction, was outlined, and led to the further conclusion that not only was every (finitary) symbolic logic incomplete relative to a certain fixed class of propositions (those stating that a given sequence was or was not an assertion in a given normal system) but that every such logic was extendable relative to that class of propositions. Since the earlier formal work made it seem obvious that the actual details of the outline could be supplied, the further efforts of the writer were directed towards establishing the universal validity of the basic identification of generated set with normal set.

RECURSIVELY ENUMERABLE SETS OF
POSITIVE INTEGERS AND THEIR
DECISION PROBLEMS

BY
E. L. POST

Reprinted from the
BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY
Vol. 50, No. 5, pp. 284-316
May, 1944

RECURSIVELY ENUMERABLE SETS OF POSITIVE INTEGERS AND THEIR DECISION PROBLEMS

EMIL L. POST

Introduction. Recent developments of symbolic logic have considerable importance for mathematics both with respect to its philosophy and practice. That mathematicians generally are oblivious to the importance of this work of Gödel, Church, Turing, Kleene, Rosser and others as it affects the subject of their own interest is in part due to the forbidding, diverse and alien formalisms in which this work is embodied. Yet, without such formalism, this pioneering work would lose most of its cogency. But apart from the question of importance, these formalisms bring to mathematics a new and precise mathematical concept, that of the general recursive function of Herbrand-Gödel-Kleene, or its proved equivalents in the developments of Church and Turing.¹ It is the purpose of this lecture to demonstrate by example that this concept admits of development into a mathematical theory much as the group concept has been developed into a theory of groups. Moreover, that stripped of its formalism, such a theory admits of an intuitive development which can be followed, if not indeed pursued, by a mathematician, layman though he be in this formal field. It is this intuitive development of a very limited portion of a sub-theory of the hoped for general theory that we present in this lecture. We must emphasize that, with a few exceptions explicitly so noted, we have obtained formal proofs of all the consequently mathematical theorems here developed informally. Yet the real mathematics involved must lie in the informal development. For in every instance the informal "proof" was first obtained; and once gotten, transforming it into the formal proof turned out to be a routine chore.²

We shall not here reproduce the formal definition of *recursive function of positive integers*. A simple example of such a function is an

An address presented before the New York meeting of the Society on February 26, 1944, by invitation of the Program Committee; received by the editors March 25, 1944.

¹ For "general recursive function" see [9] ([8] a prerequisite), [12] and [11]; for Church's " λ -defineability," [1] and [6]; for Turing's "computability," [24] and the writer's related [18]. To this may be added the writer's method of "canonical systems and normal sets" [19]. See pp. 39-42 and bibliography of [6] for a survey of the literature and further references. Numbers in brackets refer to the bibliography at the end of the paper.

² Our present formal proofs, while complete, will require drastic systematization and condensation prior to publication.

arbitrary polynomial $P(x_1, x_2, \dots, x_n)$, with say non-negative integral coefficients, and not identically zero. If the x 's are assigned arbitrary positive integral values expressed, for example, in the arabic notation, the algorithms for addition and multiplication in that notation enable us to calculate the corresponding positive integral value of the polynomial. That is, $P(x_1, x_2, \dots, x_n)$ is an *effectively calculable function of positive integers*. The importance of the technical concept recursive function derives from the overwhelming evidence that it is coextensive with the intuitive concept effectively calculable function.³

A set of positive integers is said to be *recursively enumerable* if there is a recursive function $f(x)$ of one positive integral variable whose values, for positive integral values of x , constitute the given set. The sequence $f(1), f(2), f(3), \dots$ is then said to be a *recursive enumeration* of the set. The corresponding intuitive concept is that of an *effectively enumerable* set of positive integers. To prepare us in part for our intuitive approach, consider the following three examples of recursively enumerable sets of positive integers.

- (a): $1^2, 2^2, 3^2, \dots$
- (b): $1, 2, 2^{1+2}, 2^{1+2+2^{1+2}}, \dots$
- (c): $1^2, 2^2, 3^2, \dots$
 $1^3, 2^3, 3^3, \dots$
 $1^4, 2^4, 3^4, \dots$
 $\vdots \quad \vdots \quad \vdots$
 $\vdots \quad \vdots \quad \vdots \quad \ddots$

In the first example, the set is given by a recursive enumeration thereof via the recursive function x^2 . In the second example, the set is generated in a linear sequence, each new element being effectively obtained from the elements previously generated, in this case by raising 2 to the power the sum of the preceding elements. The set is effectively enumerable, since the n th element of the sequence can be found, given n , by regenerating the sequence through its first n elements. In the third example, we rather imagine the positive integers $1, 2, 3, \dots$ generated in their natural order, and, as each positive integer n is generated, a corresponding process set up which generates n^2, n^3, n^4, \dots , all these to be in the set. Actually, the standard method for proving that an enumerable set of enumerable sets is enumerable yields an effective enumeration of the set.

³ See Kleene [13, footnote 2]. In the present paper, "recursive function" means "general recursive function."

Several more examples would have to be given to convey the writer's concept of a *generated set*, in the present instance of positive integers. Suffice it to say that each element of the set is at some time written down, and earmarked as belonging to the set, as a result of predetermined effective processes. It is understood that once an element is placed in the set, it stays there. The writer elsewhere has referred to a generalization which may be restated *every generated set of positive integers is recursively enumerable.*⁴ For comparison purposes this may be resolved into the two statements: every generated set is effectively enumerable, every effectively enumerable set of positive integers is recursively enumerable. The first of these statements is applicable to generated sets of arbitrary symbolic expressions; their converses are immediately seen to be true. We shall find the above concept and generalization very useful in our intuitive development. But while we shall frequently say, explicitly or implicitly, "set so and so of positive integers is a generated, and hence recursively enumerable set," as far as the present enterprise is concerned that is merely to mean "the set has intuitively been shown to be a generated set; it can indeed be proved to be recursively enumerable." Likewise for other identifications of informal concepts with corresponding mathematically defined formal concepts.

At a few points in our informal development we have to lean upon the formal development. The latter is actually yet another formalism, due to the writer [19] but proved completely equivalent to that of general recursive function. It will suffice to give the equivalent of "recursively enumerable set of positive integers" in this development.

A positive integer n is represented in the most primitive fashion by a succession $11 \dots 1$ of n strokes. For working purposes, we introduce the letter b , and consider "strings" of 1's and b 's such as $11b1bb1$. An operation on such strings such as " $b1bP$ produces $P1bb1$ " we term a normal operation. This particular normal operation is applicable only to strings starting with $b1b$, and the derived string is then obtained from the given string by first removing the initial $b1b$, and then tacking on $1bb1$ at the end. Thus $b1bb$ becomes $b1bb1$. " gP produces Pg' " is the form of an arbitrary normal operation. A system in normal form, or normal system, is given by an initial string A of 1's and b 's, and a finite set of normal operations " $g_i P$ produces Pg'_i ," $i = 1, 2, \dots, \mu$. The derived strings of the system are A and all strings obtainable from A by repeated applications of the μ normal

⁴ See [19, p. 201 and footnote 18]. In this connection note Kleene's use of the word "Thesis" in [14, p. 60]. We still feel that, ultimately, "Law" will best describe the situation [18].

operations. Each normal system uniquely defines a set, possibly null, of positive integers, namely the integers represented by those derived strings which are strings of 1's only. It can then be proved that every recursively enumerable set of positive integers is the set of positive integers defined by some normal system, and conversely.⁵ We here, as below, arbitrarily extend the concept recursively enumerable set to include the null set.

By the *basis* B of a normal system, and of the recursively enumerable set of positive integers it defines, we mean the string of letters and symbols here represented by

$$A; g_1 P \text{ produces } Pg'_1, \dots, g_\mu P \text{ produces } Pg'_\mu.$$

When meaningfully interpreted, B determines the normal system, and recursively enumerable set of positive integers, in question. Each basis is but a finite sequence of the symbols 1, b , P , the comma, semi-colon and the letters of the word "produces." The set of bases is therefore enumerably infinite, and can indeed be effectively generated in a sequence of distinct elements

$$O: \quad B_1, B_2, B_3, \dots$$

Since each B_i defines a unique recursively enumerable set of positive integers and each such set is defined by at least one B_i , O is also an ordering of all recursively enumerable sets of positive integers, though each set will indeed recur an infinite number of times in O . We may then say, in classical terms, that whereas there are 2^{\aleph_0} arbitrary sets of positive integers, there are but \aleph_0 recursively enumerable sets.

By the *decision problem* of a given set of positive integers we mean the problem of effectively determining for an arbitrarily given positive integer whether it is, or is not, in the set. While, in a certain sense, the theory of recursively enumerable sets of positive integers is potentially as wide as the theory of general recursive functions, the decision problems for such sets constitute a very special class of decision problems. Nevertheless they are important, as is shown by the following special and general examples.

One of the problems posed by Hilbert in his Paris address of 1900 [10, problem 10] is the problem of determining for an arbitrary diophantine equation with rational integral coefficients whether it has, or has not, a solution in rational integers. If the variables in a

⁵ We have thus restricted the normal operations and normal systems of [19] because of the following result. If in the initial string and in the normal operations of a normal system with primitive letters 1, $a'_1, \dots, a'_{\mu'}$, each a'_i , $i=1, \dots, \mu'$, is replaced by $b1 \dots 1b$ with i 1's, a normal system with primitive letters 1, b results, defining the same set of strings on 1 only as the original normal system.

diophantine equation be chosen from a given enumerably infinite set of variables, it is clear that the set of diophantine equations is enumerably infinite. Indeed they can be effectively put into one-one correspondence with the set of positive integers. Since for any one diophantine equation, and assignment of rational integral values to its variables, it can be effectively determined whether or no the equation is satisfied by those values, the set of diophantine equations having rational integral solutions can be generated. The corresponding integers under the above one-one correspondence can then also be generated, and, indeed, constitute a recursively enumerable set of positive integers.⁶ And under that correspondence, Hilbert's problem is transformed into the decision problem of that recursively enumerable set.

The assertions of an arbitrary symbolic logic⁷ constitute a generated set A of what may be called symbol-complexes or formulas. We assume that A is a subset of an infinite generated set E of symbol-complexes, which in one case may be the set of meaningful enunciations of the logic, in another the set of all symbol-complexes of a given mode of symbolization. The decision problem of the logic, more precisely its deducibility problem [3], is then the problem of determining of an arbitrary member of E whether it is, or is not, in A . Granting that every generated set is effectively enumerable, the members of E can be effectively set in one-one correspondence with the set of positive integers. The positive integers corresponding to the members of A then constitute a generated, and hence, under our generalization, a recursively enumerable set of positive integers. And under that correspondence the decision problem of the symbolic logic is transformed into the decision problem of this recursively enumerable set of positive integers.

Closely related to the technical concept recursively enumerable set of positive integers is that of a *recursive* set of positive integers. This is a set for which there is a recursive function $f(x)$ such that $f(x)$ is say 2 when x is a positive integer in the set, 1 when x is a positive integer not in the set. We may also make this the definition of the decision problem of the set being *recursively solvable*. For 2 and 1 may be regarded as the two possible truth-values, true, false, of the proposition "positive integer x is in the set," and the definition of recursive set is equivalent to this truth-value being recursively calculable for all positive integers x . If then recursive function is coextensive with

⁶ In view of [17] we inadvertently carried through our formal verification with "rational integral solution" replaced by "positive integral solution."

⁷ See Church [5, p. 225] for our omitting the qualifying "finitary."

effective calculability, recursive solvability is coextensive with solvability in the intuitive sense. In particular, the decision problem of a recursively enumerable set would be solvable or unsolvable according as the set is, or is not, recursive. More generally than in our two illustrations, through the more precise mechanism of Gödel representations [8], a wide variety of decision and other problems are transformed into problems about positive integers; and whether those problems are, or are not, solvable in the intuitive sense would be equivalent to their being, or not being, recursively solvable in the precise technical sense.

Gödel's classic theorem on the incompleteness and extendibility of symbolic logics [8] in all but wording led him to the recursive unsolvability of a generalization of the above problem of Hilbert [8, 9, 22]. Church explicitly formulated the concept of recursive unsolvability, and arrived at the unsolvability of a number of problems; certainly he proved them recursively unsolvable [1-4]. The above problem of Hilbert begs for an unsolvability proof (see [17]). Like the classic unsolvability proofs, these proofs are of unsolvability by means of given instruments. What is new is that in the present case these instruments, in effect, seem to be the only instruments at man's disposal.

Related to the question of solvability or unsolvability of problems is that of the reducibility or non-reducibility of one problem to another. Thus, if problem P_1 has been reduced to problem P_2 , a solution of P_2 immediately yields a solution of P_1 , while if P_1 is proved to be unsolvable, P_2 must also be unsolvable. For unsolvable problems the concept of reducibility leads to the concept of *degree of unsolvability*, two unsolvable problems being of the same degree of unsolvability if each is reducible to the other, one of lower degree of unsolvability than another if it is reducible to the other, but that other is not reducible to it, of incomparable degrees of unsolvability if neither is reducible to the other. A primary problem in the theory of recursively enumerable sets is the problem of determining the degrees of unsolvability of the unsolvable decision problems thereof. We shall early see that for such problems there is certainly a highest degree of unsolvability. Our whole development largely centers on the single question of whether there is, among these problems, a lower degree of unsolvability than that, or whether they are all of the same degree of unsolvability. Now in his paper on *ordinal logics* [26, section 4], Turing presents as a side issue a formulation which can immediately be restated as the general formulation of the "recursive reducibility" of one problem to another, and proves a result which immediately

generalizes to the result that for any "recursively given" unsolvable problem there is another of higher degree of unsolvability.⁸ While his theorem does not help us in our search for that lower degree of unsolvability, his formulation makes our problem precise. It remains a problem at the end of this paper. But on the way we do obtain a number of special results, and towards the end obtain some idea of the difficulties of the general problem.

1. Recursive versus recursively enumerable sets. The relationship between these two concepts is revealed by the following

THEOREM. *A set of positive integers is recursive when and only when both it and its complement with respect to the set of all positive integers are recursively enumerable.⁹*

For simplicity, we assume both the set S and its complement \bar{S} to be infinite. If, then, S is recursive, there is an effective method for telling of any positive integer n whether it is, or is not, in S . Generate the positive integers $1, 2, 3, \dots$ in their natural order, and, as a positive integer is generated, test its being or not being in S . Each time a positive integer is thus found to be in S , write it down as belonging to S . Thus, an effective process is set up for effectively enumerating the elements of S . Hence, S is recursively enumerable. Likewise \bar{S} can be shown to be recursively enumerable.

Conversely, let both S and \bar{S} be recursively enumerable, and let n_1, n_2, n_3, \dots be a recursive enumeration of S ; m_1, m_2, m_3, \dots , of \bar{S} . Given a positive integer n , generate in order $n_1, m_1, n_2, m_2, n_3, m_3$, and so on, comparing each with n . Since n must be either in S or in \bar{S} , in a finite number of steps we shall thus come across an n_i or m_j identical with n , and accordingly discover n to be in S , or \bar{S} . An effective method is thus set up for determining of any positive integer n whether it is, or is not, in S . Hence, S is recursive.

COROLLARY. *The decision problem of a recursively enumerable set is recursively solvable when and only when its complement is recursively enumerable.*

For then and only then is the recursively enumerable set recursive. It is readily proved that the logical sum and logical product of two

⁸ Both our generalization of his formulation and of his theorem have been carried through, rather hastily, by the formalism of [19], without, as yet, an actual equivalence proof. It may be that Tarski's Theorem 9.1 [23] can be transformed into a like absolute theorem.

⁹ The only portion of this theorem we can find in the literature is Rosser's Corollary II [20, p. 88].

recursively enumerable sets are recursively enumerable, the complement of a recursive set, and the logical sum, and hence logical product, of two recursive sets are recursive.

Clearly, any finite set of positive integers is recursive. For if n_1, n_2, \dots, n_r are the integers in question, we can test n being, or not being, in the set by directly comparing it with n_1, n_2, \dots, n_r .¹⁰ Likewise for a set whose complement is finite. For arbitrary infinite sets we have the following result of Kleene [12]. *An infinite set of positive integers is recursive when and only when it admits of a recursive enumeration without repetitions in order of magnitude.* Indeed, if n_1, n_2, n_3, \dots is a recursive enumeration of S without repetitions in order of magnitude, all n_i 's beyond the n th must exceed n . Hence we can test n being, or not being, in S by generating the first n members of the given recursive enumeration of S , and seeing whether n is, or is not, one of them. Conversely, if infinite S is recursive, the recursive enumeration thereof we set up in the proof of our first theorem is of the elements of S without repetition, and in order of magnitude.

A direct consequence of the first half of the last result is the following

THEOREM. *Every infinite recursively enumerable set contains an infinite recursive set.*

For, if n_1, n_2, n_3, \dots is a recursive enumeration of an infinite set S , for each n_i there must be, in this sequence, a later $n_j > n_i$. Hence, generate the elements n_1, n_2, n_3, \dots in order, and let $m_1 = n_1, m_2 = n_{i_2}$, the first n_i greater than $n_1, m_3 = n_{i_3}$, the first n_i beyond n_{i_2} greater than n_{i_2} , and so on. The sequence m_1, m_2, m_3, \dots is then a recursive enumeration of a subset of S without repetitions in order of magnitude. That subset is therefore infinite, and recursive.

Basic to the entire theory is the following result we must credit to Church, Rosser, Kleene, jointly [1, 20, 12].

THEOREM. *There exists a recursively enumerable set of positive integers which is not recursive.*¹¹

By our first theorem this is equivalent to the existence of a recursively enumerable set of positive integers whose complement is

¹⁰ The mere existence of a general recursive function defining the finite set is in question. Whether, given some definition of the set, we can actually discover what the members thereof are, is a question for a theory of proof rather than for the present theory of finite processes. For sets of finite sets the situation is otherwise, as seen in §11.

¹¹ In each of our existence theorems we show how to set up the basis of the set in question—at least, the corresponding formal proof does exactly that.

not recursively enumerable. Generate in order the distinct bases B_1, B_2, B_3, \dots of all recursively enumerable sets of positive integers as mentioned in the introduction, and keep track of these bases as the first, second, third, and so on, in this enumeration O . As the n th basis B_n is generated, with $n = 1, 2, 3, \dots$, set going the processes whereby the corresponding recursively enumerable set is generated, and whenever n is thus generated by B_n , place n in a set U . Being a generated set of positive integers, U is recursively enumerable. A positive integer n , then, is, or is not, in U according as it is, or is not, in the n th recursively enumerable set in O considered as an ordering of all recursively enumerable sets. Hence, n is, or is not, in \bar{U} , the complement of U , according as it is not, or is, in the n th set in O . We thus see that \bar{U} differs from each recursively enumerable set in the presence or absence of at least one positive integer. Hence \bar{U} is not recursively enumerable.

COROLLARY. *There exists a recursively enumerable set of positive integers whose decision problem is recursively unsolvable.*

Taken singly, finite sets, or sets whose complements are finite, are rather trivial examples of recursive sets. On the other hand, if we define two sets of positive integers to be *abstractly* the same if one can be transformed into the other by a recursive one-one transformation of the set of all positive integers into itself, then all infinite recursive sets with infinite complements are abstractly the same. Our theory being essentially an abstract theory of recursively enumerable sets, our interest therefore centers in recursively enumerable sets that are not recursive. Such sets, as well as their complements, are always infinite. We do not further pursue the question of two sets being abstractly the same, for that is but a special case of each set being one-one reducible to the other (§4).

2. A form of Gödel's theorem. Given any basis B , and positive integer n , the couple (B, n) may be used to represent the proposition, true or false, " n is in the set generated by B ." By interlacing the process for generating the distinct bases in the sequence B_1, B_2, B_3, \dots and the process for generating the positive integers in the sequence $1, 2, 3, \dots$ by the addition of 1's, we can effectively generate the distinct couples (B, n) in the single infinite sequence

$$O': (B_1, 1), (B_2, 1), (B_1, 2), (B_3, 1), (B_2, 2), (B_1, 3), \dots$$

On the one hand, the set of all couples (B, n) is thus a generated set of expressions which we shall call E . On the other hand, O' leads to an effective 1-1 correspondence between the members of E and the

set of positive integers, (B, n) corresponding to m if (B, n) is the m th member of O' . We may call m the Gödel representation¹² of (B, n) . Given a generated subset of E , the Gödel representations of its members will constitute a generated set of positive integers, and conversely. Thus, in the former case we can generate the members of the subset of E , and, as a couple (B, n) is generated, find its Gödel representation m by regenerating O' . The set of these m 's is thus a generated set. Likewise for the converse. If, therefore, we formally define a subset of E to be recursively enumerable if the set of Gödel representations of its members is recursively enumerable,¹³ we can conclude that every generated subset of E is recursively enumerable, and, of course, conversely. Similarly for a like formal definition of a recursive subset of E .

While E is just the set of couples (B, n) , it may be interpreted as the set of enunciations " n is in the set generated by B ." The subset T of E consisting of those couples (B, n) for which n is in the set generated by B may then be interpreted as the set of true propositions in E , while \bar{T} , the complement of T with respect to E , consists of the false propositions in E .

Actually, T itself can be generated as follows. Generate B_1, B_2, B_3, \dots in order. As a B is generated, set up the process for generating the set of positive integers determined by B , and, whenever a positive integer n is thus generated, write down the couple (B, n) . Each (B, n) for which n is in the set generated by B will thus be written down, and conversely. This generated set of (B, n) 's is then T . We therefore conclude that T is recursively enumerable.

Now let F be any recursively enumerable subset of \bar{T} . If (B, n) is in F , it is in \bar{T} , and hence n is certainly not in the set generated by B . Now generate the members of F , and if (B, n) is thus generated, find the n th member B_n of $O: B_1, B_2, B_3, \dots$, and if B_n is B , place n in a set of positive integers S_0 . Since S_0 is thus a generated set of positive integers, it is recursively enumerable. It will therefore be determined by some basis B . Let this basis be in the v th in O , that is, let the basis be B_v , and form the couple (B_v, v) . Now by construction, S_0 consists of those members of F of the form (B_n, n) . Suppose that (B_v, v) is in F . Then, on the one hand, proposition (B_v, v) being false,

¹² Rather is the Gödel representation in [8] not just an effectively corresponding positive integer, but one which, when expressed according to a specific algorithm, is "formally similar," in the sense of Ducasse [7, p. 51], to the symbolic expression represented.

¹³ In our own development [19], "recursively enumerable subset of E " is defined directly as a normal subset of E , or rather of the set of symbolic representations of the members of E .

ν would not be in the set generated by B_ν , that is (1): ν would not be in S_0 . But (B_ν, ν) being of the form (B_n, n) , (2): ν would be in S_0 . Our assumption thus leading to a contradiction, it follows that (B_ν, ν) is not in F . But ν can only be in S_0 by (B_ν, ν) being in F . Hence, ν is not in S_0 . Finally, (B_ν, ν) as proposition says that ν is in S_0 . The proposition (B_ν, ν) is therefore false, that is (B_ν, ν) is in \bar{T} .

For any recursively enumerable subset F of \bar{T} there is then always this couple (B_ν, ν) in \bar{T} , but not in F . On the one hand, then, \bar{T} can never be F . Hence, \bar{T} is not recursively enumerable. By the definitions of this section, and the first theorem of the last, it follows that T , while recursively enumerable, is not recursive. By the decision problem of T we mean the problem of determining for an arbitrarily given member of E whether it is, or is not, in T . But that can be interpreted as the decision problem for the class of recursively enumerable sets of positive integers, that is, the problem of determining for any arbitrarily given recursively enumerable set, that is, arbitrarily given basis B of such a set, and arbitrary positive integer n whether n is, or is not, in the set generated by B . We may therefore say that the decision problem for the class of all recursively enumerable sets of positive integers is recursively unsolvable, and hence, in all probability, unsolvable in the intuitive sense.

On the other hand, since (B_ν, ν) of \bar{T} is not in F , T and F together can never exhaust E . Now T , or any recursively enumerable subset T' of T , in conjunction with F may be called a recursively generated logic relative to the class of enunciations E . For the appearance of (B_ν, ν) in T' assures us of the truth of the proposition " n is in the set generated by B ," while its presence in F would guarantee its falseness. We can then say that no recursively generated logic relative to E is complete, since F alone will lead to the (B_ν, ν) which is neither in T' nor in F . That is, (B_ν, ν) is undecidable in this logic. Moreover, if, with a given "basis" for F , the above argument is carried through formally,¹⁴ the recursively enumerable S_0 obtained above will actually be given by a specific basis B which can be constructed by that formal argument. Having found this B , we can then regenerate $O:B_1, B_2, B_3, \dots$, until B is reached, and thus determine the ν such that $B=B_\nu$. That is, given the basis of F , the (B_ν, ν) in \bar{T} and not in F can actually be found. If then we add this (B_ν, ν) to F , a wider recursively enumerable subset F' of \bar{T} results. We may then say that every recursively generated logic relative to E can be extended. Outwardly, these two results, when formally developed, seem to be

¹⁴ Here, the basis of F may be taken to be the basis of the recursively enumerable set of Gödel representations of the members of F . But see the preceding footnote.

Gödel's theorem in miniature. But in view of the generality of the technical concept general recursive function, they implicitly, in all probability, justify the generalization that every symbolic logic is incomplete and extendible relative to the class of propositions constituting E .¹⁵ The conclusion is unescapable that even for such a fixed, well defined body of mathematical propositions, *mathematical thinking is, and must remain, essentially creative*. To the writer's mind, this conclusion must inevitably result in at least a partial reversal of the entire axiomatic trend of the late nineteenth and early twentieth centuries, with a return to meaning and truth as being of the essence of mathematics.

3. The complete set K ; creative sets. Return now to the effective 1-1 correspondence between the set E of distinct (B, n) 's and the set of positive integers obtained via the effective enumeration O' of E . Since T is a recursively enumerable subset of E , the positive integers corresponding to the elements of T constitute a recursively enumerable set of positive integers, K . We shall call K the *complete set*.¹⁶ Since \bar{T} is not recursively enumerable, \bar{K} , which consists of the positive integers corresponding to the elements of \bar{T} , is not recursively enumerable. Now let B be the basis of a recursively enumerable subset α of \bar{K} . The elements of E corresponding to the members of α constitute, then, a recursively enumerable subset F of \bar{T} . Find then the (B_v, v) of \bar{T} not in F , and, via O' , the positive integer n corresponding to (B_v, v) . This n will then be an element of \bar{K} not in α .

Actually, we have no general method of telling when a basis B defines a recursively enumerable subset of \bar{K} . Indeed, the above method will yield a unique positive integer n for any basis B of a recursively enumerable set α of positive integers. However, when α is a subset of \bar{K} , n will also be in \bar{K} , but not in α .

Furthermore, even the formal proof of this result merely gives an effective method for finding n , given B . But this method itself can be formalized, so that, as a result, n is given as a "recursive function of B ." This can mean that a recursive function $f(m)$ can be set up such that $n = f(m)$ where $B = B_m$. We now isolate this property of K by setting up the

DEFINITION. *A creative set C is a recursively enumerable set of positive integers for which there exists a recursive function giving a unique*

¹⁵ See Kleene's Theorem XIII in [12] for a mathematically stateable theorem approximating the generality of our informal generalization.

¹⁶ "A complete set" might be better. Just how to abstract from K the property of completeness is not, at the moment, clear. By contrast, see "creative set" below.

positive integer n for each basis B of a recursively enumerable set of positive integers α such that whenever α is a subset of \bar{C} , n is also in \bar{C} , but not in α .

THEOREM. *There exists a creative set; to wit, the complete set K .*

Actually, the class of creative sets is infinite, and very rich indeed as shown by the following easily proved results.¹⁷ If C is a creative set, and E a recursively enumerable set of positive integers, then if E contains \bar{C} , CE is creative, if \bar{C} contains E , $C+E$ is creative. Results of §1 enable us actually to construct creative sets according to the first method by using E 's which are the complements of recursive subsets of C . Results of the rest of this section lead to constructions using the second method.

It is convenient to talk as if the n in the definition of a creative set were determined by the α thereof instead of by the basis B of α . Clearly every creative set C is a recursively enumerable set which is not recursive. For were \bar{C} recursively enumerable, there could be no n in \bar{C} not in the recursively enumerable subset \bar{C} of \bar{C} . The decision problem of each creative set is therefore recursively unsolvable. On the other hand, the complement \bar{C} of any creative set C contains an infinite recursively enumerable set. Recall that every finite set is recursive, and hence recursively enumerable. With, then, α of the definition of creative set as the null set, find the $n = n_1$ of \bar{C} "not in α ." With α the unit set having n_1 as sole member, $n = n_2$ will be in \bar{C} , and distinct from n_1 . With α consisting of n_1 and n_2 , $n = n_3$ will be in \bar{C} , and distinct from n_1 and n_2 , and so on. The set of positive integers n_1, n_2, n_3, \dots is then an infinite generated, and hence recursively enumerable, subset of \bar{C} .

Actually, with this subset of \bar{C} as α , a new element n_ω of \bar{C} is obtained, and so on into the constructive transfinite. But this process is essentially creative. For any mechanical process could only yield n 's forming a generated, and hence recursively enumerable, subset α of \bar{C} , and hence could be transcended by finding that n of \bar{C} not in α .

4. One-one reducibility, to K ; many-one reducibility. Let S_1 and S_2 be any two sets of positive integers. One of the simplest ways in which the decision problem of S_1 would be reduced to the decision problem of S_2 would arise if we had an effective method which would determine for each positive integer n a positive integer m such that n is, or is not, in S_1 according as m is, or is not, in S_2 . For if we could

¹⁷ Of course, all sets abstractly the same as a given creative set, in the sense of §1, are creative. Likewise for our later simple and hyper-simple sets.

somehow determine whether m is, or is not, in S_2 , we would determine n to be, or not be, in S_1 correspondingly. If "effective method" be replaced by "recursive method," we shall say, briefly, that S_1 is then *many-one reducible* to S_2 . If, furthermore, different n 's always lead to different m 's, we shall say that S_1 is *one-one reducible* to S_2 .¹⁸ "Recursive method" here can mean that $m = f(n)$, where $f(n)$ is a recursive function.

THEOREM. *The decision problem of every recursively enumerable set of positive integers is one-one reducible to the decision problem of the complete set K.*

For let B' be a basis of any one recursively enumerable set S' . The effective one-one correspondence between all (B, n) 's and all positive integers yielded by the effective enumeration O' of E , the set of all (B, n) 's, then yields a unique positive integer m for each (B', n) , B' fixed, and thus a unique m for each n , different n 's yielding different m 's. Now n is, or is not, in S' according as (B', n) is in T , or \bar{T} , and hence according as m is in K , or \bar{K} , whence our result.

Since K itself is recursively enumerable, we may say that for recursively enumerable sets of positive integers with recursively unsolvable decision problems there is a *highest degree of unsolvability relative to one-one reducibility*, namely, that of K . Actually, one-one reducibility is a special case of all the more general types of reducibility later introduced, and, though the proof of this is still in the informal stage, these latter are special cases of general recursive, that is, Turing reducibility. The same result then obtains relative to these special types of reducibility and, more significantly, for reducibility in the general sense.¹⁹

We have thus far explicitly obtained two recursively enumerable sets with recursively unsolvable decision problems, the U of our first section, and K . We may note that a certain necessary and sufficient condition for the many-one reducibility of K to a recursively enumerable set, the proof of which is still in the informal stage, has as an immediate consequence that K is many-one reducible to U . It would then follow that K and U are of the same degree of unsolvability relative to many-one reducibility.

¹⁸ The resulting one-to-one correspondence is then between $S_1 + \bar{S}_1$ and a subset, recursively enumerable indeed, of $S_2 + \bar{S}_2$. Of course, both $S_1 + \bar{S}_1$ and $S_2 + \bar{S}_2$ constitute the set of all positive integers.

¹⁹ It seems rather obvious that K and the problem of Church [1] are each at least many-one reducible to the other; likewise for the problem of [1] and of [2, 3]. Had we verified this in detail, we would have called this highest degree of unsolvability of decisions problems of recursively enumerable sets the *Church degree of unsolvability*.

5. Simple sets. It is readily proved that the necessary and sufficient condition that every recursive set be one-one reducible to a given recursively enumerable set of positive integers S is that S is infinite, and \bar{S} contains an infinite recursively enumerable set. We are thus led to ask if there exist sets satisfying the following

DEFINITION. *A simple set is a recursively enumerable set of positive integers whose complement, though infinite, contains no infinite recursively enumerable set.*

We now prove the

THEOREM. *There exists a simple set.*

Recall the set T of all couples (B, n) such that positive integer n is in the recursively enumerable set of positive integers determined by basis B . Since T is recursively enumerable, we can set up an effective enumeration

$$O'': (B_{i_1}, n_1), (B_{i_2}, n_2), (B_{i_3}, n_3), \dots$$

of its members. The subscript of each B is its subscript in the effective enumeration $O: B_1, B_2, B_3, \dots$ of all distinct B 's. Now the complement of a set containing no infinite recursively enumerable set is equivalent to the set itself having an element in common with each infinite recursively enumerable set. Generate then the distinct bases B_1, B_2, B_3, \dots , and as a B_i is generated, regenerate the sequence O'' of (B, n) 's in T , and the first time, if ever, B is B_i , and n is greater than $2i$, place n in a set S . The resulting set S is then a generated, and hence recursively enumerable, set of positive integers. We proceed to prove it simple.

If S' is an infinite recursively enumerable set of positive integers, it will be determined by some basis B_i , and will have some element m greater than $2i$. Since (B_i, m) , being then in T , will appear in O'' , our construction will place m in S , if some earlier (B_i, n) of O'' has not already contributed an element of S' to S . That is, S has an element in common with each infinite recursively enumerable S' . As for \bar{S} being infinite, note that each B_i contributes at most one element to S . The first n B 's in O therefore contribute at most n elements to S . Each B_i with $i \geq n+1$ can only contribute to S an element greater than $2n+2$. Of the first $2n+2$ positive integers, at most n are therefore in S , and hence at least $n+2$ are in the consequently infinite \bar{S} .²⁰

²⁰ $n > i$ can replace $n > 2i$ in the above construction, but the proof will then depend on there being an infinite number of bases defining the null set.

Having one simple set, the method of our succeeding §8 can be modified to yield a rich infinite class of simple sets. Clearly, *every simple set S is a recursively enumerable set that is not recursive*. For were S recursive, \bar{S} would be an infinite recursively enumerable subset of \bar{S} . The decision problem of each simple set is therefore recursively unsolvable. We thus have obtained two infinite mutually exclusive classes of recursively enumerable sets with recursively unsolvable decision problems, the class of creative sets, and the class of simple sets. They are poles apart in that the complements of creative sets have a creative infinity of infinite recursively enumerable subsets, those of simple sets, not one.

In passing, we may note that every recursively enumerable set of positive integers S with recursively unsolvable decision problem leads to an incompleteness theorem for symbolic logics relative to the class of propositions $n \in S$, n an arbitrary positive integer. Creative sets S are then exactly those recursively enumerable sets of this type each of which admits a universal extendibility theorem as well, simple sets S those for which, given S , each logic can prove the falsity of but a finite number of the infinite set of false propositions $n \in S$.

It is readily seen that no creative set C can be one-one reducible to a simple set S . For under such a reduction, each infinite recursively enumerable subset of C , proved above to exist, would be transformed into an infinite recursively enumerable subset of \bar{S} , contradicting the simplicity of S . Simple sets thus offer themselves as *candidates* for recursively enumerable sets with decision problems of lower degree of unsolvability than that of the complete set K . Even for many-one reducibility the situation is no longer immediately obvious; for an infinite recursively enumerable subset of C could thus be transformed into a finite subset of \bar{S} , the complement of simple S , without contradiction. However we can actually go much further than that.

6. Reducibility by truth-tables. If S_1 is many-one reducible to S_2 , positive integer n being, or not being, in S_1 may be said to be determined by its correspondent m being, or not being, in S_2 in accordance with the truth-table

(S_2)	m	n	(S_1)
	+	+	
	-	-	

Here, the two signs +, - under m represent the two possibilities m is in S_2 , m is not in S_2 , respectively. And by the sign under n in the

same horizontal row as the corresponding sign under m the table in the same language tells whether n correspondingly is (+), or is not (-), in S_1 . The table then says that when m is in S_2 , n is in S_1 , when m is not in S_2 , n is not in S_1 , as required by many-one reducibility. Now there are altogether four ways in which n being, or not being, in S_1 can be made to depend solely on m being, or not being, in S_2 , the signs under n being +, - as above; or +, +; -, -; -, +. If then we have an effective method which for each positive integer n will not only determine a unique corresponding positive integer m , but also one of these four "first order" truth-tables, and if in each case the table is such that for the correct statement of membership or non-membership of m in S_2 , it gives the correct statement of membership or non-membership of n in S_1 , then the decision problem of S_1 will thus be reduced to the decision problem of S_2 . For here also, given n , if we could somehow determine whether m is, or is not, in S_2 , we could thereby determine which row of the corresponding table correctly describes the membership or non-membership of m in S_2 , and from that row correctly determine whether n is, or is not, in S_1 .

More generally, let there be an effective method which for each positive integer n determines a finite sequence of positive integers m_1, m_2, \dots, m_v as well as the m 's depending on n . Let that method correspondingly determine for each n a " v th order" truth-table of the form

(S_2)	m_1	m_2	\dots	m_v	$ $	n	(S_1)
+	+	...	+			-	
+	+	...	-			+	
.	
+	+	...	+			-	

Each horizontal row, to the left of the vertical bar, specifies one of the 2^v possible ways in which the v m_i 's may, or may not, be in S_2 , to the right of the bar correspondingly commits itself to one of the statements n is in S_1 , n is not in S_1 . If then for each n that row of the corresponding table which gives the correct statements for the m 's being or not being in S_2 also gives the correct statement regarding the membership or non-membership of n in S_1 , the decision problem of S_1 is again thereby reduced to the decision problem of S_2 .

If such a situation obtains with "effective method" replaced by "recursive method," we shall say that S_1 is *reducible to S_2 by truth-tables*. "Recursive method" here can mean that a suitable Gödel representation of the couple consisting of the sequence $m_1, m_2, \dots,$

m , and the truth-table of order ν is a recursive function of n . If the orders of the truth-tables arising in such a reduction are bounded, we shall say that S_1 is *reducible to S_2 by bounded truth-tables*. Since there are 2^{ω} distinct truth-tables of order ν , reducibility by bounded truth-tables is equivalent to reducibility by truth-tables in which but a finite number of distinct tables arise.

7. Non-reducibility of creative sets to simple sets by bounded truth-tables. Let us suppose that creative set C is reducible to simple set S by bounded truth-tables. Let $T_1, T_2, \dots, T_\kappa$ be the finite set of distinct truth-tables entering into such a reduction. That reduction then effectively determines for each positive integer n a finite sequence of positive integers m_1, m_2, \dots, m_ν , and a unique T_i , $1 \leq i \leq \kappa$.

The gist of our reductio-ad-absurdum proof consists in showing that under the assumed reduction we can obtain for each natural number p a sequence of m 's at least p of which are in S . We then immediately have our desired contradiction. For in each case $p \leq \nu$. The finite set of ν 's, the orders of the T_i 's, being bounded, p cannot then be arbitrarily large as stated.

More precisely we prove by mathematical induction that under the assumed reduction the following would be true. *For each natural number p an effective process Π_p can be set up which will determine for each recursively enumerable subset α of \bar{C} an element n of \bar{C} not in α , and which for the corresponding m_1, m_2, \dots, m_ν and T_i yielded by the assumed reduction will correctly designate p of these m 's as belonging to S .* The mode of designation may be assumed to be by specifying the sequence of subscripts, i_1, i_2, \dots, i_p , of the m 's to be designated, with say $i_1 < i_2 < \dots < i_p$. With the assumed reduction adjoined to this process, Π_p then determines for each α in question the quadruplet (n, M, T_i, I) , M being the sequence of m 's, I the sequence of subscripts of the p designated m 's.

For $p=0$, Π_p is immediately given by the creative character of C . For that immediately gives us for each recursively enumerable subset α of \bar{C} a definite element n of \bar{C} not in α . The assumed reduction yields the corresponding M and T_i ; and with no members of M designated as being in S , I is the null sequence.

Inductively, assume that we have the process Π_p for $p=k$. Let α be any given recursively enumerable subset of \bar{C} , and let $(n', M', T_{i'}, I')$ be the corresponding quadruplet yielded by Π_k . Now suppose n is a positive integer for which the assumed reduction yields the same table $T_{i'}$ as it did for n' , and a sequence of m 's, M , consequently of the same length as M' , having the following property. For each un-

designated element of M' , the correspondingly placed element of M is identical with that of M' ; for each element of M' designated as being in S , the corresponding element of M is also in S . Such an n must then be in \bar{C} along with n' . For that row of T_v which correctly tells of the m 's of M' whether they are, or are not, in S will also be the correct row for M . And since in the former case that row must say that n' is in \bar{C} , in the latter case it will say that n is in \bar{C} , and correctly so. We proceed to show how all such n 's may be generated.

We first show how to generate all M 's obtainable from M' by replacing the designated elements of M' by arbitrary elements of S . For any one such M , the replacing elements, being finite in number, will be among the first N elements, for some positive integer N , of a given recursive enumeration of S . Generate then the positive integers $1, 2, 3, \dots$, and as a positive integer N is generated, generate the first N elements of the given recursive enumeration of S . For each N place in a set β the at most N^k sequences M that can be obtained from M' by replacing the designated elements of M' by elements chosen from the first N elements of S . The generated set of sequences β then consists of all M 's obtainable from M' by replacing the designated elements of M' by arbitrary elements of S .

The n 's we wish to generate are then those positive integers for which the assumed reduction yields the table T_v and a sequence of m 's, M , such that M is a member of β . Generate then the elements of β . As an element M of β is generated, generate the positive integers $1, 2, 3, \dots$, and as a positive integer n is generated, find the corresponding sequence of m 's and table yielded by the reduction of C to S . If then that sequence of m 's is M , and the table is T_v , add n to the given set α . As seen above, each such n will be in \bar{C} . Hence the resulting generated, and hence recursively enumerable, set α' is a subset of \bar{C} containing α . Our reason for thus adding the desired n 's to α instead of just forming the class thereof is that the iterative process we are about to set up requires a cumulative effect.

As a result of our hypothesis and construction we thus have a derived process Π'_k which for every recursively enumerable subset α of \bar{C} yields a definite recursively enumerable subset α' of \bar{C} containing α . Starting with α , we may then iterate the process Π'_k to obtain the infinite sequence $A : \alpha_1, \alpha_2, \alpha_3, \dots$, where $\alpha_1 = \alpha$, $\alpha_{n+1} = (\alpha_n)'$. Each member of A is thus a recursively enumerable subset of \bar{C} , and contained in the next member of A . By applying the original process Π_k to the members of A we correspondingly obtain the infinite sequence $\Sigma : \sigma_1, \sigma_2, \sigma_3, \dots$, where σ_i is the quadruplet $(n^{(i)}, M^{(i)}, T_i^{(i)}, I^{(i)})$ yielded by Π_k for α_i . We then observe the following. If for $j_1 \neq j_2$

the T 's of σ_{i_1} and σ_{i_2} are the same, and the I 's are the same, then the sequences obtained from the M 's by deleting the designated m 's cannot be identical. For if they also were identical, then, with say $j_1 < j_2$, $n^{(j_2)}$ would have been assigned to $\alpha^{(i_1+1)}$, whereas it actually is outside of $\alpha^{(i_2)}$ which contains $\alpha^{(i_1+1)}$. Hence, the infinite sequence Σ' , obtained from Σ by deleting from each σ , the integer $n^{(i)}$ and the designated m 's of $M^{(i)}$, itself consists of distinct elements.

It follows that *there are an infinite number of distinct undesignated m's appearing in Σ* . Indeed, the distinct $T_i^{(j)}$'s of Σ are at most κ in number. With $T_i^{(j)}$ fixed, the order $\nu^{(i)}$ of $T_i^{(j)}$ is fixed; and since $1 \leq i_1^{(j)} < i_2^{(j)} < \dots < i_k^{(j)} \leq \nu^{(i)}$, the number of distinct I 's is finite. Finally, with T and I fixed, were the total number of distinct undesignated m 's finite, the number of distinct ways in which those $\nu^{(i)} - k$ undesignated m 's could assume values would be finite. Hence Σ' would be finite, not infinite.

Now were each of this infinite set of undesignated m 's in \bar{S} , we could regenerate the elements of Σ , and as an element σ , thereof is generated, place all of its undesignated m 's in a set γ , and thus obtain an infinite generated, and hence recursively enumerable, subset of \bar{S} . As this contradicts the simplicity of S , it follows that *at least one undesignated m arising in Σ is in S* .

We can then find a unique such m , as well as a σ in which it occurs, as follows. With $N = 1, 2, 3, \dots$, generate the first N elements of the given recursive enumeration of S , and the first N elements of Σ , and test the latter in order to see if any undesignated m is among those first N elements of S . If a particular undesignated m of Σ in S , proved above to exist, is the L th member of S , and in the K th member σ_K of Σ , then an affirmative answer to the above test will certainly be obtained for $N = \max(L, K)$. Find then the first N for which an affirmative answer is obtained, and let (m, M, T_i, I) be the first σ to yield the affirmative answer for this N , m ; the first undesignated m of M thus found to be in S . We can then add m to the designated m 's of M , thus obtaining a quadruplet (n, M, T_i, I_1) , where I_1 designates $(k+1)$ of the m 's of M as being in S , and where n is certainly a member of \bar{C} not in the originally given α . But the whole process leading up to (n, M, T_i, I_1) is determined by that α . It is therefore the desired process Π_p for $p = k+1$.

Under the assumed reduction of C to S , Π_p would therefore exist for every natural number p . With α say the null set, we would thus obtain for every natural number p a quadruplet (n_p, M_p, T_{i_p}, I_p) such that p of the members of the sequence M_p are in S . Yet the total length of M_p is the order of T_{i_p} , and hence bounded. Hence the

THEOREM. *No creative set is reducible to a simple set by bounded truth-tables.*

We recall that every recursively enumerable set of positive integers is one-one reducible to the creative set K , the complete set. Hence the

COROLLARY. *Every simple set is of lower degree of unsolvability than the complete set K relative to reducibility by bounded truth-tables.*

8. Counter-example for unbounded truth-tables. We recall that for the particular simple set S constructed in §5, of the first $2m+2$ positive integers at most m were in S , m being any positive integer. Hence, of the $m+1$ integers $m+2, m+3, \dots, 2m+2$, at least one is in \bar{S} . By setting $m = 2^n - 1$, with $n = 1, 2, 3, \dots$, we can effectively generate the infinite sequence of mutually exclusive finite sequences

$$\sigma: (3, 4), (5, 6, 7, 8), \dots, (2^n + 1, 2^n + 2, \dots, 2^{n+1}), \dots$$

such that each sequence in σ has at least one member thereof in \bar{S} . An effective one-one correspondence between the positive integers $1, 2, 3, \dots$ and the elements of σ is then obtained by making the positive integer n correspond to the sequence $(2^n + 1, 2^n + 2, \dots, 2^{n+1})$ constituting the n th element of σ .

Given a creative set C , regenerate the elements of S , placing each in a set S_1 . Furthermore, regenerate the elements of C , and as an element n thereof is generated, place all of the positive integers in the n th sequence of σ in S_1 . The resulting set S_1 is a generated, and hence recursively enumerable, set of positive integers. Since S_1 contains S , \bar{S} contains \bar{S}_1 . As S is simple, \bar{S} , and hence \bar{S}_1 , does not have an infinite recursively enumerable subset. Moreover, \bar{S}_1 is also infinite. For \bar{C} is infinite. And, for each element of \bar{C} , the corresponding sequence in σ has only those of its members that are already in S also in S_1 , and hence at least one element in \bar{S}_1 . Hence, S_1 is simple.

Likewise we see that a positive integer n is in C , or \bar{C} , according as all of the integers in the n th sequence of σ are in S_1 , or at least one is in \bar{S}_1 . If then we make correspond to each positive integer n the sequence of 2^n positive integers $(2^n + 1, 2^n + 2, \dots, 2^{n+1})$, and the truth-table of order 2^n in which the sign under n is $+$ in that row in which the signs under the 2^n "m's" are all $+$, and in every other row the sign under n is $-$, we have a reduction of C to S_1 by truth-tables. Hence the

THEOREM. *For each creative set C a simple set S can be constructed such that C is reducible to S by unbounded truth-tables.*

COROLLARY. *A simple set S can be constructed which is of the same degree of unsolvability as the complete set K relative to reducibility by truth-tables unrestricted.*

Simple sets as such do not therefore give us the absolutely lower degree of unsolvability than that of K we are seeking.

9. Hyper-simple sets. The counter-example of the last section suggests that we seek a set satisfying the following

DEFINITION. *A hyper-simple set H is a recursively enumerable set of positive integers whose complement \bar{H} is infinite, while there is no infinite recursively enumerable set of mutually exclusive finite sequences of positive integers such that each sequence has at least one member thereof in \bar{H} .*²¹

In this definition we may use the original Gödel method for representing a finite sequence of positive integers m_1, m_2, \dots, m_r by the single positive integer $2^{m_1} 3^{m_2} \cdots p^{m_r}$, where $2, 3, \dots, p_r$ are the first r primes in order of magnitude. A set of finite sequences of positive integers is then recursively enumerable if the set of Gödel representations of those sequences is recursively enumerable.

THEOREM. *A hyper-simple set exists.*

Our intuitive argument must again draw upon the formal development to the effect that each recursively enumerable set of finite sequences of positive integers will be determined by a "basis" B^* , and that all such bases can be generated in a single infinite sequence of distinct bases

$$O^*: B_1^*, B_2^*, B_3^*, \dots$$

As in §2, generate the elements of O^* , and as an element B^* is generated, set up the process for generating the set of sequences determined by B^* , and as a sequence s is thus generated, write down the couple (B^*, s) . The resulting set of couples is then a generated set, and can indeed be effectively ordered in a sequence of distinct couples

$$O_1^*: (B_{i_1}^*, s_1), (B_{i_2}^*, s_2), (B_{i_3}^*, s_3), \dots$$

²¹ Mutually exclusive sequences here mean no element of one sequence is an element of another. Curry suggests that "hyper-simple" is linguistically objectionable, and should be replaced by "super-simple." But we would not then know what to use in place of the letter H .

O_1^* then consists of all distinct couples (B^*, s) such that finite sequence s is a member of the recursively enumerable set of finite sequences of positive integers determined by basis B^* .

Now the condition that no infinite recursively enumerable set of mutually exclusive finite sequences of positive integers has the property that each sequence has at least one positive integer thereof in \bar{H} is equivalent to each such set of sequences having at least one sequence all of whose members are in H . Our method of constructing the desired hyper-simple set H will then consist in placing in H for certain B^* 's in O^* all of the positive integers in a sequence in the set of sequences determined by B^* . For purposes of presentation we shall call each such basis B^* a *contributing basis*, while every B^* determining an infinite recursively enumerable set of mutually exclusive sequences will be called a *relevant basis*. Set H , if recursively enumerable, will then be hyper-simple if each relevant basis is a contributing basis, and \bar{H} is infinite.

If B^* is a relevant basis, then among the infinite number of mutually exclusive sequences generated by B^* there must be a sequence each of whose elements exceeds an arbitrarily given positive integer N . For did every sequence generated by B^* have as element one of the integers $1, 2, \dots, N$, for any $N+1$ of these sequences at least two would have one of these integers in common. We shall then generate H by regenerating sequence O_1^* , and, as an element $(B_{t_n}^*, s_n)$ thereof is generated, we shall place all the elements of s_n in H if $B_{t_n}^*$ has not thus been made a contributing basis earlier in the process, while the elements of s_n are all greater than a certain positive integer N_n , about to be determined; otherwise none. Inductively, assume N_m to have been determined for $1 \leq m < n$, and thus the entire process up to the time $(B_{t_n}^*, s_n)$ was brought up for consideration. Let $B_{j_1}^*, B_{j_2}^*, \dots, B_{j_\nu}^*$ be the bases that have thus far contributed to H , and in the order in which they became contributing bases. These bases are then distinct, and hence their subscripts, which give their position in the sequence O^* of all distinct bases, are distinct. Let k_1, k_2, \dots, k_ν be the largest integer placed in H by the first contributing basis, by the first two, \dots , by the first ν . The result being cumulative, $k_1 \leq k_2 \leq \dots \leq k_\nu$. The crux of our construction is to make N_n depend not on the history of all these ν contributions to H , but only on that part of that history up to and including the last contribution, if any, made by a B^* preceding $B_{t_n}^*$ in O^* . Specifically, if $B_{j_\mu}^*$ is the last of the above ν contributing bases preceding $B_{t_n}^*$ in O^* , that is, with $j_\mu < i_n$, N_n is to be one more than the largest integer present in H as a result of all the contributions made up to and in-

cluding the contribution made by $B_{j_\mu}^*$. That is, $N_n = k_\mu + 1$. Actually, if none of the ν contributing bases precede $B_{i_n}^*$ in O^* , no condition is to be placed on s_n , and all of its elements are placed in H so long as $B_{i_n}^*$ is distinct from the ν contributing bases obtained thus far.

Furthermore, in our induction assume that we have been able to keep a record of the sequence $B_{j_1}^*, B_{j_2}^*, \dots, B_{j_\nu}^*$, of k_1, k_2, \dots, k_ν , and also of j_1, j_2, \dots, j_ν , up to the time $(B_{i_n}^*, s_n)$ was about to be generated. We then generate $(B_{i_n}^*, s_n)$, and by regenerating O^* find the place of $B_{i_n}^*$ in O^* thus determining the subscript i_n . Our criterion for determining whether, or no, the elements of s_n are to be placed in H then becomes effective. In the latter case, the record is unchanged as we generate $(B_{i_{n+1}}^*, s_{n+1})$. In the former, $B_{i_n}^*$ is written into the record as $B_{j_{\nu+1}}^*$, i_n as $j_{\nu+1}$ while we can write in for $k_{\nu+1}$ the maximum of k , and the largest integer in s_n . The entire process is thus effective at each stage, and H is thus a generated, and hence recursively enumerable, set of positive integers. We proceed to prove it hyper-simple.

Let B^* be any relevant basis. Of the finite number of bases preceding B^* in O^* , but a finite number can be contributing bases. Let $B_{j_\mu}^*$ be the last of these contributing bases, if any, appearing in the sequence $B_{j_1}^*, B_{j_2}^*, B_{j_3}^*, \dots$ of distinct contributing bases determined by the above generation of H . There will then be a sequence s generated by B^* each of whose elements is greater than $k_\mu + 1$. When then (B^*, s) , a definite element of O_1^* , is generated in the course of generating H , B^* will contribute each element of s to H unless it became a contributing basis earlier in the process. Hence, every relevant basis is a contributing basis.

It also follows, or is easily seen directly, that the number of contributing bases is infinite. Consider then the infinite sequence of contributing bases $B_{j_1}^*, B_{j_2}^*, B_{j_3}^*, \dots$, the corresponding infinite sequence of subscripts j_1, j_2, j_3, \dots , and the associated infinite sequence k_1, k_2, k_3, \dots . Since the contributing bases are distinct, so are their subscripts. Hence, for each j_m , among the infinite set of j 's following j_m there is a unique least $j, j_{m'}$. Consider then the resulting infinite sequence $j_{\lambda_1}, j_{\lambda_2}, j_{\lambda_3}, \dots$, where j_{λ_1} is the least j in the whole infinite sequence of j 's, while $\lambda_2 = (\lambda_1)', \lambda_3 = (\lambda_2)', \dots$. Now k_{λ_n} is the largest integer contributed to H through the contributing basis with subscript j_{λ_n} . Since j_{λ_n} is the smallest j following $j_{\lambda_{n-1}}$ it is less than all succeeding j 's. Hence B^* with subscript j_{λ_n} precedes in O^* all bases following that B^* in the above infinite sequence of contributing bases. Hence, each element added to H by contributing bases thus following B^* with subscript j_{λ_n} must exceed $k_{\lambda_n} + 1$. It fol-

lows, on the one hand, that for each positive integer n , $k_{\lambda_n} + 1$ is in \bar{H} . On the other hand, $k_{\lambda_{n+1}}$ itself exceeds $k_{\lambda_n} + 1$ so that $k_{\lambda_{n+1}} + 1 > k_{\lambda_n} + 1$. These members of \bar{H} therefore constitute an infinite subset of the consequently infinite \bar{H} . Hence, H is hyper-simple.

Clearly, *every hyper-simple set H is simple*. For an infinite recursively enumerable subset of \bar{H} , as set of unit sequences, would contradict H being hyper-simple. Our construction of §6, in view of §8, gives us, however, *a simple set which is not hyper-simple*. Hyper-simple sets thus constitute a third class of recursively enumerable sets with recursively unsolvable decision problems—a class which is a proper subclass of the class of simple sets.

10. Non-reducibility of creative sets to hyper-simple sets by truth-tables unrestricted. Let creative set C be reducible by truth-tables to a recursively enumerable set of positive integers H . The given reduction will again determine for each positive integer n a finite sequence of positive integers m_1, m_2, \dots, m_v , and a truth-table T of order v such that that row of the table which correctly tells of the m 's whether they are, or are not, in H will correctly tell of n whether it is, or is not, in C . Of course v and T as well as the m 's depend on n , and the set of distinct T 's now entering into our reduction may be infinite, and hence the set of distinct v 's unbounded.

Let l_1, l_2, \dots, l_μ be any given finite sequence of distinct positive integers. A particular hypothesis on the l 's being, or not being, in H may then be symbolized by a sequence of μ signs, each + or -, such as $+ - \dots +$, such that the i th sign is +, or -, according as the hypothesis says that l_i is in H , or \bar{H} , respectively. We shall speak of such a sequence of signs as a *truth-assignment* for the l 's, the i th sign in that sequence as the *sign of l_i* in that truth-assignment. Of the 2^μ possible truth-assignments for the l 's, constituting a set V_1 , one and only one correctly tells of each l_i whether it is, or is not, in H . Every set V of truth-assignments for the l 's is then a subset of V_1 , and will be called a *possible set* of truth-assignments if it includes this *correct* truth-assignment.

Let then V be any given possible set of truth-assignments for the l 's. Let n be a positive integer with corresponding m_1, m_2, \dots, m_v , T yielded by the given reduction of C to H such that *each m not an l is in H*. The correct row of table T must then have the following two properties. First, the sign under each m not an l must be +. Second, the signs under those m 's which are l 's must be the same as the signs of those integers in some one and the same truth-assignment for the l 's in V , in fact, as in the correct truth assignment for the l 's. Any row

of T having these two properties, given the l 's, m 's and V , will be called a *relevant row* of T . Since for our n the correct row of T is thus a relevant row, it follows that n will surely be in \bar{C} if for each relevant row of T the sign under n is $-$.

Generate then the positive integers 1, 2, 3, \dots , and as a positive integer N is generated, generate the first N members of a given recursive enumeration of H , and for each n , with $1 \leq n \leq N$, find the corresponding m_1, m_2, \dots, m_r, T yielded by the given reduction of C to H . Of those m 's, if any, which are not l 's, see if each is one of those first N members of H . If they all are, see if for each relevant row of T the sign under n is $-$. If that also is true, place n in a set α_V . Since each such n must be in \bar{C} , as seen above, α_V is a subset of \bar{C} . And being a generated set, α_V is therefore a recursively enumerable subset of \bar{C} .

C being creative, we can therefore find a definite positive integer n' in \bar{C} but not in α_V , and, by the given reduction, the corresponding $m'_1, m'_2, \dots, m'_{r'}, T'$. Let $p_1, p_2, \dots, p_{\lambda}$ be those m 's, if any, which are not l 's. Now suppose that each p is in H . Then for at least one relevant row of T' the sign under n' must be $+$. For otherwise, if p is say the k_i th element in the given recursive enumeration of H , n' would have been placed in α_V in the above generation thereof for $N = \max(k_1, k_2, \dots, k_{\lambda}, n')$. Since n' is in \bar{C} , such a relevant row cannot be the correct row. But, with each p in H , the signs in that row under m 's that are not l 's are correctly $+$. Hence the sign under at least one m' that is an l must be incorrect. But, by our definition of a relevant row, the signs under all such m 's are the same as the signs of those integers in at least one truth-assignment in V . Such a truth-assignment in V cannot therefore be the correct truth-assignment for the l 's, and hence may be deleted from V . Perform this deletion for all such truth-assignments in V , and for all such relevant rows of T' , to obtain the set of truth-assignments V' . Under our hypothesis that each p is in H , V' will then be a proper subset of V , and yet a possible set of truth-assignments for the l 's.

Actually, let V be any given set of truth-assignments for the l 's, possible or not. Each step of the above construction can then still be carried out, though the constructed entities need not now have all the properties they otherwise possess.²² In particular, the set of integers, possibly null, $p_1, p_2, \dots, p_{\lambda}$ can be found, all different from any l . Likewise, whether the p 's are, or are not, all in H , the subset V' of V can be found. What we can say is that if V is a possible set,

²² Recall that in the definition of creative set, §3, each B determines an n , whether the α determined by B is, or is not, a subset of \bar{C} .

and if furthermore each p is in H , then V' is a proper subset of V , and itself is also a possible set of truth-assignments for the l 's.

For the given sequence of l 's, start then with $V = V_1$, the possible set of all 2^μ truth-assignments for the l 's, obtain the corresponding p 's, $p'_1, p'_2, \dots, p'_{\lambda'}$, and corresponding V' , $V_2 = (V_1)'$. With $V = V_2$, likewise find $p''_1, p''_2, \dots, p''_{\lambda''}$, and $V_3 = (V_2)'$, and so on. Now each V_{i+1} is a subset of V_i , while V_1 is but a finite set of 2^μ members. Hence in at most 2^μ steps we shall come across a V_k such that either V_{k+1} is identical with V_k , or is null. But if all the p'' 's, p''' 's, $\dots, p^{(\kappa)}$'s were in H , V_1 being a possible set, V_2, \dots, V_k as well as V_{k+1} would all be possible sets, each a proper subset of the preceding. V_{k+1} could not then either be identical with V_k , or null. It follows that at least one of the $p_j^{(j)}$'s with $1 \leq j \leq \kappa$ is in \bar{H} . Each $p_j^{(j)}$ is an integer that is not one of the l 's. If then we take this finite set of $p_j^{(j)}$'s and arrange them in a sequence of distinct elements in say order of magnitude, we obtain for our arbitrarily given sequence of distinct positive integers l_1, l_2, \dots, l_μ a sequence of distinct positive integers k_1, k_2, \dots, k_ν , having no element in common with the former sequence, and having at least one element in \bar{H} .

Starting with the null sequence as the sequence of l 's, we can thus find the sequence of k 's, $(k'_1, k'_2, \dots, k'_{\nu'})$ of distinct positive integers at least one of which is in \bar{H} . Inductively, let us have thus generated the sequences $(k'_1, k'_2, \dots, k'_{\nu'}), \dots, (k^{(\mu)}, k^{(\mu)}_2, \dots, k^{(\mu)}_{\nu^{(\mu)}})$, mutually exclusive, of distinct positive integers, each having at least one element in \bar{H} . With the single sequence $k'_1, \dots, k^{(\mu)}_{\nu^{(\mu)}}$ as the sequence of l 's, we can find the corresponding sequence of k 's, $(k^{(\mu+1)}, k^{(\mu+1)}_2, \dots, k^{(\mu+1)}_{\nu^{(\mu+1)}})$ of distinct positive integers with no element in common with any of the preceding sequences, and having at least one element in \bar{H} .

With creative C reducible to recursively enumerable H by truth-tables we can thus obtain an infinite generated, and hence recursively enumerable, set of mutually exclusive finite sequences of positive integers each having an element in \bar{H} . The set H is therefore not hyper-simple. Hence the

THEOREM. *No creative set is reducible to a hyper-simple set by truth-tables.*

COROLLARY. *Every hyper-simple set is of lower degree of unsolvability than the complete set K relative to reducibility by truth-tables.*

Despite this result, the brief discussion of Turing reducibility, still in the informal stage, entered into in the next section makes it dubious that hyper-simple sets as such will give us the desired absolutely

lower degree of unsolvability than that of K . But, in the absence of a counter-example, they remain candidates for this position.

11. General (Turing) reducibility. The process envisaged in our concept of a generated set may be said to be *polygenic*. In a *monogenic* process act succeeds act in one time sequence. The intuitive picture is that of a machine grinding out act after act (Turing [24]) or a set of rules directing act after act (Post [18]). The actual formulations are in terms of "atomic acts," the first leading to a development proved by Turing [25] equivalent to those arising from general recursive function or λ -definability, and hence of the same degree of generality. In our intuitive discussion the acts may be "molecular."

An effective solution of the decision problem for a recursively enumerable set S_1 of positive integers may therefore be thought of as a machine, or set of rules, which, given any positive integer n , will set up a monogenic process terminating in the correct answer, "yes" or "no," to the question "is n in S_1 ." Now suppose instead, says Turing [26] in effect, this situation obtains with the following modification. That at certain times the otherwise machine determined process raises the question is a certain positive integer in a given recursively enumerable set S_2 of positive integers, and that the machine is so constructed that were the correct answer to this question supplied on every occasion that arises, the process would automatically continue to its eventual correct conclusion.²³ We could then say that the machine effectively reduces the decision problem of S_1 to the decision problem of S_2 . Intuitively this should correspond to the most general concept of the reducibility of S_1 to S_2 . For the very concept of the decision problem of S_2 merely involves the answering for an arbitrarily given single positive integer m of the question is m in S_2 ; and in finite time but a finite number of such questions can be asked. A corresponding formulation of "Turing reducibility" should then be the same degree of generality for effective reducibility as say general recursive function is for effective calculability.²⁴

We may note that whereas in reducibility by truth-tables the posi-

²³ Turing picturesquely suggests access to an "oracle" which would supply the correct answer in each case. The "if" of mathematics is however more conducive to the development of a theory.

²⁴ A reading of McKinsey [16] suggested generalizing the reducibility of a recursively enumerable set S to a recursively enumerable set S' to the reducibility of S to a finite set of recursively enumerable sets S_1, S_2, \dots, S_n . However, no absolute gain in generality is thus achieved, as a single recursively enumerable set S' can be constructed such that reducing S to (S_1, S_2, \dots, S_n) is equivalent to reducing S to S' . Points of interest, however, do arise.

tive integers m of which we ask the questions "is m in S_2 " are effectively determined, for given n , by the reducibility process, in Turing reducibility, except for the first such m , the very identity of the m 's for which this question is to be asked depends, in general, on the correct answers having been given to these questions for all preceding m 's. The mode of this dependence is, however, effective, hence we still have effective reducibility in the intuitive sense.

Let now creative set C be Turing reducible to a recursively enumerable set S of positive integers. We shall talk as if our intuitive discussion has already been formalized. Generate the positive integers $1, 2, 3, \dots$, and as a positive integer N is generated, for each n with $1 \leq n \leq N$ proceed as follows. Set going the reducibility process of C to S for n . Each time a question of the form "is m in S " is met, see if m is among the first N integers in a given recursive enumeration of S . If it is, supply the answer "yes," thus enabling the reducibility process to continue. Finally, if under these circumstances the process terminates in a "no" for the initial question of n being in C , place n in a set α_0 . This α_0 is then a recursively enumerable subset of \bar{C} consisting of all members thereof for which the given Turing reduction of C to S leads only to questions of the form "is m in S " whose answer is "yes."

Find then n_0 of \bar{C} not in α_0 , and set the reducibility process going for n_0 . Now if at any time a wrong answer is supplied to a question "is m in S ," we can nevertheless expect our machine for reducing C to S either to effectively pick up the wrong answer and operate on it to give a next step in the process, or to cease operating. Generate then the positive integers $1, 2, 3, \dots$, and as a positive integer N is generated, generate the first N members of the given recursive enumeration of S , and make the reducibility process for n_0 *effective* though perhaps *incorrect* as follows. Each time a question of the form "is m in S " is reached, see if m is among the first N members of S . If it is, answer the question "yes," and correctly so; if not, answer the question "no," whether that answer be correct or no. If now this *pseudo-reduction* terminates in a "no," place the finite number of m 's thus arising in a set β_{n_0} . Note that β_{n_0} consists of all such m 's for all such pseudo-reductions for the given n_0 . Being a generated set of positive integers, β_{n_0} is *recursively enumerable*.

Now let the correct, though possibly non-effective, reducibility process for n_0 involve the μ questions "is m_i in S ," $i=1, 2, \dots, \mu$. Let $m_{i_1}, m_{i_2}, \dots, m_{i_\mu}$ be those of these m 's actually in S , and let them be the n_1 st, n_2 nd, \dots , n_μ th members of the given recursive enumeration of S . If then $N \geq M = \max(n_1, n_2, \dots, n_\mu)$, or $M=1$ if

$\nu = 0$, the corresponding psuedo-reduction for n becomes the correct reduction. For, inductively, if that be so through the time a question "is m in S " is raised, m will be m_1, m_2, \dots, m_μ , hence will, or will not, be in S according as it is, or is not, one of the first N members of S . The answer is then correctly given by that pseudo-reduction, which therefore continues to be correct through the raising of the next question. Finally, since n_0 is in \bar{C} , the correct reduction, now the pseudo-reduction, must terminate with a "no."

It follows that all N 's with $N > M$ merely repeat the contribution to β_{n_0} made by $N = M$, that is, of the integers m_1, m_2, \dots, m_μ . Since but a finite number of m 's are contributed by N 's with $N < M$, it follows that β_{n_0} is a finite set. Finally, were each of the integers m_1, m_2, \dots, m_μ in S , n_0 would be in α_0 . Hence, at least one member of β_{n_0} is in \bar{S} .

Formally, we would thus obtain a basis for a finite recursively enumerable set of positive integers at least one of whose members is in \bar{S} . Instead of recursively enumerable sets of finite sequences of positive integers, we would thus be led to consider recursively enumerable sets of bases for finite recursively enumerable sets of positive integers. Though, in the last analysis, each sequence in the former case must be generated atom by atom, there will come a time for each sequence when the process will say "this sequence is completed." In the latter case, in general, we cannot have an effective method which, for each basis, will give a point in the ensuing process at which it can say all members of the finite set in question have already been obtained, even though, with the process made monogenic, there always is such a stage in the process.

This suggests, then, that we strengthen the condition of hypersimplicity still further by replacing "infinitive recursively enumerable set of mutually exclusive finite sequences of positive integers" in the definition of §9 by "infinite recursively enumerable set of bases defining mutually exclusive finite recursively enumerable sets of positive integers." Whether such a "hyper-hyper-simple" set exists, or whether, if it exists, it will lead to a stronger non-reducibility result than that of the last section we do not know.

On the other hand, an equivalent definition of hyper-simple set is obtained if, for example, we replace the quoted phrase by "recursively enumerable set of finite sequences of positive integers having for each positive integer n a member each of whose elements exceeds n ." We now can say that with this as the definition of a hyper-simple set, the corresponding extension to a hyper-hyper-simple set cannot be made. For we prove the

THEOREM. *For any recursively enumerable set of positive integers S , with infinite \bar{S} , there exists a recursively enumerable set of bases defining finite recursively enumerable sets of positive integers, each set having at least one element in \bar{S} , and at least one set having each of its elements greater than an arbitrarily given positive integer n .*

Briefly, with n given, for each positive integer N , and each positive integer m , place all of the integers $n+1, n+2, \dots, n+m$ in a set α_n if all, or all but the last, are among the first N members of a given recursive enumeration of S . It is readily seen that α_n is a generated, and hence recursively enumerable, set of positive integers. A corresponding basis $B^{(n)}$ can actually be found, and the set of $B^{(n)}$'s, $n = 1, 2, 3, \dots$, being a generated set, is therefore recursively enumerable. Moreover, if v_n is the smallest integer in the infinite \bar{S} greater than n , α_n will consist of exactly the integers $n+1, n+2, \dots, v_n$, and hence will be finite, with indeed v_n as the only element in \bar{S} , and with each element greater than n .

As a result we are left completely on the fence as to whether there exists a recursively enumerable set of positive integers of absolutely lower degree of unsolvability than the complete set K , or whether, indeed, all recursively enumerable sets of positive integers with recursively unsolvable decision problems are absolutely of the same degree of unsolvability. On the other hand, if this question can be answered, that answer would seem to be not far off, if not in time, then in the number of special results to be gotten on the way.²⁵

Such then is the portion of "Recursive theory" we have thus far developed. In fixing our gaze in the one direction of answering the lower degree of unsolvability question, we have left unanswered many questions that stud even the short path we have traversed. Moreover, both our special, and the general Turing, definitions of reducibility are applicable to arbitrary decision problems whose questions in symbolic form are recursively enumerable, and indeed to problems with recursively enumerable set of questions whose answers belong to a recursively enumerable set. Thus, only partly leaving the field of decisions problems of recursively enumerable sets, work of Turing [26] suggests the question is the problem of determining of an arbitrary basis B whether it generates a finite, or infinite, set of positive

²⁵ This is a matter of practical concern as well as of theoretical interest. For according as the second or first of the above alternatives holds will the method of reducing new decision problems to problems previously proved unsolvable be, or not be, the general method for proving the unsolvability of decision problem either of recursively enumerable sets of positive integers or of problems equivalent thereto.

integers of absolutely higher degree of unsolvability than K . And if so, what is its relationship to that decision problem of absolutely higher degree of unsolvability than K yielded by Turing's theorem.

Actually, the theory of recursive reducibility can be but one chapter in the theory of recursive unsolvability, and that, but one volume of the theory and applications of general recursive functions. Indeed, if general recursive function is the formal equivalent of effective calculability, its formulation may play a rôle in the history of combinatorial mathematics second only to that of the formulation of the concept of natural number.

BIBLIOGRAPHY

1. Alonzo Church, *An unsolvable problem of elementary number theory*, Amer. J. Math. vol. 58 (1936) pp. 345–363.
2. ———, *A note on the Entscheidungsproblem*, Journal of Symbolic Logic vol. 1 (1936) pp. 40–41.
3. ———, *Correction to a note on the Entscheidungsproblem*, ibid. pp. 101–102.
4. ———, *Combinatory logic as a semi-group*, Preliminary report, Bull. Amer. Math. Soc. abstract 43-5-267.
5. ———, *The constructive second number class*, ibid. vol. 44 (1938) pp. 224–232.
6. ———, *The calculi of lambda-conversion*, Annals of Mathematics Studies, no. 6, Princeton University Press, 1941.
7. C. J. Ducasse, *Symbols, signs, and signals*, Journal of Symbolic Logic vol. 4 (1939) pp. 41–52.
8. Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik vol. 38 (1931) pp. 173–198.
9. ———, *On undecidable propositions of formal mathematical systems*, mimeographed lecture notes, The Institute for Advanced Study, 1934.
10. David Hilbert, *Mathematical problems*. Lecture delivered before the International Congress of Mathematicians at Paris in 1900. English translation by Mary Winston Newsom, Bull. Amer. Math. Soc. vol. 8 (1901–1902) pp. 437–479.
11. David Hilbert and Paul Bernays, *Grundlagen der Mathematik*, vol. 2, Julius Springer, Berlin, 1939.
12. S. C. Kleene, *General recursive functions of natural numbers*, Math. Ann. vol. 112 (1936) pp. 727–742.
13. ———, *On notation for ordinal numbers*, Journal of Symbolic Logic vol. 3 (1938) pp. 150–155.
14. ———, *Recursive predicates and quantifiers*, Trans. Amer. Math. Soc. vol. 53 (1943) pp. 41–73.
15. C. I. Lewis, *A survey of symbolic logic*, Berkley, 1918, chap. 6, §3.
16. J. C. C. McKinsey, *The decision problem for some classes of sentences without quantifiers*, Journal of Symbolic Logic vol. 8 (1943) pp. 61–76.
17. Rozsa Péter, *Az axiomatikus módszer korlátai* (The bounds of the axiomatic method), Review of, Journal of Symbolic Logic vol. 6 (1941) pp. 111.
18. Emil L. Post, *Finite combinatory processes—formulation 1*, ibid. vol. 1 (1936) pp. 103–105.

19. ———, *Formal reductions of the general combinatorial decision problem*, Amer. J. Math. vol. 65 (1943) pp. 197–215.
20. J. B. Rosser, *Extensions of some theorems of Gödel and Church*, Journal of Symbolic Logic vol. 1 (1936) pp. 87–91.
21. ———, *An informal exposition of proofs of Gödel's theorems and Church's theorem*, ibid. vol. 4 (1939) pp. 53–60.
22. Th. Skolem, *Einfacher beweis der unmöglichkeit eines allgemeinen losungsverfahrens für arithmetische probleme*, Review of, Mathematical Reviews vol. 2 (1941) p. 210.
23. Alfred Tarski, *On undecidable statements in enlarged systems of logic and the concept of truth*, Journal of Symbolic Logic vol. 4 (1939) pp. 105–112.
24. A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, Proc. London Math. Soc. (2) vol. 42 (1937) pp. 230–265.
25. ———, *Computability and λ -definability*, Journal of Symbolic Logic vol. 2 (1937) pp. 153–163.
26. ———, *Systems of logic based on ordinals*, Proc. London Math. Soc. (2) vol. 45 (1939) pp. 161–228.

THE CITY COLLEGE,
NEW YORK CITY.

A VARIANT OF A RECURSIVELY UNSOLVABLE PROBLEM

BY

E. L. POST

Reprinted from the
BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY
Vol. 52, No. 4, pp. 264-268
April, 1946

A VARIANT OF A RECURSIVELY UNSOLVABLE PROBLEM

EMIL L. POST

By a string on a, b we mean a row of a 's and b 's such as $baabbbab$. It may involve only a , or b , or be null. If, for example, g_1, g_2, g_3 represent strings bab, aa, b respectively, string $g_2g_1g_1g_3g_2$ on g_1, g_2, g_3 will represent, in obvious fashion, the string $aabbabbbaa$ on a, b . By the *correspondence decision problem* we mean the problem of determining for an arbitrary finite set $(g_1, g'_1), (g_2, g'_2), \dots, (g_\mu, g'_\mu)$ of pairs of corresponding non-null strings on a, b whether there is a solution in n, i_1, i_2, \dots, i_n of equation

$$(1) \quad g_{i_1}g_{i_2}\cdots g_{i_n} = g'_{i_1}g'_{i_2}\cdots g'_{i_n}, \quad n \geq 1, i_j = 1, 2, \dots, \mu.$$

That is, whether some non-null string on g_1, g_2, \dots, g_μ , and corresponding string on $g'_1, g'_2, \dots, g'_\mu$, represent identical strings on a, b .

In special cases, of course, the question posed by (1) may be answerable. Thus, if, with $\mu=3$, $(g_1, g'_1), (g_2, g'_2), (g_3, g'_3)$ are $(bb, b), (ab, ba), (b, bb)$ respectively, $g_1g_2g_3 = bbababb = g'_1g'_2g'_3$, and (1) has a solution. Again, if each g_i is of greater length than the corresponding g'_i , or if each g_i starts with a different letter than the corresponding g'_i , (1) has no solution. We proceed to prove, on the other hand, that in its full generality the *correspondence decision problem is recursively unsolvable*,¹ and hence, no doubt, unsolvable in the intuitive sense.

We start with the known recursive unsolvability of the decision problem for the class of normal systems on a, b .² A normal system S on

Received by the editors October 20, 1945.

¹ It suffices here to consider "recursively unsolvable" to mean unsolvable in the sense of Church [1]. Of course the general problem remains recursively unsolvable if we allow null g 's and g' 's. Numbers in brackets refer to the references cited at the end of the paper.

² See [4, §2] for an informal proof. As far as the printed literature is concerned, we must refer to [2] for a formal proof, though there then remains the actual verification, via Gödel representations, that the reduction effected is indeed recursive. This verification, at least for the reduction of S' to S'' [2, p. 51], is immediate if we use the following simpler method of reducing S' to a system S'' in canonical form than that given by Church. The primitive symbols of our S'' are those of S' and one additional primitive symbol α . The basis of S'' in part consists of the two primitive assertions $\alpha I, \alpha J$, and the operation $\alpha P, \alpha Q$ produce $\alpha(PQ)$. It will follow that αP is asserted in S'' when and only when P is a combination without free variables. The remainder of the basis of S'' consists of the primitive assertion of S' as primitive assertion, and the thirty-eight operations of S' each modified as follows. For each operational variable P occurring in the operation, αP is introduced as additional premise.

a, b is given by a basis consisting of an initial non-null string A on a, b , and a finite set of operations $\alpha_i P$ produces $P\alpha'_i$, $i=1, 2, \dots, v$, where the α 's and α' 's are given strings on a, b , while the operational variable P represents an arbitrary string on a, b , possibly null. The assertions of S consist of A and all non-null strings obtainable from A by repeated use of the v operations. The known recursively unsolvable problem is then the problem of determining for arbitrary S , as given by a basis therefore, and arbitrary non-null string B on a, b , whether B is an assertion of S . This unsolvability is undisturbed if the α 's and α' 's are all non-null,³ a condition which automatically excludes the possibility of null assertions, and will henceforth be assumed.

Referring to operation $\alpha_i P$ produces $P\alpha'_i$ by the subscript i , string B on a, b will be an assertion of S when and only when some finite sequence of operations i_1, i_2, \dots, i_n leads from A to B . Now operation $\alpha_i P$ produces $P\alpha'_i$ can be applied to string C to yield string D when and only when for some string P , possibly null, $C = \alpha_i P$, $P\alpha'_i = D$. Hence B is an assertion of S when and only when the following set of equations has a solution in n, i_1, i_2, \dots, i_n , and the P 's.

$$(2) \quad A = \alpha_{i_1} P_1, P_1 \alpha'_{i_1} = \alpha_{i_2} P_2, \dots, P_{n-1} \alpha'_{i_{n-1}} = \alpha_{i_n} P_n, P_n \alpha'_{i_n} = B.$$

Here n may be 0, (2) then becoming $A = B$. We proceed to show that (2) is equivalent to a single equation somewhat like (1) subject, however, to certain length conditions.

Given (2), we can eliminate the P 's by forming $A\alpha'_{i_1}\alpha'_{i_2} \dots \alpha'_{i_n}$ and successively substituting for the left members of (2) the right to obtain

$$(3) \quad A\alpha'_{i_1}\alpha'_{i_2} \dots \alpha'_{i_n} = \alpha_{i_1}\alpha_{i_2} \dots \alpha_{i_n} B.$$

Likewise, starting with $A\alpha'_{i_1}\alpha'_{i_2} \dots \alpha'_{i_{m-1}}$, we obtain

$$(4) \quad A\alpha'_{i_1} \dots \alpha'_{i_{m-1}} = \alpha_{i_1} \dots \alpha_{i_m} P_m,$$

whence,

$$(5) \quad \text{length}(A\alpha'_{i_1} \dots \alpha'_{i_{m-1}}) \geq \text{length}(\alpha_{i_1} \dots \alpha_{i_m}), \quad m = 1, 2, \dots, n,$$

the length of a string being the total number of occurrences of letters therein, here a 's and b 's. Conversely, let (3) be given, with (5) satisfied. With the length of $\alpha_{i_1} \dots \alpha_{i_m}$ less than or equal to that of $A\alpha'_{i_1} \dots \alpha'_{i_{m-1}}$, (3) shows that the former must be identical with an

³ It suffices to modify the production starting on page 214 of [3] in accordance with footnote 3 thereof to insure that the final normal system has no g or g' null.

"initial segment" of the latter. Hence, P_m can be determined so that (4) is satisfied, and for $m = 1, 2, \dots, n$. For $m = 1$, (4) yields the first equation of (2). By substituting the right side of (4), with $m = j$, for the left, (4) for $m = j+1$ becomes $\alpha_{i_1} \dots \alpha_{i_j} P_i \alpha_{i_j} = \alpha'_{i_1} \dots \alpha_{i_j} \alpha_{i_{j+1}} P_{j+1}$, whence $P_j \alpha'_j = \alpha_{i_{j+1}} P_{j+1}$, $j = 1, 2, \dots, n-1$. Likewise, the last equation of (2) is obtained from (3) via (4) for $m = n$. Hence, (2) has a solution when and only when (3) has a solution subject to (5). That is, B is in normal system S when and only when (3) has a solution in n, i_1, i_2, \dots, i_n subject to (5). Comparing (3) with (1), we see that to reduce the decision problem for the class of normal systems on a, b to the correspondence decision problem, and thus have the unsolvability of the former lead to the unsolvability of the latter, we must on the one hand eliminate the length condition (5), on the other, the A and B of (3).

We achieve the first aim by reducing normal system S in three stages to a normal system in which (3) implies (5). If $C = x_1 x_2 \dots x_n$, the x 's a 's or b 's, let $\bar{C} = x_n \dots x_2 x_1$. For a letter with subscript, superscript, we shall only bar the letter. Now for the normal system S on a, b with initial string A and operations $\alpha_i P$ produces $P\alpha'_i$, $i = 1, 2, \dots, v$, form the system S' , not normal, with initial string \bar{A} , and operations $P\bar{\alpha}_i$ produces $\bar{\alpha}'_i P$. Clearly, string B on a, b will be an assertion of S when and only when \bar{B} is an assertion of S' . Next form S'' with initial string $\bar{A}h$, and operations $P\bar{\alpha}_i h$ produces $\bar{\alpha}'_i Ph$. String \bar{B} is then in S' when and only when $\bar{B}h$ is in S'' . We finally form a normal system S''' , though on the three letters a, b, h , whose assertions are the assertions of S'' and all cyclic permutations thereof, a cyclic permutation of string $x_1 \dots x_i x_{i+1} \dots x_n$ here meaning any string $x_{i+1} \dots x_n x_1 \dots x_i$. The initial string of S''' is again $\bar{A}h$. For its first v operations we take $\bar{\alpha}_i h P$ produces $Ph\bar{\alpha}'_i$, premise and conclusion being a cyclic permutation of the premise and conclusion of the corresponding operation of S'' . We finally add the operations aP produces Pa , bP produces Pb , hP produces Ph which serve to transform a string on a, b, h into any of its cyclic permutations. System S''' is therefore normal, and, by induction, is easily seen to have the stated property.⁴ It follows that B is an assertion of normal system S when and only when $\bar{B}h$ is an assertion of normal system S''' .

Let the operations of S''' be resymbolized $\beta_i P$ produces $P\beta'_i$, $i = 1, 2, \dots, v+3$. Though S''' is a normal system on three letters, the discussion of equations (2)–(5) is equally applicable to it. Hence B is an assertion of S when and only when the following equation (6) has a solution subject to (7).

⁴ Cf. [3], final reduction.

$$(6) \quad \bar{A}h\beta'_{i_1}\beta'_{i_2} \cdots \beta'_{i_n} = \beta_{i_1}\beta_{i_2} \cdots \beta_{i_n} \bar{B}h.$$

$$(7) \quad \text{length } (\bar{A}h\beta'_{i_1} \cdots \beta'_{i_{m-1}}) \geq \text{length } (\beta_{i_1} \cdots \beta_{i_m}), \quad m = 1, 2, \dots, n.$$

Suppose (6) had a solution with (7) not satisfied for a certain m . For that m , the length of $\beta_{i_1} \cdots \beta_{i_m}$ would exceed the length of $\bar{A}h\beta'_{i_1} \cdots \beta'_{i_{m-1}}$, and hence, in virtue of (6), we would have

$$(8) \quad \beta_{i_1} \cdots \beta_{i_m} = \bar{A}h\beta'_{i_1} \cdots \beta'_{i_{m-1}} Q$$

with non-null Q . Recall that (β_i, β'_i) is $(\bar{\alpha}h, h\bar{\alpha}')$ for $i \leq v$, (a, a) , (b, b) , (h, h) for the three remaining i 's. With α 's and α' 's on a, b only, β_i and β'_i are then either both free from h , or have exactly one h apiece. Were the β_{i_m} of (8) a or b , the right side of (8) would have at least one more occurrence of h than the left, which is impossible. In any other case, β_{i_m} ends with h . Non-null Q therefore ends with h , and again the right side of (8) would have at least one more h than the left. Hence, every solution of (6) must satisfy (7). That is, B is an assertion of S when and only when (6) has a solution.

The elimination of $\bar{A}h$ and $\bar{B}h$ from (6) is more easily effected. Corresponding to the $v+3$ couples (β_i, β'_i) , $i=1, 2, \dots, v+3$, and $\bar{A}h, \bar{B}h$, we introduce $v+5$ couples (γ_i, γ'_i) as follows. With x 's and y 's representing letters, in this case a, b , or h , if β_i and β'_i are $x_1x_2 \cdots x_k$ and $y_1y_2 \cdots y_\lambda$ respectively, γ_i and γ'_i are to be $x_1kx_2k \cdots x_kk$ and $ky_1ky_2 \cdots ky_\lambda$ respectively. If $\bar{A}h$ and $\bar{B}h$ are $y_1y_2 \cdots y_\lambda$ and $x_1x_2 \cdots x_k$ respectively, γ_{v+4} and γ'_{v+4} are to be kk and $kky_1ky_2 \cdots ky_\lambda$, γ_{v+5} and γ'_{v+5} , $x_1kx_2k \cdots x_kkk$ and kk , respectively. It then follows that (6) has a solution in n, i_1, i_2, \dots, i_n , $n \geq 0$, $i_p = 1, 2, \dots, v+3$, when and only when the equation

$$(9) \quad \gamma'_{j_1}\gamma'_{j_2} \cdots \gamma'_{j_m} = \gamma_{j_1}\gamma_{j_2} \cdots \gamma_{j_m}$$

has a solution in m, j_1, j_2, \dots, j_m , $m \geq 1, j_q = 1, 2, \dots, v+5$. In fact, if i_1, i_2, \dots, i_n makes both sides of (6) equal to $z_1z_2 \cdots z_l$, $(j_1, j_2, \dots, j_m) = (v+4, i_1, i_2, \dots, i_n, v+5)$ makes both sides of (9) equal to $kkz_1kz_2 \cdots kz_1kk$. On the other hand, suppose (9) has a solution. Then since γ'_{j_1} and γ_{j_1} must start with the same letter, $j_1 = v+4$. For in every other case γ'_{j_1} starts with k , γ_{j_1} with a, b , or h . Similarly, $j_m = v+5$, (9) forcing γ'_{j_m} and γ_{j_m} to end with the same letter. If, now, the intermediate j 's are all different from $v+4$ and $v+5$, they directly give a sequence of i 's satisfying (6). Otherwise, let j_μ be the first j beyond j_1 that is $v+4$ or $v+5$. Were $j_\mu = v+4$, $\gamma'_{j_1}\gamma'_{j_2} \cdots \gamma'_{j_\mu}$ and $\gamma_{j_1}\gamma_{j_2} \cdots \gamma_{j_\mu}$ would take the forms $kkx_1kx_2 \cdots kx_pkkx_{p+1}k \cdots x_q$ and $kky_1ky_2 \cdots y_rkkk$ respectively, with x 's and y 's a, b , or h . But then the second occurrence of kk in the left side of (9) would be im-

mediately followed by a , b , or h , in the right side, by k , contradicting (9). Hence $j_\mu = \nu + 5$, and $\gamma'_{j_1}\gamma'_{j_2} \cdots \gamma'_{j_\mu}$ and $\gamma_{j_1}\gamma_{j_2} \cdots \gamma_{j_\mu}$ consequently take the form $kkx_1kx_2 \cdots kx_pkk$ and $kky_1ky_2 \cdots ky_qkk$. But the left side of (9) through the second occurrence of kk must equal the right side of (9) through its second occurrence of kk . Hence $\gamma'_{j_1}\gamma'_{j_2} \cdots \gamma'_{j_\mu} = \gamma_{j_1}\gamma_{j_2} \cdots \gamma_{j_\mu}$ and we have a solution of (9) of the type previously seen to lead to a solution of (6). It follows that B is an assertion of S when and only when (9) has a solution.

In the reduction thus effected we have introduced the new letters h and k . But now in (γ_i, γ'_i) replace the letters a, b, h, k by $bab, baab, baaab, baaaab$ respectively,⁵ and call the resulting pair of strings on $a, b, (\delta_i, \delta'_i)$. Then (9) is seen to be equivalent to

$$(10) \quad \delta'_{j_1}\delta'_{j_2} \cdots \delta'_{j_m} = \delta_{j_1}\delta_{j_2} \cdots \delta_{j_m},$$

immediately so in passing from (9) to (10), and conversely. For if, for example, δ'_{j_1} starts with $baab$, δ_{j_1} must also start with $baab$, and likewise for the next group of letters, and so on till (10) is seen to be a translation of (9).

Given normal system S on a, b with basis A , $\alpha_i P$ produces $P\alpha'_i$, $i = 1, 2, \dots, \nu$, and string B on a, b , the above gives an effective method for forming the pairs of strings on $a, b, (\delta_i, \delta'_i)$, $i = 1, 2, \dots, \nu + 5$, such that B is an assertion of S when and only when (10) has a solution. But (10), with left and right hand members interchanged, is a case of (1). We have therefore effectively reduced the decision problem for the class of normal systems on a, b to our correspondence decision problem. If, then, the former is unsolvable in the intuitive sense, so must be the latter. Actually, by introducing Gödel representations, we readily verify that the above effective reduction is indeed recursive, the recursive unsolvability of the former problem then leading to the recursive unsolvability of the correspondence decision problem.

REFERENCES

1. Alonzo Church, *An unsolvable problem of elementary number theory*, Amer. J. Math. vol. 58 (1936) pp. 345–363.
2. ———, Review of [3], Journal of Symbolic Logic vol. 8 (1943) pp. 50–52.
3. Emil L. Post, *Formal reductions of the general combinatorial decision problem*, Amer. J. Math. vol. 65 (1943) pp. 197–215.
4. ———, *Recursively enumerable sets of positive integers and their decision problems*, Bull. Amer. Math. Soc. vol. 50 (1944) pp. 284–316.

THE CITY COLLEGE,
NEW YORK CITY

⁵ Cf. [4], footnote 5.

NOTE ON A CONJECTURE OF SKOLEM

EMIL L. POST

In his excellent review of four notes of Skolem on recursive functions of natural numbers¹ Bernays states: "The question whether every relation $y = f(x_1, \dots, x_n)$ with a recursive function f is primitive recursive remains undecided." Actually, the question is easily answered in the negative by a form of the familiar diagonal argument.

We start with the ternary recursive relation R , referred to in the review, such that $R(x, y, 0), R(x, y, 1), \dots$ is an enumeration of all binary primitive recursive relations. Define $f(x)$ to be $x + 1$ when $R(x, x, x)$ is true, x , when $R(x, x, x)$ is false. Then, on the one hand, $y = f(x)$ differs extensionally from each primitive recursive relation $R(x, y, n)$, $n = 0, 1, \dots$. For with $x = n$, $y = n$, if $R(x, y, n)$, that is, $R(n, n, n)$, is true, $y = f(x)$ becomes $n = n + 1$ and is false; if $R(x, y, n)$ is false, $y = f(x)$ becomes $n = n$ and is true. Hence, $y = f(x)$ is not primitive recursive. On the other hand $f(x)$ is recursive as shown by its formal expression:

$$f(x) = \epsilon y \{ [R(x, x, x) \rightarrow y = x + 1] \& [\overline{R(x, x, x)} \rightarrow y = x] \}.$$

The Skolem-Gödel *abschätzung* $y \leq x + 1$ is not needed to insure the recursiveness of $f(x)$ since the $\phi(x, y)$ constituting the brace satisfies the Kleene condition $(x)(Ey)\phi(x, y)$.

For the rôle played by the above question in Skolem's notes we refer the reader to Bernays's very clear and detailed review.

On being acquainted with the above result, Bernays suggested the following reformulation of its proof in terms of functions instead of relations. "Let $\phi(x, y, n)$ be the enumeration, with repetitions, of all primitive recursive functions, definable by a two-fold recursion (as stated by Rózsa Péter in Math. Annalen 111, 1935), and put $\chi(x) = x + (1 - \phi(x, x, x))$; then the function $\chi(x)$ is general recursive, but the relation $y = \chi(x)$ is not primitive recursive. For if there existed a primitive recursive function $\psi(x, y)$ such that $\psi(x, y) = 0 \Leftrightarrow y = \chi(x)$, there would be a number n such that $\phi(x, y, n) = 0 \Leftrightarrow y = \chi(x)$, and we should have for this number n : $\phi(n, n, n) = 0 \Leftrightarrow n = n + (1 - \phi(n, n, n))$, and thus $\phi(n, n, n) = 0 \Leftrightarrow (1 - \phi(n, n, n) = 0)$," which "leads to a contradiction."

Actually, we may think of $R(x, y, n)$ as being given by the equation $\phi(x, y, n) = 0$, in which case our $f(x)$ and Bernays's $\chi(x)$ are indeed identical.

The above example is a little misleading as to method since it diagonalizes both the x and y of $R(x, y, n)$, whereas it suffices to diagonalize x only. Thus, another example of a recursive $f(x)$ with $y = f(x)$ not primitive recursive is obtained by defining $f(x)$ to be 1 when $R(x, 0, x)$ is true, 0, when $R(x, 0, x)$ is false. In general, for $y = f(x)$ not to be primitive recursive it is necessary and sufficient

Received July 6, 1946.

¹ In this JOURNAL, vol. 11 (1946), pp. 26-28.

that it differ extensionally from each $R(x, y, n)$, $n = 0, 1, \dots$; hence, that for each natural number n there be a corresponding (x_n, y_n) such that $R(x_n, y_n, n)$ and $y_n = f(x_n)$ have opposite truth-values. For the immediate consistency of the definition of $f(x)$ we shall want the x_n 's distinct. The corresponding y_n 's can then be chosen arbitrarily. For x not an x_n , $f(x)$ may be defined arbitrarily. For x an x_n , we define $f(x)$ so that if $R(x_n, y_n, n)$ is true, $y_n \neq f(x_n)$, if $R(x_n, y_n, n)$ is false, $y_n = f(x_n)$. If the various processes involved are recursive, $f(x)$ will then be recursive, yet $y = f(x)$ not primitive recursive.

It would seem, however, that the x_n 's being distinct, for immediate consistency, is the distinguishing feature arising from the special form $y = f(x)$ rather than that only x in $R(x, y, n)$ need be diagonalized. Thus, a binary relation $S(x, y)$ of arbitrary form will fail to be primitive recursive when and only when for each natural number n there is a corresponding (x_n, y_n) such that $R(x_n, y_n, n)$ and $S(x_n, y_n)$ have opposite truth-values. To obtain such an $S(x, y)$, the procedure outlined for $y = f(x)$ could again be followed. But it could also be generalized; since for the immediate consistency of the definition of $S(x, y)$, we need not have the x_n 's distinct, but only the pairs (x_n, y_n) .

THE CITY COLLEGE
COLLEGE OF THE CITY OF NEW YORK

RECURSIVE UNSOLVABILITY OF A PROBLEM OF THUE

EMIL L. POST

Alonzo Church suggested to the writer that a certain problem of Thue [6]¹ might be proved unsolvable by the methods of [5]. We proceed to prove the problem recursively unsolvable, that is, unsolvable in the sense of Church [1], but by a method meeting the special needs of the problem.

Thue's (general) problem is the following. Given a finite set of symbols a_1, a_2, \dots, a_n , we consider arbitrary *strings* (Zeichenreihen) on those symbols, that is, rows of symbols each of which is in the given set. Null strings are included. We further have given a finite set of pairs of corresponding strings on the a_i 's, $(A_1, B_1), (A_2, B_2), \dots, (A_n, B_n)$. A string R is said to be a *substring* of a string S if S can be written in the form URV , that is, S consists of the letters, in order of occurrence, of some string U , followed by the letters of R , followed by the letters of some string V . Strings P and Q are then said to be *similar* if Q can be obtained from P by replacing a substring A_i or B_i of P by its correspondent B_i, A_i . Clearly, if P and Q are similar, Q and P are similar. Finally, P and Q are said to be *equivalent* if there is a finite set R_1, R_2, \dots, R_r of strings on a_1, \dots, a_n such that in the sequence of strings $P, R_1, R_2, \dots, R_r, Q$ each string except the last is similar to the following string. It is readily seen that this relation between strings on a_1, \dots, a_n , is indeed an equivalence relation. Thue's problem is then the problem of determining for arbitrarily given strings A, B on a_1, \dots, a_n whether, or no, A and B are equivalent.

This problem, at least for the writer, is more readily placed if it is restated in terms of a special form of the canonical systems of [3]. In that notation, strings C and D are similar if D can be obtained from C by applying to C one of the following operations:

$$PA_iQ \text{ produces } PB_iQ, PB_iQ \text{ produces } PA_iQ, i = 1, 2, \dots, n. \quad (1)$$

In these operations the operational variables P, Q represent arbitrary strings. Strings A and B will then be equivalent if B can be obtained from A by starting with A , and applying in turn a finite sequence of operations (1). That is, A and B are equivalent if B is an assertion in the "canonical system"² with primitive assertion A and operations (1). Thue's general problem thus becomes the decision problem for the class of all canonical systems of this "Thue type."

This general problem could easily be proved recursively unsolvable if, instead of the pair of operations for each i of (1), we merely had the first operation of each pair.³ In fact, by direct methods such as those of [3], we easily reduce the decision problem of an arbitrary "normal system" [3] to the decision problem of such a system of "semi-Thue type," the known recursive unsolvability of the

Received October 26, 1946. Presented to the American Mathematical Society November 2, 1946.

¹ Numbers in brackets refer to the bibliography at the end of the paper.

² Null assertions, however, now being allowed.

³ That is, using the language of propositions instead of operations, if we merely had an implication where (1) has an equivalence.

decision problem for the class of all normal systems then, no doubt, leading to the recursive unsolvability of the decision problem for the class of all semi-Thue systems. The crux of our method for handling the Thue systems themselves is to find such a reduction of a known unsolvable problem to a system of semi-Thue type that when, for each i , the second of the two operations in (1) is added to the semi-Thue system, no new assertions are thereby added to the system. The known unsolvable problem is thus reduced to the resulting Thue system, as desired. Such a reduction turns out to be possible for a certain unsolvable problem arising in the theory of Turing machines.

We shall adopt the following formulation of a Turing machine [7].⁴ A two-way infinite linear tape is provided, ruled off into squares. Time is a one-way infinite sequence of discrete moments. A square will either be blank, or have at most one symbol printed upon it. At any moment the machine "scans" one of the squares. At such a moment the machine is capable of performing one of the following atomic acts: moving one square to the left, moving one square to the right, printing on the scanned square one of a given finite number of symbols S_1, \dots, S_m , or a blank. Following Turing, we take "printing" here to mean "overprinting," that is, the letter or blank printed replaces any letter that may have been on the scanned square. Printing a blank is then equivalent to erasing, when the scanned square is not blank. The machine, furthermore, is capable of assuming but a finite number of internal states, internal configurations or m -configurations with Turing, q_1, q_2, \dots, q_R . At any moment, the letter or blank on the scanned square together with the internal configuration of the machine determines the atomic act to be performed by the machine and the new internal configuration of the machine, or else, the machine then stops. At the initial moment a finite, possibly null, number of squares have S 's printed on them, the machine scans a particular square and has a particular internal configuration.

Symbolically, the machine may be given as follows. Let S_o be used to represent a blank square. For the start of the action of the machine we need only consider the smallest unbroken piece of the tape containing the initially marked squares and the scanned square, replace these squares by their markings, or by S_o if blank, and insert the symbol q_{i_1} of the initial internal configuration prior to the S of the scanned square to yield the representation

$$S_{j_1} S_{j_2} \cdots S_{j_{k-1}} q_{i_1} S_{j_k} \cdots S_{j_\kappa}. \quad (2)$$

A finite number of quadruplets of symbols of the three forms,

$$q_i S_j L q_i, \quad q_i S_j R q_i, \quad q_i S_j S_k q_i, \quad (3)$$

will then determine the behavior of the machine. Here, q_i and S_j represent the internal configuration of the machine, and the symbol or blank on the scanned

⁴ Apart from the Turing convention, discussed in the appendix, this differs from Turing's formulation of an automatic machine in the nature of the tape, and in Turing's use, in his standard form [7, p. 240], of the composite operation "print and move" where we just have "move." A number of comparisons with [2] will occur to a reader of that note.

square, at any moment; L , R , or S_k the correspondingly determined atomic act of motion left, right, or printing of S_k ; q_i the determined new internal configuration of the machine. It is fundamental that the pairs q_i, S_j of the several quadruplets are distinct, for they are to determine uniquely the consequent behavior of the machine. The machine will then continue acting deterministically from the initial moment on unless, and until, a $q_i S_j$ is reached for which there is no quadruplet (3), in which case it will stop.

We can now readily set up a semi-Thue system whose assertions will represent the successive states of the tape, and the relation of the machine thereto, as (2) represented these at the initial moment. However, for simplicity, the portion of the tape represented, while including the marked squares and the scanned square, need not now be the smallest such portion.⁵ Because of the particular needs of the semi-Thue form, we introduce a new symbol h . Each assertion of the semi-Thue system will then be in the form hPh with P free from h . If A represents the string (2) of S 's and one q , the initial assertion of the semi-Thue system will be hAh . For each quadruplet (3) corresponding to moving one square to the left, we introduce the operations

$$PS_n q_i S_j Q \text{ produces } Pq_i S_n S_j Q, n = 0, 1, \dots, m, \quad (4)$$

$$Phq_i S_j Q \text{ produces } Phq_i S_o S_j Q. \quad (5)$$

Note that (4) takes care of all cases where the scanned square is not the leftmost square of the part of the tape represented at the given moment, (5) where the scanned square is that leftmost square. Due to the form hPh of all assertions of the system, when (5) is applicable, the h of the premise thereof must be the leftmost of these two h 's, so that P will be identified with the null string. The S_o of the conclusion then takes care of the necessary extension of the portion of the tape represented when the motion is one square to the left of that portion. Likewise, for each quadruplet (3) corresponding to motion of one square to the right we introduce

$$Pq_i S_j S_n Q \text{ produces } PS_j q_i S_n Q, n = 0, 1, \dots, m, \quad (6)$$

$$Pq_i S_j h Q \text{ produces } PS_j q_i S_o h Q; \quad (7)$$

while for each quadruplet (3) corresponding to the printing of S_k over the scanned square we have

$$Pq_i S_j Q \text{ produces } Pq_i S_k Q. \quad (8)$$

Clearly both premise and conclusion of each operation thus introduced is of the form PBQ with fixed B , so that we do thus have a semi-Thue system. An obvious induction yields the form hPh with P free from h for each assertion. Likewise, each assertion has one and only one q therein. Finally, it is readily

⁵ It could be made the smallest such portion by using more operations. There would then be a 1-1 correspondence between the intrinsic states of tape versus machine and the representations thereof.

verified from the deterministic character of the Turing machine, and from the forms of the above operations, that at most one of these operations is applicable to any string having no more than one occurrence of a q therein, and then in only one way.

The unsolvable problem that is to yield the unsolvability of the problem of Thue would seem to be furnished by the following result of Turing's [7, p. 248]: "There can be no machine \mathfrak{E} which, when supplied with the S.D. of an arbitrary machine \mathfrak{M} , will determine whether \mathfrak{M} ever prints a given symbol (0 say)." There are, however, difficulties in using this result as given due to peculiarities of Turing's development. (The matter is discussed in the appendix.) We therefore proceed independently of Turing as follows.

We start with the known recursive unsolvability of the decision problem for the class of normal systems on two letters a, b .⁶ It suffices here to think of this problem as consisting of a class of questions, each question Q being symbolized by a string on a given finite set of letters. By methods such as those used by Turing in setting up his universal computing machine [7], we then set up the quadruplets (3) of a fixed Turing machine with certain letters S_1, S_2, \dots, S_m , and internal configurations q_1, q_2, \dots, q_n , and give an effective method for translating each question Q into a Q' of form (2) to serve as the initial state of tape versus machine, the construction being such that the following is true. The answer to question Q is yes, or no, according as the constructed machine, when applied to Q' , does, or does not, in the course of its operation print a certain fixed letter S_p . This letter S_p is not present in the Q' of any Q . Since such methods are fully exploited by Turing in [7], we do not give the details of this construction.⁷

Now, given Q , form the semi-Thue system T' with initial assertion $hQ'h$, and operations (4)–(8) corresponding to this Turing machine. Then, the answer to Q is yes, or no, according as some assertion of T' involves the letter S_p , or no assertion involves that letter. We now modify T' as follows. Delete all operations in T' such that the S_j of the premise, the symbol on the scanned square of the Turing machine, is S_p . Since, when S_p is first printed, it can appear so only as the S_k of (8), and thus would be the S_j of a next operation, the deductive processes of the semi-Thue system will now stop the first time S_p appears in an assertion. We now add operations which, in deterministic fashion, will erase all of this assertion except for the two h 's and the q , while changing this q . For this purpose, we introduce two new "internal configurations"

⁶ See [5, footnote 2]. The specific form of this problem, however, need not be known by the reader for an understanding of the present argument.

⁷ This work was carried through before the definitive study of Turing's paper [7], referred to in the appendix, was made. As a result, some differences of method appear. A minor difference is that where Turing uses the method of "marking" a sequence of symbols [7, p. 235] to distinguish it, we introduce the effect of movable physical markers; two, indeed, suffice. A major difference is that instead of the m -configuration functions of Turing's skeleton tables [7, p. 236], we introduce a symbolism and technique based on the concept of a subset of directions of a given set of directions. Both differences were suggested by [2]. They may, perhaps, better be exploited in a more general setting.

q_{R+1} and q_{R+2} . We further alter the operations of T' by changing the q_i of each operation (8) for which S_i is S_p to q_{R+1} , and add the following operations:

$$PS_nq_{R+1}Q \text{ produces } Pq_{R+1}Q, n = 0, 1, \dots, m; n \neq p. \quad (9)$$

$$Phq_{R+1}Q \text{ produces } Phq_{R+2}Q. \quad (10)$$

$$Pq_{R+2}S_nQ \text{ produces } Pq_{R+2}Q, n = 0, 1, \dots, m. \quad (11)$$

Note that as a result of the previous changes, when S_p first appears in an assertion, the q therein is q_{R+1} . Operations (9) then serve to erase the S 's of that assertion to the left of q_{R+1} , (10) then changes q_{R+1} to q_{R+2} , (11) erases the S 's to the right of q_{R+2} . Finally, therefore, the assertion becomes $hq_{R+2}h$, to which no further operation is applicable. Call the resulting semi-Thue system T'' . Clearly, for T'' it is also true that at most one of its operations is applicable to any string having no more than one occurrence of a q therein, and then in only one way. It follows that the answer to Q is yes, or no, according as $hq_{R+2}h$ is, or is not, an assertion in T'' , the operations of T'' operating one by one in deterministic fashion, and, in the former case, terminating in $hq_{R+2}h$.

The proof of the reducibility of our initial unsolvable problem to the problem of Thue essentially becomes the proof of the following two lemmas.⁸ By the inverse of an operation of the form PAQ produces PBQ we shall mean the operation PBQ produces PAQ . Let T''' be the semi-Thue system with primitive assertion $hq_{R+2}h$ and operations the inverses of those of T'' . We then have:

LEMMA I. The primitive assertion $hq_{R+2}h$ of T''' is an assertion of T'' when, and only when, the primitive assertion $hQ'h$ of T'' is an assertion of T''' .

Proof. D is a result of applying "PAQ produces PBQ" to C when, and only when, C is a result of applying the inverse operation "PBQ produces PAQ" to D . For both statements are equivalent to the existence of strings P and Q such that $PAQ = C$, $PBQ = D$. If, then, operations O_1, O_2, \dots, O_n of T'' lead from its primitive assertion $hQ'h$ through assertions C_1, C_2, \dots, C_{n-1} to the assertion $hq_{R+2}h$, the inverses of these operations, all in T''' , will in reverse order lead from $hq_{R+2}h$, the primitive assertion of T''' , through C_{n-1}, \dots, C_2, C_1 to $hQ'h$; and conversely.

As a result of Lemma I, the answer to question Q is yes, or no, according as $hQ'h$ is, or is not, an assertion of T''' . Note that while the initial assertion of T'' depended on Q , T''' is the same for all Q 's. Now let T be the Thue system obtained from the semi-Thue system T''' by adding to the latter the inverse of each of its operations. We then have:

LEMMA II. The class of assertions of T is identical with the class of assertions of T''' .

Proof. Each assertion of T''' is, of course, an assertion of T . For the converse, let operations O_1, O_2, \dots, O_n of T lead from its primitive assertion $hq_{R+2}h$, through assertions C_1, C_2, \dots, C_{n-1} , to an assertion C of T . If n is zero, C is $hq_{R+2}h$, and hence an assertion of T''' . Otherwise, note that the operations of T , being those of T''' and their inverses, are the combined operations of T'''

⁸ These lemmas can be made more general.

and of T'' . Now we saw that no operation of T'' is applicable to $hq_{R+2}h$, the deductive processes of T'' terminating in $hq_{R+2}h$ if leading thereto. Hence, operation O_1 must be in T''' . Assume O_{m+1} to be the first O not in T''' , and hence in T'' . Since O_m is in T''' , its inverse is in T'' . As O_m operates on $C_{m-1}(hq_{R+2}h)$, if m is one) to yield C_m , the inverse of O_m is applicable to C_m yielding C_{m-1} . That is, both the inverse of O_m , and O_{m+1} , are operations in T'' applicable to C_m , and yielding C_{m-1} and C_{m+1} (C , if $m + 1 = n$) respectively. Now, since the premise and conclusion of each operation in T , and the primitive assertion of T , have exactly one occurrence of a q therein, the same is true of every assertion of T . But we saw that at most one operation of T'' is applicable to a string with a single occurrence of a q therein, and then in only one way. It follows that O_{m+1} is in fact the inverse of O_m , and hence C_{m+1} is C_{m-1} all over again. We may therefore delete operations O_m and O_{m+1} from the given sequence of operations, and still have a sequence of operations leading from $hq_{R+2}h$ to C . By repeating this process, we finally obtain such a sequence of O 's with each O in T''' . The arbitrary assertion C of T is therefore also an assertion of T''' .⁹

Hence, $hQ'h$ is an assertion of T''' when and only when it is an assertion of T . Finally, then, the answer to Q is yes, or no, according as $hQ'h$ is, or is not, an assertion of the Thue system T . In terms of the language of Thue, we have then a fixed set of pairs of strings $(A_1, B_1), \dots, (A_n, B_n)$ leading to a definition of equivalence of strings such that the answer to Q is yes, or no, according as $hQ'h$ is, or is not, equivalent to the fixed string $hq_{R+2}h$.¹⁰ Certainly, then, a solution of the problem of Thue in its full generality would thus lead to a solution of the "decision problem for the class of normal systems on a, b ." By the use of Gödel representations, the recursive unsolvability of the latter problem then easily leads to the recursive unsolvability of the problem of Thue.

A few concluding remarks may be in order. The methods of [5], and of the present paper, do have something in common, a something we may call *the method of the irrelevant modification*. Once an unsolvable problem has been obtained by a *reductio ad absurdum* argument based on the definition of solvability, the usual method of proving a new problem unsolvable is to reduce a known unsolvable problem to this given problem. In the method of the irrelevant modification, the known unsolvable problem is reduced to a problem which on modification becomes the given problem, while that modification does not affect the answers to the individual questions. In [5] the modification is a simplification, the existence of a solution of a certain string equation subject to

⁹ Briefly, then, the effect of operations of T'' on deductive processes of T is to unravel work done by operations of T''' . Note that while the deductive processes of T'' give rise to a single sequence of assertions starting with $hQ'h$ and terminating in $hq_{R+2}h$, if leading thereto, the deductive processes of T''' give rise to a tree of assertions, elements not necessarily distinct, stemming from $hq_{R+2}h$, and containing the above sequence in reverse when that sequence terminates in $hq_{R+2}h$.

¹⁰ By the method of the next to the last paragraph of [5], this definition of equivalence could be transformed into a definition of equivalence for strings on the two letters a, b . We have not paused to prove the recursive unsolvability of the resulting special case of the problem of Thue.

certain length conditions being equivalent to the mere existence of a solution of that string equation. In the present paper the modification is a complication, the answer to $hQ'h$ being or not being an assertion of T''' being unaffected by adding to the operations of T''' their inverses.

The writer has often felt that the multiplicity of equivalent formulations of recursiveness has been a deterrent to the general promulgation of this discipline. Yet, the writer's normal systems naturally lead to the unsolvable problem of [5], while the deterministic character of the Turing machine is basic to the above unsolvability proof. From this point of view, the several formulations of recursiveness are so many different instruments for tackling new unsolvability proofs.

Though we have not paused to verify this formally, it seems rather obvious that when the problem of [5], and the problem of Thue, are translated via positive integers as suggested in [4], they become decision problems of recursively enumerable sets of positive integers of the same degree of unsolvability as the complete set K , at worst, with respect to many-one reducibility [4]. This indicates how far practice lags behind theory in this field.

Appendix. The following critique of Turing's "computability" paper [7] concerns only pp. 230-248 thereof. We have checked the work through the construction of the "universal computing machine" in detail;¹¹ but the proofs of the two theorems in the section following are there given in outline only, and we have not supplied the formal details. *We have therefore also left in intuitive form the proofs of the statements on recursiveness, and alternative procedures, we make below.*

Turing's definition of an arbitrary machine is not completely given in his paper, and, at a number of points, has to be inferred from his development. In the first instance his machine is a "computing machine" for obtaining the successive digits of a real number in dyadic notation, and, in that case, starts operating on a blank tape. Where explicitly stated, however, the machine may

¹¹ One major correction is needed. To the instructions for $\text{con}_1(\mathcal{C}, \alpha)$ p. 244, add the line: None $PD, R, P\alpha, R, R, R \quad \mathcal{C}$. This is needed to introduce the representation D of the blank scanned square when, as at the beginning of the action of the machine, or due to motion right beyond the rightmost previous point, the complete configuration ends with a q , and thus make the fmp of p. 244 correct. We may also note the following minor slips and misprints in pp. 230-248. Page 236, to the instructions for $f(\mathcal{C}, \mathcal{B}, \alpha)$ add the line: None $L \quad f(\mathcal{C}, \mathcal{B}, \alpha)$; p. 240 and p. 241, the S.D should begin, but not end, with a semicolon; p. 242, omit the first D in (C_2) ; p. 243, last paragraph, add ":" to the first list of symbols; pp. 244-246, replace g by q ; p. 245, in the instruction for mf , mf should be mf_1 ; p. 245, in the second instruction for smt_2 , replace the first R by L ; p. 245, in the first instruction for sh_2 , replace sh_2 by sh_3 . A reader of the paper will be helped by keeping in mind that the "examples" of pages 236-239 are really parts of the table for the universal computing machine, and accomplish what they are said to accomplish not for all possible printings on the tape, but for certain ones that include printings arising from the action of the universal computing machine. In particular, the tape has \circ printed on its first two squares, the occurrence of two consecutive blank squares insures all squares to the right thereof being blank, and, usually, symbols referred to are on "F-squares," and obey the convention of p. 235.

start operating on a tape previously marked. From Turing's frequent references to the beginning of the tape, and the way his universal computing machine treats motion left, we gather that, unlike our tape, this tape is a one-way infinite affair going right from an initial square.

Primarily as a matter of practice, Turing makes his machines satisfy the following convention. Starting with the first square, alternate squares are called *F*-squares, the rest, *E*-squares. In its action the machine then never directs motion left when it is scanning the initial square, never orders the erasure, or change, of a symbol on an *F*-square, never orders the printing of a symbol on a blank *F*-square if the previous *F*-square is blank and, in the case of a computing machine, never orders the printing of 0 or 1 on an *E*-square. This convention is very useful in practice. However the actual performance, described below, of the universal computing machine, coupled with Turing's proof of the second of the two theorems referred to above, strongly suggests that Turing makes this convention part of the definition of an arbitrary machine. We shall distinguish between a Turing machine and a Turing convention-machine.

By a uniform method of representation, Turing represents the set of instructions, corresponding to our quadruplets,¹² which determine the behavior of a machine by a single string on seven letters called the standard description (S.D) of the machine. With the letters replaced by numerals, the S.D of a machine is considered the arabic representation of a positive integer called the description number (D.N) of the machine. If our critique is correct, a machine is said to be circle-free if it is a Turing computing convention-machine which prints an infinite number of 0's and 1's.¹³ And the two theorems of Turing's in question are really the following. There is no Turing convention-machine which, when supplied with an arbitrary positive integer n , will determine whether n is the D.N of a Turing computing convention-machine that is circle-free. There is no Turing convention-machine which, when supplied with an arbitrary positive integer n , will determine whether n is the D.N of a Turing computing convention-machine that ever prints a given symbol (0 say).¹⁴

In view of [8], these "no machine" results are no doubt equivalent to the re-

¹² Our quadruplets are quintuplets in the Turing development. That is, where our standard instruction orders either a printing (overprinting) or motion, left or right, Turing's standard instruction always orders a printing and a motion, right, left, or none. Turing's method has certain technical advantages, but complicates theory by introducing an irrelevant "printing" of a symbol each time that symbol is merely passed over.

¹³ "Genuinely prints," that is, a genuine printing being a printing in an empty square. See the previous footnote.

¹⁴ Turing in each case refers to the S.D of a machine being supplied. But the proof of the first theorem, and the second theorem depends on the first, shows that it is really a positive integer n that is supplied. Turing's proof of the second theorem is unusual in that while it uses the unsolvability result of the first theorem, it does not "reduce" [4] the problem of the first theorem to that of the second. In fact, the first problem is almost surely of "higher degree of unsolvability" [4] than the second, in which case it could not be "reduced" to the second. Despite appearances, that second unsolvability proof, like the first, is a *reductio ad absurdum* proof based on the definition of unsolvability, at the conclusion of which, the first result is used.

cursive unsolvability of the corresponding problems.¹⁵ But both of these problems are infected by the spurious Turing convention. Actually, the set of n 's which are D.N.'s of Turing computing machines as such is recursive, and hence the condition that n be a D.N. offers no difficulty. But, while the set of n 's which are not D.N.'s of convention-machines is recursively enumerable, the complement of that set, that is, the set of n 's which are D.N.'s of convention-machines, is not recursively enumerable. As a result, in both of the above problems, neither the set of n 's for which the question posed has the answer yes, nor the set for which the answer is no, is recursively enumerable.

This would remain true for the first problem even apart from the convention condition. But the second would then become that simplest type of unsolvable problem, the decision problem of a non-recursive recursively enumerable set of positive integers [4]. For the set of n 's that are D.N.'s of unrestricted Turing computing machines printing 0, say, is recursively enumerable, though its complement is not. The Turing convention therefore prevents the early appearance of this simplest type of unsolvable problem.

It likewise prevents the use of Turing's second theorem in the above unsolvability proof of the problem of Thue. For in attempting to reduce the problem of Turing's second theorem to the problem of Thue, when an n leads to a Thue question for which the answer is yes, we would still have to determine whether n is the D.N. of a Turing convention-machine before the answer to the question posed by n can be given, and that determination cannot be made recursively for arbitrary n . If, however, we could replace the Turing convention by a convention that is recursive, the application to the problem of Thue could be made. An analysis of what Turing's universal computing machine accomplishes when applied to an arbitrary machine reveals that this can be done.

The universal computing machine was designed so that when applied to the S.D. of an arbitrary computing machine it would yield the same sequence of 0's and 1's as the computing machine as well as, and through the intervention of, the successive "complete configurations"—representations of the successive states of tape versus machine—yielded by the computing machine. This it does for a Turing convention-machine.¹⁶ For an arbitrary machine, we have to interpret a direction of motion left at a time when the initial square of the tape is scanned as meaning no motion.¹⁷ The universal computing machine will then yield again the correct complete configurations generated by the given machine. But *the space sequence of 0's and 1's printed by the universal computing machine will now be identical with the time sequence of those printings of 0's and 1's by the given machine that are made in empty squares.* If, now, instead of Turing's

¹⁵ Our experience with proving that "normal unsolvability" in a sense implicit in [3] is equivalent to unsolvability in the sense of Church [1], at least when the set of questions is recursive, suggests that a fair amount of additional labor would here be involved. That is probably our chief reason for making our proof of the recursive unsolvability of the problem of Thue independent of Turing's development.

¹⁶ Granted the corrections given in footnote 11.

¹⁷ This modification of the concept of motion left is assumed throughout the rest of the discussion, with the exception of the last paragraph.

convention we introduce the convention that the instructions defining the machine never order the printing of a 0 or 1 except when the scanned square is empty, or 0, 1 respectively, and never order the erasure of a 0 or 1, Turing's arguments again can be carried through. And this "(0, 1) convention," being recursive, allows the application to the problem of Thue to be made.¹⁸ Note that if a machine is in fact a Turing convention-machine, we could strike out any direction thereof which contradicts the (0, 1) convention without altering the behavior of the machine, and thus obtain a (0, 1) convention-machine. But a (0, 1) convention-machine need not satisfy the Turing convention. However, by replacing each internal-configuration q_i of a machine by a pair q_i, q'_i to correspond to the scanned square being an *F*- or an *E*-square respectively, and modifying printing on an *F*-square to include testing the preceding *F*-square for being blank, we can obtain a "(q, q') convention" which is again recursive, and usable both for Turing's arguments and the problem of Thue, and has the property of, in a sense, being equivalent to the Turing convention. That is, every (q, q') convention-machine is a Turing convention-machine, while the directions of every Turing convention-machine can be recursively modified to yield a (q, q') convention-machine whose operation yields the same time sequence and spatial arrangement of printings and erasures as does the given machine, except for reprintings of the same symbol in a given square.

These changes in the Turing convention, while preserving the general outline of Turing's development and at the same admitting of the application to the problem of Thue, would at least require a complete redoing of the formal work of the proof of the second Turing theorem. On the other hand, very little added formal work would be required if the following changes are made in the Turing argument itself, though there would still remain the need of extending the equivalence proof of [8] to the concept of unsolvability. By using the above result on the performance of the universal computing machine when applied to the S.D of an arbitrary machine, we see that Turing's proof of his first theorem, whatever the formal counterpart thereof is, yields the following theorem. There is no Turing convention-machine which, when supplied with an arbitrary positive integer n , will determine whether n is the D.N of an arbitrary Turing machine that prints 0's and 1's in empty squares infinitely often. Now given an arbitrary positive integer n , if that n is the D.N of a Turing machine \mathfrak{M} , apply the universal computing machine to the S.D of \mathfrak{M} to obtain a machine \mathfrak{M}^* . Since \mathfrak{M}^* satisfies the Turing convention, whatever Turing's formal proof of his second theorem is, it will be usable intact in the present proof, and, via the new form of his first theorem, will yield the following usable result. There is no machine which, when supplied with an arbitrary positive integer n , will

¹⁸ So far as recursiveness is concerned, the distinction between the Turing convention and the (0, 1) convention is that the former concerns the history of the machine in action, the latter only the instructions defining the machine. Likewise, despite appearances, the later (q, q') convention.

determine whether n is the D.N. of an arbitrary Turing machine that ever prints a given symbol (0 say).¹⁹

These alternative procedures assume that Turing's universal computing machine is retained. However, in view of the above discussion, it seems to the writer that Turing's preoccupation with computable numbers has marred his entire development of the Turing machine. We therefore suggest a redevelopment of the Turing machine based on the formulation given in the body of the present paper. This could easily include computable numbers by defining a computable sequence of 0's and 1's as the *time sequence* of printings of 0's and 1's by an arbitrary Turing machine, provided there are an infinite number of such printings. By adding to Turing's complete configuration a representation of the act last performed, a few changes in Turing's method would yield a universal computing machine which would transform such a time sequence into a space sequence. Turing's convention would be followed as a matter of useful practice in setting up this, and other, particular machines. But it would not infect the theory of arbitrary Turing machines.

REFERENCES

- [1] Alonzo Church, *An unsolvable problem of elementary number theory*, *American journal of mathematics*, vol. 58 (1936), pp. 345-363.
- [2] Emil L. Post, *Finite combinatory processes—formulation 1*, this JOURNAL, vol. 1 (1936), pp. 103-105.
- [3] Emil L. Post, *Formal reductions of the general combinatorial decision problem*, *American journal of mathematics*, vol. 65 (1943), pp. 197-215.
- [4] Emil L. Post, *Recursively enumerable sets of positive integers and their decision problems*, *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 284-316.
- [5] Emil L. Post, *A variant of a recursively unsolvable problem*, *ibid.*, vol. 52 (1946), pp. 264-268.
- [6] Axel Thue, *Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln*, *Skrifter utgit av Videnskapsseksjonet i Kristiania*, I. Matematisk-naturvidenskabelig klasse 1914, no. 10 (1914), 34 pp.
- [7] A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, *Proceedings of the London Mathematical Society*, ser. 2 vol. 42 (1937), pp. 230-265.
- [8] A. M. Turing, *Computability and λ -definability*, this JOURNAL, vol. 2 (1937), pp. 153-163.

THE CITY COLLEGE
COLLEGE OF THE CITY OF NEW YORK

¹⁹ It is here assumed that the suggested extension of [8] includes a proof to the effect that the existence of an arbitrary Turing machine for solving a given problem is equivalent to the existence of a Turing convention-machine for solving that problem.

THE UPPER SEMI-LATTICE OF DEGREES OF RECURSIVE UNSOLVABILITY

S. C. KLEENE AND EMIL L. POST

(Received June 22, 1953)

The concept 'degree of recursive unsolvability' was introduced briefly in Post [16]. In his abstract [17] the concept was formulated precisely via an extension of [15], and a resulting partial scale of degrees of recursive unsolvability was applied to strengthen Theorem II of Kleene [8]. In the present paper our interest is in the abstract structure of the system of the degrees of recursive unsolvability.

The concept 'degree of recursive unsolvability' is based on that of reducibility of decision problems. Three precise formulations of the latter concept have appeared, more or less completely, in the literature.

In 'Turing reducibility' [19] §4, the concept of a Turing machine [18] is generalized to that of a Turing reducibility machine. A Turing reducibility machine \mathfrak{M} reduces the decision problem for a set S to that for a set T , if for each positive integer n , the machine \mathfrak{M} applied to n terminates in the correct answer to the question whether n is in S , via ordinary Turing machine acts and the hypothetically correct answering of such questions of the form "Is m in T ?" as may arise in the process (finite in number).

In 'general recursive reducibility', the representing function of S is general recursive in that of T in the sense of Kleene [8].¹

In Post's 'canonical reducibility' [17], his canonical sets [15] are generalized to T -canonical sets by hypothetically adding primitive assertions expressing the membership or non-membership of 1, 2, 3, ... in T . Then S is canonically reducible to T , if both S and its complement \bar{S} with respect to the set of all positive integers are T -canonical sets. Although an infinite collection of hypothetical primitive assertions are added to an otherwise ordinary canonical system, in any derivation of an assertion in the T -canonical system but a finite number of them enter.

Turing reducibility was proved equivalent to canonical reducibility by Post (unpublished), and canonical reducibility to general recursive reducibility by Martin Davis in his typed thesis [5]. A direct proof of the equivalence of Turing reducibility to general recursive reducibility was published (independently) by Kleene in his book [10] end §§68, 69.

Theoretically, no increase in generality is obtained by the use of Kleene's concept of a function general recursive in other functions (as may be seen from [10] pp. 307, 291); and by talking about just sets S and T a greater simplicity of concept results.

Practically, when it comes to giving rigorous demonstrations under one or another of the equivalent concepts of reducibility, there are advantages in using

¹ Further references are given in 1.1 below.

number-theoretic functions freely, and not simply sets (or functions of one variable only admitting only two numbers as values). Moreover, a detailed development of the theory of reducibility, in a form convenient for our applications, is in print at this writing only in [10] in terms of functions recursive in other functions. Accordingly the body of this paper is written in terms of that notion. Although in Post's papers the range of the independent variables has been the positive integers $1, 2, 3, \dots$, we shall here use instead the natural numbers $0, 1, 2, \dots$ for conformity with the notations of [10].

In the preliminary §1 we abstract from degrees of unsolvability of sets to degrees as such, and recognize the degrees as forming an upper semi-lattice under a least upper bound operation $a \cup b$. We also introduce a jump operation a' , which corresponds essentially to an added alternating quantifier of [8]. The previous work of Kleene [6], [8], [10], Mostowski [13], [14], Post [17] and Davis [5] merely singled out the degrees yielded by this jump operation, starting with the degree $\mathbf{0}$ of solvability, and extended them into the transfinite.² Now the incomparability and betweenness results of §2 yield an unbounded finite, and therefore an infinite, number of degrees passed over by each jump, as well as a profusion of incomparable degrees flanked by it. In §3 further information is obtained by undertaking a simultaneous construction of infinitely many degrees passed over by a jump; we find such a set of degrees which is dense. The present paper may therefore be said to give us some knowledge of the fine structure of the upper semi-lattice of degrees; yet many questions will be pointed out which remain unsettled concerning this structure. In §4 it is shown that the upper semi-lattice is not a lattice. Classical methods are used.³

1. Degrees of recursive unsolvability

1.1. Hereafter we cite [10] as "IM". The (general) recursiveness of a function ϕ in functions ψ_1, \dots, ψ_i is defined on IM p. 275. The notion extends to predicates P, Q_1, \dots, Q_i in place of $\phi, \psi_1, \dots, \psi_i$, and to mixed cases in which some but not all of $\phi, \psi_1, \dots, \psi_i$ are replaced by predicates, by use of the representing functions of the predicates; i.e. a statement of general recursiveness holds among predicates or predicates and functions, if the statement holds when each of the predicates is replaced by its representing function, which takes 0 or 1 as value according as the value of the predicate is true or false (IM pp. 276, 227). Thence it extends to sets by considering for each set S the predicate $a \in S$ or directly by use of representing functions of the sets (IM p. 307).

We use some notations of IM and the accompanying facts concerning recursiveness. In particular (noting that primitive recursive functions are general recursive, Theorem II p. 275), we use $* * 1-13$ pp. 222-223; the intuitive sym-

² In [11] Kleene considers certain greater jumps.

³ A manuscript of this paper was made available to the Seminar on the Foundations of Mathematics at the University of Wisconsin in February 1953, with the result that a number of the questions pointed out as open have been answered by Mr. Clifford Spector. Publication of his results is planned.

bolism of p. 225; * * A-F pp. 224, 228, 229; * * 14-21 pp. 227, 229, 230; $\ddot{\phi}$ pp. 231, 291; μ p. 279.⁴

1.2. Consider any two objects A and B , each either a number-theoretic function, predicate or set.⁵

The relation ' A is recursive in B ' is reflexive and transitive (by IM Theorem II p. 275).

Hence the relation ' A is recursive in B and B is recursive in A ' is reflexive, symmetric and transitive. So the latter relation divides the class of all functions, predicates and sets into disjoint non-empty equivalence classes, such that each two objects are in the same equivalence class if and only if they are in the relation. We define A and B to have the *same degree of unsolvability*, if they stand in the latter relation; this enables the degrees to be identified with the equivalence classes. Now the original relation ' A is recursive in B ' can be construed as a relation $a \leq b$ between the degrees a of A and b of B ; i.e. if A is recursive in B , then each A_1 of the same degree as A is recursive in each B_1 of the same degree as B . Under this relation, the degrees are a partially ordered system.

In symbols, we thus have, for arbitrary degrees a and b ,

- (1) $a \leq a,$
- (2) $a \leq b \ \& \ b \leq c \rightarrow a \leq c,$
- (3) $a \leq b \ \& \ b \leq a \equiv a = b.$

We define $a < b \equiv a \leq b \ \& \ a \neq b$ (so $a \leq b \equiv a < b \vee a = b$), $a \geq b \equiv b \leq a$, $a > b \equiv b < a$, and read the symbols as usual. Also we say $a \mid b$ (a is *incomparable* with b) if neither $a \leq b$ nor $b \leq a$. On the ground that either A is recursive in B or not, and vice versa, exactly one of

$$a < b, \quad a = b, \quad a > b, \quad a \mid b$$

must hold for each two degrees a and b ; that the fourth case $a \mid b$ actually occurs will appear in §2.

The name "degree of recursive unsolvability", which we take from [17], is appropriate to the case the degree is that of a non-recursive set or predicate; such a set or predicate has an unsolvable decision problem ([16] Introduction, IM p. 301).

However, since any recursive B is a fortiori recursive in any A , the recursive functions, predicates and sets constitute a degree 0 ("solvability"), and for every degree a

$$(4) \quad 0 \leq a,$$

so 0 is the lowest degree. That there is no highest degree will appear from the end of this subsection or from (11) in 1.4.

To any function or predicate there is a set having the same degree (IM p.

⁴ These notations are indexed in IM bottom p. 538.

⁵ It is only for terminological convenience that we mention predicates and sets separately from their representing functions.

307), so no more degrees arise by allowing functions and predicates as well as sets. By the definition of recursiveness (1.1), inversely any degree of a set or predicate is a degree of a function. To any degree one can choose a function or predicate of that degree of any number ≥ 1 of variables.

Given any function ψ , by definition (IM p. 275) a function ϕ is recursive in ψ if ϕ is defined recursively from ψ by some system E of equations in the formalism of recursive functions; the same E cannot define recursively different ϕ 's from the given ψ . There are only a countable infinity (\aleph_0) of systems E of equations in the fixed formalism of recursive functions. Hence there are at most \aleph_0 functions, and thence predicates and sets, recursive in a given function ψ , i.e. of degree $\leq a$ given degree a , a fortiori of the same degree a . But changing a finite number of the values of a function does not change its degree, so each degree actually consists of \aleph_0 functions etc. With the 2^{\aleph_0} functions etc. thus separated into mutually exclusive countably infinite equivalence classes, it follows that there are 2^{\aleph_0} of the classes, i.e. 2^{\aleph_0} degrees. But there are at most \aleph_0 degrees $\leq a$ given degree a , since there are only \aleph_0 functions etc. of these degrees.

1.3. Given two sets etc. A and B , consider any third set etc. D in the following relationship to A and B :

- (i) A is recursive in D , and B is recursive in D ,
- (ii) D is recursive in A , B .

First, we observe that to any A and B there are such D 's. Say A and B are sets, let $A = \delta A(a)$ and $B = \delta B(a)$, and write α and β for the respective representing functions. Then the following function $\delta(a)$ and predicates $D(a)$ and $D(a, k)$ are three examples of D 's for the given A and B :

$$\delta(a) = 2^{\alpha(a)} \cdot 3^{\beta(a)}, \quad \begin{cases} D(2a) \equiv A(a), \\ D(2a + 1) \equiv B(a), \end{cases} \quad \begin{cases} D(a, 0) \equiv A(a), \\ D(a, k + 1) \equiv B(a). \end{cases}$$

Secondly, by (ii) D is recursive in any C in which both A and B are recursive. So in particular, if (i) and (ii) both hold of A_1, B_1, D_1 and also of A_2, B_2, D_2 , where A_1 is of the same degree as A_2 and B_1 as B_2 , then D_1 is of the same degree as D_2 . So all the D 's which satisfy (i) and (ii) for a given A and B are of one degree (and in fact constitute that degree, since any D_1 of the same degree as D also satisfies (i) and (ii)); and this degree is determined by the degrees a of A and b of B . We write this degree $a \cup b$. Then, in symbols,

$$(5) \quad a \leq a \cup b \text{ and } b \leq a \cup b,$$

$$(6) \quad a \leq c \& b \leq c \equiv a \cup b \leq c.$$

For arbitrary degrees a and b , the degree $a \cup b$ is thus their least upper bound (or "join"); so the partially ordered system of the degrees constitutes an upper semi-lattice.⁶

⁶ Using (7), (7a), (8) and (9), the degrees come under the definition of semi-lattice as formulated in Birkhoff [1] p. 18 Ex. 1 and Footnote 6 when his " \sqcap " and " \geq " are changed to " \cup " and " \leq ".

Algebraically from (1)–(6), or directly from (i) and (ii),

$$(7) \quad a \leq b \equiv a \cup b = b,$$

$$(7a) \quad a \cup a = a, \quad (7b) \quad 0 \cup a = a,$$

$$(8) \quad a \cup b = b \cup a,$$

$$(9) \quad a \cup (b \cup c) = (a \cup b) \cup c.$$

By (9) parentheses can be omitted in writing continued joins. In fact $a_1 \cup \dots \cup a_n$ ($n \geq 1$) is then exactly the degree of the sets etc. D in the following relationship to given sets etc. A_1, \dots, A_n of degrees a_1, \dots, a_n , respectively:

$$(i_n) \quad A_i \text{ is recursive in } D \quad (i = 1, \dots, n),$$

$$(ii_n) \quad D \text{ is recursive in } A_1, \dots, A_n.$$

Consistently with this, we define $a_1 \cup \dots \cup a_n$ for $n = 0$ to be 0.

Now, if b, a_1, \dots, a_n ($n \geq 0$) are the degrees of B, A_1, \dots, A_n , respectively: B is recursive in A_1, \dots, A_n , if and only if $b \leq a_1 \cup \dots \cup a_n$.

Sometimes we shall write $A_1 \cup \dots \cup A_n$ ($n \geq 0$) for some set etc. of the degree $a_1 \cup \dots \cup a_n$, where a_1, \dots, a_n are the degrees of A_1, \dots, A_n , respectively. We do this only in contexts where it either is immaterial, or has been indicated, which particular such set etc. is to be used.

We say that n sets A_1, \dots, A_n ($n \geq 1$) are *recursively independent*, or that their degrees a_1, \dots, a_n are *independent*, if none of A_1, \dots, A_n is recursive in the rest, or equivalently if for no k ($k = 1, \dots, n$) is $a_k \leq a_1 \cup \dots \cup a_{k-1} \cup a_{k+1} \cup \dots \cup a_n$.

If a_1, \dots, a_n are independent, so are any subset of them.

For $n = 1$, independence is equivalent to non-recursiveness; for $n = 2$, to incomparability. For $n > 2$, independence implies pairwise incomparability, but in view of the following not conversely.

In §2 we will show that for every $n \geq 1$ there exist n independent degrees a_1, \dots, a_n . Form the joins b_1, \dots, b_{2^n} of these degrees in all the 2^n possible combinations. It is easily seen that then $b_i \leq b_j$ if and only if b_i is a subcombination of b_j . So for example (with $n \geq 3$), $a_2 \cup a_3, a_1 \cup a_3, a_1 \cup a_2$ are pairwise incomparable (e.g. were $a_2 \cup a_3 \leq a_1 \cup a_3$, using (5) and (2) we could infer $a_2 \leq a_1 \cup a_3$, contradicting the independence of a_1, a_2, a_3), but not independent (e.g. $a_2 \cup a_3 \leq (a_1 \cup a_3) \cup (a_1 \cup a_2)$, by (5), (6), (2)).

1.4. Given any set etc. A , consider the 1-place predicates $(Ex)R^A(a, x)$ where $R^A(a, x)$ is any 2-place predicate recursive in A . (These may be called the *A-generable* predicates.) It is known that in this class of predicates there are par-

⁷ The non-empty sets $\{\}(Ex)R^A(a, x)$ are familiar for A recursive as the “recursively enumerable sets” [6], and in general are the “sets enumerable recursively in A ” IM p. 308. Post [16] included the empty set by fiat. In this paper when the empty set is included we are changing the term “enumerable” to “generable”; “ A -generable” is then short for “generable recursively in A ”.

ticular predicates $C^A(a)$ with the property that each predicate $(Ex)R^A(a, x)$ of the class is expressible as $C^A(\phi(a))$ for some recursive function ϕ (in fact, ϕ can be primitive recursive and such that $a \neq b \rightarrow \phi(a) \neq \phi(b)$); a fortiori each $(Ex)R^A(a, x)$ is recursive in $C^A(a)$. We call such predicates *complete A-generable* predicates; three examples are

$$(Ex)T_1^A(a, a, x), \quad (Ex)T_0^A(a, x), \quad (Ex)T_1^A((a)_0, (a)_1, x)$$

(cf. IM pp. 281, 291, 343).⁸

Now suppose A is recursive in B , and let $(Ex)R^A(a, x)$ be any A -generable predicate, and $C^B(a)$ be any complete B -generable predicate. Then $R^A(a, x)$, being recursive in A , is recursive in B ; so by the completeness property of $C^B(a)$, $(Ex)R^A(a, x)$ is recursive in $C^B(a)$. In particular, any complete A -generable predicate $C^A(a)$ is recursive in $C^B(a)$.

For the case A_1 and A_2 are of the same degree, by applying this remark first with $A = A_1$ and $B = A_2$ and then vice versa, we conclude that $C^{A_1}(a)$ and $C^{A_2}(a)$ are of the same degree. Thus all complete A -generable predicates $C^A(a)$ for a given A are of one degree (although they do not constitute this degree⁹); and this degree is determined by the degree a of A . We write this degree a' . Then also by the remark,

$$(10) \quad a \leq b \rightarrow a' \leq b'.$$

Writing $A(a) \equiv (Ex)(A(a) \& x = x)$ and using the completeness property of $C^A(a)$, $a \leq a'$. But $C^A(a)$ is not recursive in $A(a)$ (e.g. with $C^A(a) \equiv (Ex)T_1^A(a, a, x)$, by IM Theorem V* pp. 283, 292), so

$$(11) \quad a < a'.$$

By (5), (10) and (6),

$$(12) \quad a' \cup b' \leq (a \cup b)'.$$

Sometimes we shall write A' for some set etc. of degree a' , where a is the degree of A .

While the operation $a \cup b$ is characterizable intrinsically from the abstract partially ordered system of the degrees as the l.u.b. of a and b , the operation a' may so far as we know merely be superimposed upon this ordering.

We know nothing of the converse of (10).⁸ Even with $a' = b'$, for all we know

⁸ Any one of these can play the role here of what Post in [16] called the "complete set" in the case of A recursive, and in [17] the "complete A -canonical set", though the construction is strictly analogous only in the case of the third example (which was used by Kleene in the case of A recursive in lectures in 1941).

⁹ $\bar{C}^A(a)$, though recursive in $C^A(a)$, is not A -generable, by IM Theorem V* pp. 283, 292. The definition of a' could of course have been given equally well from the predicates $\bar{C}^A(a)$, which are complete for the class of the predicates $(x)R^A(a, x)$. The subject should not be left without mentioning that Mostowski [13] has an elegant development and notation.

any of the four possibilities $a < b$, $a = b$, $a > b$, $a \mid b$ may hold for different a 's and b 's. Of course for $a' < b'$, (10) tells us that we cannot have $a \geq b$.

In (12) when a and b are comparable the equality holds trivially; e.g. for $a \leq b$, both sides reduce by (7) and (10) to b' . We can only conjecture that when $a \mid b$ the inequality holds.³

Any degree b for which there is an a with $a' = b$ may be called a *complete* degree. A question which is left here almost untouched is the distribution of the complete degrees among all the degrees. By (4) and (10), there is a least complete degree $0'$. We have not established the existence of incomparable complete degrees, nor do we know except for $a = 0$ whether there is a complete degree among the infinite number of degrees between any a and a' established below. From $a < b < a'$ we are able to conclude about b' only that $a' \leq b' \leq a''$. By (11) each degree a determines an infinite sequence a, a', a'', \dots of degrees, where our ignorance on the above questions allows us only to say that if two such infinite sequences have a common element they are identical from there on.

References to the literature concerning the particular scale $0, 0', 0'', \dots$ with its extension into the transfinite were given in the concluding paragraph of the introduction. We shall have some remarks to make on the degree $0^{(\omega)}$ (or in general, $a^{(\omega)}$) in §4 below.

2. Construction of n independent degrees between a and a' ¹⁰

2.1. THEOREM 1. Given any sets A_1, \dots, A_m ($m \geq 0$), for any $n \geq 1$ sets B_1, \dots, B_n can be found such that: (A) For $k = 1, \dots, n$, B_k is recursive in $(A_1 \cup \dots \cup A_m)'$. (B) For $k = 1, \dots, n$, B_k is not recursive in $A_1, \dots, A_m, B_1, \dots, B_{k-1}, B_{k+1}, \dots, B_n$. (C) For each j, k_1, \dots, k_p ($p \geq 0$) for which A_j is not recursive in A_{k_1}, \dots, A_{k_p} , A_j is not recursive in $A_{k_1}, \dots, A_{k_p}, B_1, \dots, B_n$.

In brief, (B) and (C) assert that all non-recursivities should hold among the $m + n$ sets $A_1, \dots, A_m, B_1, \dots, B_n$ which are not prohibited by recursivities already holding among the m given sets A_1, \dots, A_m .

PROOF. To simplify the notation, we take an example with $m = n = 2$, A_2 not recursive in A_1 , and A_1 recursive in A_2 but not recursive. Then (since A_1 is recursive in A_2) the set $(A_1 \cup A_2)'$ in (A) can be written simply A_2' ; the two conditions (B) can be written simply

$$\{0\} \quad B_1 \text{ is not recursive in } A_2, B_2,$$

$$\{1\} \quad B_2 \text{ is not recursive in } A_2, B_1;$$

and for the set of conditions (C) it suffices to take

$$\{2\} \quad A_2 \text{ is not recursive in } A_1, B_1, B_2,$$

$$\{3\} \quad A_1 \text{ is not recursive in } B_1, B_2.$$

¹⁰ The results of this section are included in those of §3, but we prefer to exhibit the construction in this simpler finite case first.

The method of the illustration will be general.¹¹

By the normal form theorem (IM Theorem IX* p. 292), each of the conditions $\{c\}$ ($c = 0, 1, 2, 3$) is equivalent to an infinity of conditions $\{c, 0\}$, $\{c, 1\}$, $\{c, 2\}$, \dots , where $\{c, e\}$ consists in the recursive relationship in question not holding with Gödel number e .

Let $\alpha_1, \alpha_2, \beta_1, \beta_2$ be the representing functions of A_1, A_2, B_1, B_2 , respectively.

We shall define β_1, β_2 by stages, so that at stage g ($g = 0, 1, 2, \dots$) exactly the first $\nu(g)$ values of both functions will have been chosen, where ν will be a function such that

$$(13) \quad \nu(g) < \nu(g + 1).$$

Instead of saying that the first $\nu(g)$ values of β_k ($k = 1, 2$) have been chosen, we can say equivalently that the number $\kappa_k(g) = \beta_k(\nu(g))$ (IM p. 231) has been chosen. Then

$$(14) \quad \beta_k(a) = (\kappa_k(g))_a \quad \text{if } a < \nu(g),$$

whence, since by (13) $a < \nu(a + 1)$,

$$(15) \quad \beta_k(a) = (\kappa_k(a + 1))_a.$$

The definitions will be in proper sequence, if we regard κ_k as being defined first, so that each value $\kappa_k(g)$ represents a sequence of $\nu(g)$ numbers, and so that with each increase in g the new sequence is an extension of the old, which permits defining β_k from κ_k by (14), or by (15), so then indeed $\kappa_k(g) = \beta_k(\nu(g))$. However to explicate the construction we shall speak of $\kappa_k(g)$ and $\beta_k(\nu(g))$ interchangeably throughout.

At stage 0 no values of β_k shall have been chosen, so

$$(16) \quad \nu(0) = 0,$$

$$(17) \quad \kappa_k(0) = 1.$$

In the step from stage g to stage $g + 1$, for any g , the $\nu(g + 1) - \nu(g)$ additional values which are chosen for β_1, β_2 will be picked so that, no matter how the definitions of β_1, β_2 are subsequently completed, the $g + 1^{\text{st}}$ of the conditions $\{0, 0\}, \{1, 0\}, \{2, 0\}, \{3, 0\}, \{0, 1\}, \{1, 1\}, \{2, 1\}, \{3, 1\}, \dots, \{0, e\}, \{1, e\}, \{2, e\}, \{3, e\}, \dots$ will be satisfied. Thus when the definitions of β_1, β_2 are completed, all of these conditions will be satisfied, and so $\{0\}, \{1\}, \{2\}, \{3\}$, or (B) and (C), will be.

Careful attention will be given to the form of the operations employed, so that when we are done we shall be able to say that κ_1, κ_2 (and ν) are recursive in α_2' , whence it will follow by (15) that β_1, β_2 are, i.e. that (A) holds.

¹¹ If instead we did not know whether A_2 is recursive in A_1 , the sets B_1 and B_2 defined below would still satisfy (A)-(C), the choices made in Case 2 serving to prevent the recursiveness of A_2 in A_1, B_1, B_2 if A_2 is not recursive in A_1 . Given m, n and the method of forming $(A_1 \cup \dots \cup A_m)'$ from A_1, \dots, A_m , by using all disjoint lists of indices $j, k_1, \dots, k_p \leq m$, one can find fixed systems E_1, \dots, E_n of equations, which for every m sets A_1, \dots, A_m define recursively from $(A_1 \cup \dots \cup A_m)'$ n respective sets B_1, \dots, B_n with the properties (B) and (C).

Now we shall describe by cases how the $\nu(g+1) - \nu(g)$ additional values of β_1, β_2 are to be chosen in proceeding from stage g to stage $g+1$. We can think of the process as "extending", for $k = 1$ and 2 , $\kappa_k(g)$ (which incorporates the first $\nu(g)$ values to be given to β_k) to $\kappa_k(g+1)$ (which incorporates also the next $\nu(g+1) - \nu(g)$ values).

At any stage g (when only $\nu(g)$ of the values to be given finally to β_k have been chosen), it will be convenient to call any function β_k which represents a set and possesses these values as its first $\nu(g)$ values an "extension of $\kappa_k(g)$ ".

CASE 0: $\text{rm}(g, 4) = 0$. Then $g = 4e$ where $e = [g/4]$. We must extend $\kappa_k(g)$ to $\kappa_k(g + 1)$ so that $\{0, e\}$ will hold for all extensions β_1 of $\kappa_1(g + 1)$ and β_2 of $\kappa_2(g + 1)$. That is (IM p. 292), we must render it impossible, for every such pair of extensions, that

$$(a) \quad \beta_1(a) = U(\mu y T_1^{1,1}(\tilde{\alpha}_2(y), \tilde{\beta}_2(y), e, a, y))$$

hold for all a ; i.e. that for every a , the right member of (a) be defined, for which the condition is $(Ey)T_1^{1,1}(\tilde{\alpha}_2(y), \tilde{\beta}_2(y), e, a, y)$, and equal in value to the left member. In fact, we shall render this impossible for $a = v(g)$.

SUBCASE 0.1: for some extension β_2 of $\kappa_2(g)$ and some y ,

$$T_1^{1,1}(\tilde{\alpha}_2(y), \tilde{\beta}_2(y), e, v(g), y)$$

This is equivalent to saying that there exist a y and also a number b ($= \beta_2(y)$) having the four properties $b \neq 0$ (so b will be a value $\beta_2(y)$ of a course-of-values function¹²), $(i)_{i < y}[(b)_i < 2]$ (so β_2 will be the representing function of a set¹³), $(i)_{i < \min(y, \nu(g))}[(b)_i = (\kappa_2(g))_i]$ (so β_2 will be an extension of $\kappa_2(g)$) and $T_1^{1,1}(\tilde{\alpha}_2(y), b, e, \nu(g), y)$. If we put $x = 2^y \cdot 3^b$ (so $y = (x)_0$, $b = (x)_1$), and recall that $e = [g/4]$, the subcase hypothesis becomes in symbols

$$(Ex) \{ (x)_1 \neq 0 \ \& \ (i)_{i < (x)_0} [(x)_{1,i} < 2] \ \& \ (i)_{i < \min((x)_0, \nu(g))} [(x)_{1,i} = (\kappa_2(g))_i] \\ \& T_1^{1,1}(\tilde{\alpha}_2((x)_0), (x)_1, [g/4], \nu(g), (x)_0) \}.$$

This is of the form $(Ex)R_0^{\alpha_2}(g, \nu(g), \kappa_2(g), x)$ where $R_0^{\alpha_2}(g, u, w, x)$ is (primitive) recursive in α_2 . In this subcase, let us put for abbreviation

$$(18) \quad X_0 = \mu x R_0^{\alpha_2}(g, v(g), \kappa_2(g), x),$$

so $(X_0)_0$ and $(X_0)_1$ are a y and b with the above four properties. Now we first extend (if necessary, i.e. if $(X_0)_0 > \nu(g)$) the choice of values for β_2 so that $\beta_2((X_0)_0) = (X_0)_1$, and secondly we choose for $\beta_1(\nu(g))$ a value ≤ 1 different from that given then by the right member of (a) when $a = \nu(g)$, say 1 if that is 0 and 0 otherwise. Then as desired (a) will be false when $a = \nu(g)$ for all β_1, β_2 having the values thus far chosen (including those already chosen at stage g). What this gives as value for $\beta_1(\nu(g))$ is¹⁴

¹² We could omit this, by the proof of the normal form theorem IM p. 290.

¹³ We could omit this here, by taking advantage of the remark in 1.2 that to any function there is a set of the same degree.

¹⁴ For $(X_0)_0$ and $(X_0)_1$ are a y and $\tilde{\beta}_2(y)$ which make $T_1^{1,1}(\tilde{\alpha}_2(y), \tilde{\beta}_2(y), e, \nu(g), y)$ true. But then $(X_0)_0 = \mu y T_1^{1,1}(\tilde{\alpha}_2(y), \tilde{\beta}_2(y), e, \nu(g), y)$, by the definition (18) of X_0 as the least x (or indeed also by the analog for $T_1^{\tilde{\alpha}_2, \tilde{\beta}_2}$ of IM p. 291 (37) for T_1^ψ).

$$(19) \quad \beta_1(\nu(g)) = \overline{\text{sg}}(U((X_0)_0)).$$

After taking

$$(20) \quad \nu(g + 1) = \max(\nu(g) + 1, (X_0)_0)$$

(which ensures (13)), thirdly we choose all values (if any) of $\beta_1(a)$ and $\beta_2(a)$ for $a < \nu(g + 1)$ not already chosen (either at the g^{th} stage or in the further choices just described) to be 0. Then

$$(21) \quad \begin{aligned} \tilde{\beta}_1(\nu(g + 1)) &= \mu t \{ t \neq 0 \ \& (i)_{i < \nu(g)} [(t)_i = (\tilde{\beta}_1(\nu(g)))_i] \\ &\quad \& (t)_{\nu(g)} = \overline{\text{sg}}(U((X_0)_0)) \}, \end{aligned}$$

$$(22) \quad \begin{aligned} \tilde{\beta}_2(\nu(g + 1)) &= \mu t \{ t \neq 0 \ \& (i)_{i < \nu(g)} [(t)_i = (\tilde{\beta}_2(\nu(g)))_i] \\ &\quad \& (i)_{i < (X_0)_0} [(t)_i = (X_0)_{1,i}] \}. \end{aligned}$$

Writing $\tilde{\beta}_k(\nu(w))$ as $\kappa_k(w)$, these (with (18)) give us equations of the form¹⁵

$$(23) \quad \kappa_k(g + 1) = \chi_{k0}^{\alpha_2}(g, \nu(g), \kappa_1(g), \kappa_2(g))$$

where $\chi_{k0}^{\alpha_2}(g, u, v, w)$ is partial recursive in α_2 , being defined exactly when $(Ex)R_0^{\alpha_2}(g, u, w, x)$ (cf. IM pp. 326–330).

SUBCASE 0.2: otherwise. Then the values already chosen at stage g render (a) impossible for $a = \nu(g)$ no matter how β_1, β_2 are completed as extensions of $\kappa_1(g), \kappa_2(g)$. However to ensure (13) we take

$$(24) \quad \nu(g + 1) = \nu(g) + 1;$$

and extend β_1, β_2 by choosing $\beta_1(\nu(g)) = \beta_2(\nu(g)) = 0$, so that

$$(25) \quad \kappa_k(g + 1) = \kappa_k(g).$$

Combining the two subcases, using (23) and (25) with our formulation of the Subcase 0.1 hypothesis,

$$(26) \quad \kappa_k(g + 1) = \begin{cases} \chi_{k0}^{\alpha_2}(g, \nu(g), \kappa_1(g), \kappa_2(g)) & \text{if } (Ex)R_0^{\alpha_2}(g, \nu(g), \kappa_2(g), x), \\ \kappa_k(g) & \text{otherwise.} \end{cases}$$

Applying IM p. 337 Theorem XX (c), and noting that α_2 and $(Ex)R_0^{\alpha_2}(g, u, w, x)$ are both recursive in α_2' (1.4 and IM p. 291 line 20), (26) is of the form

$$(27) \quad \kappa_k(g + 1) = \chi_{k0}^{\alpha_2'}(g, \nu(g), \kappa_1(g), \kappa_2(g))$$

where $\chi_{k0}^{\alpha_2'}(g, u, v, w)$ is partial recursive in α_2' , and hence (since it is completely defined) general recursive in α_2' . Similarly, using (20) and (24) (with (18)),

$$(28) \quad \nu(g + 1) = \chi_{00}^{\alpha_2'}(g, \nu(g), \kappa_2(g))$$

where $\chi_{00}^{\alpha_2'}(g, u, w)$ is (general) recursive in α_2' .

¹⁵ For $k = 2$, the $\kappa_1(g)$ is superfluous.

CASE 1: $\text{rm}(g, 4) = 1$. Similar to Case 0.

CASE 2: $\text{rm}(g, 4) = 2$. Then $g = 4e + 2$ where $e = [g/4]$. We must extend $\kappa_k(g)$ to $\kappa_k(g + 1)$ so that $\{\kappa_1(g), \kappa_2(g)\}$ will hold for all extensions β_1 of $\kappa_1(g + 1)$ and β_2 of $\kappa_2(g + 1)$; i.e. we must render it impossible, for every such pair of extensions, that

$$(b) \quad \alpha_2(a) = U(\mu y T_1^{1,1,1}(\tilde{\alpha}_1(y), \tilde{\beta}_1(y), \tilde{\beta}_2(y), e, a, y))$$

hold for all a .

SUBCASE 2.1: for some a and some extensions β_1, β_2 of $\kappa_1(g), \kappa_2(g)$, the right side of (b) is defined with value opposite to the left $\alpha_2(a)$. This is equivalent to the existence of numbers $y, b_1 (= \tilde{\beta}_1(y)), b_2 (= \tilde{\beta}_2(y))$ and a such that

$b_k \neq 0, \quad (i)_{i < y}[(b_k)_i < 2], \quad (i)_{i < \min(y, \nu(g))}[(b_k)_i = (\kappa_k(g))_i] \quad (k = 1, 2),$
 $T_1^{1,1,1}(\tilde{\alpha}_1(y), b_1, b_2, e, a, y)$ and $U(y) \neq \alpha_2(a)$. Putting $x = 2^y \cdot 3^{b_1} \cdot 5^{b_2} \cdot 7^a$, the subcase hypothesis takes the form $(Ex)R_2^{\alpha_2}(g, \nu(g), \kappa_1(g), \kappa_2(g), x)$ with an $R_2^{\alpha_2}$ (general) recursive in α_2 (since α_1 is recursive in α_2). Put

$$(29) \quad X_2 = \mu x R_2^{\alpha_2}(g, \nu(g), \kappa_1(g), \kappa_2(g), x).$$

We first extend (if necessary) the choice of values of β_1, β_2 so that $\tilde{\beta}_k((X_2)_0) = (X_2)_k$. Then (b) will be false for $a = (X_2)_3$ for all β_1, β_2 having the values thus far chosen. After taking

$$(30) \quad \nu(g + 1) = \max(\nu(g) + 1, (X_2)_0),$$

secondly we choose the further values (if any) of $\beta_k(a)$ for $a < \nu(g + 1)$ to be 0. Then

$$(31) \quad \tilde{\beta}_k(\nu(g + 1)) = \mu t \{t \neq 0 \& (i)_{i < \nu(g)}[(t)_i = (\beta_k(\nu(g)))_i] \\ \& (i)_{i < (X_2)_0}[(t)_i = (X_2)_{k,i}]\},$$

whence

$$(32) \quad \kappa_k(g + 1) = \chi_{k2}^{\alpha_2}(g, \nu(g), \kappa_1(g), \kappa_2(g))$$

where $\chi_{k2}^{\alpha_2}(g, u, v, w)$ is partial recursive in α_2 , being defined exactly when $(Ex)R_2^{\alpha_2}(g, u, v, w, x)$.

SUBCASE 2.2: otherwise. We shall show that in this subcase, for some a , the right side of (b) is undefined for every pair of extensions β_1, β_2 of $\kappa_1(g), \kappa_2(g)$, after showing which the subcase can be treated similarly to Subcase 0.2. Accordingly, suppose (for reductio ad absurdum) that, for every a , the right side of (b) is defined for some extensions β_1, β_2 of $\kappa_1(g), \kappa_2(g)$. That is, for each a , there exist $y, b_1 (= \tilde{\beta}_1(y))$ and $b_2 (= \tilde{\beta}_2(y))$ such that $b_k \neq 0, (i)_{i < y}[(b_k)_i < 2]$ and $(i)_{i < \min(y, \nu(g))}[(b_k)_i = (\kappa_k(g))_i]$ ($k = 1, 2$), and $T_1^{1,1,1}(\tilde{\alpha}_1(y), b_1, b_2, e, a, y)$. Putting $x = 2^y \cdot 3^{b_1} \cdot 5^{b_2}$, all of this can be expressed in the form $(a)(Ex)R^{\alpha_1}(g, \nu(g), \kappa_1(g), \kappa_2(g), a, x)$ where R^{α_1} is (primitive) recursive in α_1 . For a given a , put $X = \mu x R^{\alpha_1}(g, \nu(g), \kappa_1(g), \kappa_2(g), a, x)$. Then $U((X)_0)$ is the value

the right side of (b), for some extensions β_1, β_2 of $\kappa_1(g), \kappa_2(g)$ for which that right side is defined. But since Subcase 2.1 is excluded, the right side of (b) when defined has the value $\alpha_2(a)$. Thus, writing out the X in $U((X)_0)$ in full, for all a

$$\alpha_2(a) = U((\mu x R^{\alpha_1}(g, \nu(g), \kappa_1(g), \kappa_2(g), a, x))_0)$$

for the fixed g under consideration, which makes α_2 recursive in α_1 . This contradicts our hypothesis that A_2 is not recursive in A_1 .

Combining the two subcases,

$$(3) \quad \nu(g + 1) = \chi_0^{\alpha_2'}(g, \nu(g), \kappa_1(g), \kappa_2(g)),$$

$$(4) \quad \kappa_k(g + 1) = \chi_{k2}^{\alpha_2'}(g, \nu(g), \kappa_1(g), \kappa_2(g)),$$

here $\chi_0^{\alpha_2'}, \chi_{k2}^{\alpha_2'}$ are (general) recursive in α_2' .

CASE 3: $\text{rm}(g, 4) = 3$. Similar to Case 2.

PROOF OF THEOREM (CONCLUDED). Combining the four cases,

$$(5) \quad \nu(g + 1) = \chi_0^{\alpha_2'}(g, \nu(g), \kappa_1(g), \kappa_2(g)),$$

$$(6) \quad \kappa_k(g + 1) = \chi_k^{\alpha_2'}(g, \nu(g), \kappa_1(g), \kappa_2(g)),$$

here $\chi_0^{\alpha_2'}, \chi_k^{\alpha_2'}$ are recursive in α_2' . These equations with (16) and (17) define κ_1 and κ_2 simultaneously by recursion on g from $\chi_0^{\alpha_2'}, \chi_1^{\alpha_2'}, \chi_2^{\alpha_2'}$. By first setting up a recursion for the "joint function" $\psi(g) = 2^{\nu(g)} \cdot 3^{\kappa_1(g)} \cdot 5^{\kappa_2(g)}$, and then taking $\nu(g) = (\psi(g))_0$ and $\kappa_k(g) = (\psi(g))_k$, it follows that ν, κ_1 and κ_2 are recursive in $\alpha_2', \chi_1^{\alpha_2'}, \chi_2^{\alpha_2'}$. Since the latter are recursive in α_2' , so are ν, κ_1 and κ_2 , and by (5) so are β_1 and β_2 .

2.2. COROLLARY 1. (a) To each degree $a \neq 0$ and positive integer n , there exist degrees $b_1, \dots, b_n < a'$ such that a, b_1, \dots, b_n are independent. (b) Hence: To each degree $a \neq 0$, there are infinitely many degrees less than a' and incomparable with a .

PROOF. (a) Use of theorem for $m = 1$ and A ($= A_1$) of degree a to construct sets B_1, \dots, B_n of degrees b_1, \dots, b_n . Since $a \neq 0$, A is not recursive in the empty list, so as (C) A is not recursive in B_1, \dots, B_n . By (A), $b_k \leq a'$; but were $b_k = a'$, then it would follow by (11) that $a < b_k$, contradicting (C). So by (B) and (C), a, b_1, \dots, b_n are independent.¹⁶

COROLLARY 2. (a) To each degree a and positive integer n , there exist independent degrees c_1, \dots, c_n such that $a < c_k < a'$ ($k = 1, \dots, n$). (b) Hence: To each degree a , there are infinitely many degrees between a and a' .

PROOF. (a) Without loss of generality, we can take $n \geq 2$. After using the theorem as for Corollary 1 (but without supposing $a \neq 0$), let $c_1 = a \cup b_1, \dots, c_n = a \cup b_n$. To prove $a < c_k$: By (5), $a \leq a \cup b_k = c_k$; but were $a = c_k$, then $b_k \leq a \cup b_k = c_k = a$, contradicting (B). To prove $c_k < a'$: By (11) and (A) with (6), $c_k = a \cup b_k \leq a'$; but were $c_k = a'$, then for a $j \neq k$ (which exists, since $n \geq 2$) $b_j \leq a' = c_k = a \cup b_k$, contradicting (B). To prove

¹⁶ So far as we know, $a \cup b_1 \cup \dots \cup b_n$ might $= a'$.

the independence of c_1, \dots, c_n : Were $c_k \leq c_1 \cup \dots \cup c_{k-1} \cup c_{k+1} \cup \dots \cup c_n$, then $b_k \leq a \cup b_k = c_k \leq c_1 \cup \dots \cup c_{k-1} \cup c_{k+1} \cup \dots \cup c_n = a \cup b_1 \cup \dots \cup b_{k-1} \cup b_{k+1} \cup \dots \cup b_n$ (using $c_j = a \cup b_j$ with (8), (9) and (7a)), contradicting (B).

DISCUSSION. According to the case of Corollary 2 for $a = 0$, there exist infinitely many degrees less than the degree $0'$ of Post's complete set.⁵ But we do not know that any of these degrees are degrees of recursively generable sets; those just now constructed are probably not. So the main question of [16] remains unanswered. Similarly for an arbitrary set A of degree a , it is an open question whether among the infinite number of degrees between a and a' there are any which belong to A -generable sets. Another question which the above method does not answer is whether the degrees are dense, i.e. whether for arbitrary degrees a and b such that $a < b$, there exists a degree c such that $a < c < b$.⁶

COROLLARY 3. *To each degree a and positive integer n , there exist degrees d_1, \dots, d_n such that $a < d_1 < d_2 < \dots < d_n < a'$.*

PROOF. In Corollary 2 (a) for $n + 1$, take $d_1 = c_1, d_2 = c_1 \cup c_2, \dots, d_{n+1} = c_1 \cup \dots \cup c_{n+1}$. Then (cf. end 1.3) $a < d_1 < d_2 < \dots < d_n < d_{n+1} \leq a'$.

COROLLARY 4. *To each three degrees a_1, a_2, a_3 such that $a_1 < a_2 \leq a_3$ and each positive integer n , there exist degrees c_1, \dots, c_n such that $a_1 < c_k < a_3$ ($k = 1, \dots, n$) and c, c_1, \dots, c_n are independent for each degree c such that $a_2 \leq c \leq a_3$.*

PROOF. Again take $n \geq 2$. Let A_1, A_2, A_3 have the degrees a_1, a_2, a_3 , and use the theorem to construct sets B_1, \dots, B_n of degrees b_1, \dots, b_n . Then let $c_1 = a_1 \cup b_1, \dots, c_n = a_1 \cup b_n$. Now (for the independence) were $c \leq c_1 \cup \dots \cup c_n$, then $a_2 \leq c \leq c_1 \cup \dots \cup c_n = a_1 \cup b_1 \cup \dots \cup b_n$, contradicting (C) in view of the hypothesis that $a_1 < a_2$. The other n statements for the independence, and the inequalities $a_1 < c_k < a_3$ are proved by contradicting (B) essentially as for Corollary 2.

DISCUSSION. In particular, to each two non-consecutive degrees in the scale $0, 0', 0'', \dots$, an infinite number of degrees can be found which lie between those two and are incomparable to the other degrees in that scale between those two.

3. Simultaneous construction of \aleph_0 degrees between a and a'

3.1. Theorem 1 Corollary 3 suggests asking whether to any degree a we can have a linearly ordered infinite sequence of degrees between a and a' , and moreover whether we can have such a sequence which is dense. The question will be answered affirmatively in this section by extending Theorem 1 to n infinite.

In 1.3 to any sets etc. A_1, \dots, A_n we found a single set etc. $A_1 \cup \dots \cup A_n$ (call it the "recursive join" of A_1, \dots, A_n) which can replace the finite sequence A_1, \dots, A_n in statements of the form 'B is recursive in A_1, \dots, A_n '.¹⁷ Such

¹⁷ And of the form 'B is recursive in $C_1, \dots, C_m, A_1, \dots, A_n$ '.

statements are not usually employed for infinite sequences A_0, A_1, A_2, \dots ;¹⁸ but instead of making such statements we can employ an "infinite recursive join" $A_0 \cup A_1 \cup A_2 \cup \dots$, which we define now.

If A_0, A_1, A_2, \dots are 1-place predicates $A_0(a), A_1(a), A_2(a), \dots$, we can take as their recursive join $A_0 \cup A_1 \cup A_2 \cup \dots$ simply the 2-place predicate $A(a, k) \equiv A_k(a)$. Another possibility is to contract the variables of the latter (cf. IM p. 291 line 20) to obtain as the recursive join the 1-place predicate $A(a) \equiv A((a)_0, (a)_1) \equiv A_{(a)_1}((a)_0)$. The predicates $A(a, k)$ and $A(a)$ are recursive each in the other, i.e. have the same degree, so when we are interested only in the degree of the infinite join it is immaterial which we use. In this paper it is simpler to use $A(a, k)$ avoiding the extra contraction operation, but if the infinite joins are in turn to be made members of new infinite joins (as e.g. in [11]) it is better to use $A(a)$ so as to work throughout with 1-place predicates. For this reason we refrain at least in this paper from choosing from among the various objects of the same degree in defining the infinite join, just as we did with $P \cup Q$ and P' ; but in fact we shall be using $A(a, k)$.

If the A_k are given as 1-place functions $\alpha_k(a)$, we can define their recursive join likewise as the 2-place function $\alpha(a, k) = \alpha_k(a)$. Via representing functions the preceding case, as well as the case of sets, can be considered to come under this; and n -place functions or predicates for $n > 1$ we can first contract to 1-place functions or predicates.

In working with predicates and functions, Church's λ -notation (IM p. 34) can be used to emphasize which are the independent variables for the questions of recursiveness (the other independent variables being parameters). Thus given a 2-place predicate $A(a, k)$, we can write $A = \lambda a \lambda k A(a, k)$ and $A_k = \lambda a A(a, k)$,¹⁹ so $A = A_0 \cup A_1 \cup A_2 \cup \dots$.

Unlike the finite case, the degree of $A_0 \cup A_1 \cup A_2 \cup \dots$ is not necessarily

¹⁸ Under the usual analogy between 'recursive in' and ' \vdash ', such statements would mean ' B is recursive in some finite sublist of A_0, A_1, A_2, \dots ' (IM pp. 224, 264, 391, 425); but this notion is not what is desired for the present purpose.

Say A_0, A_1, A_2, \dots are 1-place predicates with representing functions $\alpha_0, \alpha_1, \alpha_2, \dots$. If instead we attempt to generalize from the finite case by allowing $E_{g_0}^{\alpha_0} g_1^{\alpha_1} g_2^{\alpha_2} \dots, E$ in place of $E_{g_1}^{\alpha_1} \dots g_n^{\alpha_n}, E$ in the definition of relative recursiveness (cf. IM bottom p. 266 and p. 275, where " ψ ", " l " are written instead of " α ", " n "), we get no more because the equations $E_{g_i}^{\alpha_i}$ are utilizable only for the function symbols g_i which occur in E , which are only finitely many since E is a finite list of equations.

To make all of the equations $E_{g_0}^{\alpha_0} g_1^{\alpha_1} g_2^{\alpha_2} \dots$ utilizable, one proposal would be to replace the function symbols g_i used with one argument by a single function symbol g used with two arguments, the second being the numeral for i . This indeed is what we do in defining the infinite recursive join.

Another proposal would be, after choosing the g_i so that their Gödel numbers are a monotone recursive function of i , to allow E to be an infinite set of equations the Gödel numbers of which are recursively enumerable. With the help of a modification of IM pp. 289–292, this proposal can be seen to be equivalent to the preceding one.

¹⁹ For questions of recursiveness, it is of course immaterial whether we employ the predicate $\lambda a A(a, k)$ or the set $\delta A(a, k)$.

preserved when A_0, A_1, A_2, \dots are changed to objects of the same respective degrees; so we have the infinite join only as an operation on the sequence of say predicates A_0, A_1, A_2, \dots , and not on the set or even the sequence of their degrees. (Hence when A_0, A_1, A_2, \dots are to be taken as the members of an infinite join, an ambiguity between objects of the same degree is material, and so in each case that A_0, A_1, A_2, \dots are formed using the operations $P \cup Q$, P' or $P_0 \cup P_1 \cup P_2 \cup \dots$ the ambiguity must be removed by specifying our particular method of forming the latter.) We postpone the discussion of this to 3.6. However most of what we did in §§1 and 2 can be paralleled now.

Thus we say now that A_0, A_1, A_2, \dots (in the given sequence) are *recursively independent*, if for no k ($k = 0, 1, 2, \dots$) is A_k recursive in $A_0 \cup \dots \cup A_{k-1} \cup A_{k+1} \cup \dots$.

Obviously from the definition of the infinite join, A_k is recursive in $A_0 \cup A_1 \cup A_2 \cup \dots$ ($k = 0, 1, 2, \dots$). Thence by (6), any join of a finite subset of A_0, A_1, A_2, \dots is recursive in the infinite join $A_0 \cup A_1 \cup A_2 \cup \dots$. Hence if A_0, A_1, A_2, \dots are recursively independent, so are any finite subset of them; but the converse is not true, as we shall illustrate in 3.6.

Say $A = \lambda a A(a)$, $B = \lambda a k B(a, k)$, $B_k = \lambda a B(a, k)$. What $(A \cup B_0) \cup (A \cup B_1) \cup (A \cup B_2) \cup \dots$ shall be depends on the method of forming the joins $A \cup B_k$ ($k = 0, 1, 2, \dots$). Say for each k we use the particular method given second in 1.3, which is convenient now since it operates on 1-place predicates to form a 1-place predicate. Then $(A \cup B_0) \cup (A \cup B_1) \cup (A \cup B_2) \cup \dots = D = \lambda a k D(a, k)$ where

$$\begin{cases} D(2a, k) \equiv A(a), \\ D(2a + 1, k) \equiv B(a, k). \end{cases}$$

Clearly A is recursive in D and B is recursive in D , and D is recursive in A, B . So D is of the same degree as $A \cup B$ whatever our method of forming the latter; i.e. $(A \cup B_0) \cup (A \cup B_1) \cup (A \cup B_2) \cup \dots$ and $A \cup (B_0 \cup B_1 \cup B_2 \cup \dots)$ are of the same degree.

3.2. THEOREM 2. Given any 1-place predicates A_1, \dots, A_m ($m \geq 0$), a 2-place predicate $B(a, k)$ can be found such that, putting $B = \lambda a k B(a, k)$, $B_k = \lambda a B(a, k)$, $B^k = \lambda a n B(a, n + \text{sg}((n + 1) \dot{-} k))$: (A) B is recursive in $(A_1 \cup \dots \cup A_m)'$. (B) For $k = 0, 1, 2, \dots$, B_k is not recursive in A_1, \dots, A_m, B^k . (C) For each j, k_1, \dots, k_p ($p \geq 0$) for which A_j is not recursive in A_{k_1}, \dots, A_{k_p} , A_j is not recursive in $A_{k_1}, \dots, A_{k_p}, B$.

To see that this is the generalization of Theorem 1 from finite n to \aleph_0 , note that $B = B_0 \cup B_1 \cup B_2 \cup \dots$ and $B^k = B_0 \cup \dots \cup B_{k-1} \cup B_{k+1} \cup \dots$.

PROOF. We use the same illustration as for Theorem 1. Now for (B) and (C) we have to meet an infinity of conditions,

- | | | |
|--------|--------------------------------------|--------------------------|
| {0, k} | B_k is not recursive in A_2, B^k | $(k = 0, 1, 2, \dots)$, |
| {1} | A_2 is not recursive in A_1, B , | |
| {2} | A_1 is not recursive in B , | |

each of which can be separated into an infinity, consisting in the recursive relationship in question not holding with Gödel number e ($e = 0, 1, 2, \dots$); thus for (B) and (C) it will suffice to meet conditions $\{0, k, e\}$, $\{1, e\}$, $\{2, e\}$ ($k, e = 0, 1, 2, \dots$).

We shall define β by stages, so that at stage g exactly the values $\beta(a, n)$ for $a, n < \nu(g)$ will have been chosen, where ν will be such that

$$(37) \quad \nu(g) < \nu(g + 1).$$

These $[\nu(g)]^2$ values will be incorporated into the number $\kappa(g) = \tilde{\beta}(\nu(g), \nu(g))$ (IM p. 291), so

$$(38) \quad \beta(a, n) = (\kappa(g))_{a,n} \text{ if } a, n < \nu(g),$$

$$(39) \quad \beta(a, n) = (\kappa(\max(a, n) + 1))_{a,n}.$$

Rather than order the conditions $\{0, k, e\}$, $\{1, e\}$, $\{2, e\}$ ($k, e = 0, 1, 2, \dots$) without repetitions, we shall use another device to insure that each will be met eventually; in fact $\{0, k, e\}$ ($\{1, e\}$, $\{2, e\}$) will be met in particular on the basis of the values chosen at stage $g + 1$ for $g = 2^0 \cdot 3^k \cdot 5^e$ ($g = 2^1 \cdot 3^e$, $g = 2^2 \cdot 3^e$). Starting with

$$(40) \quad \nu(0) = 0,$$

$$(41) \quad \kappa(0) = 1,$$

we describe by cases the step from stage g to stage $g + 1$.

CASE 0: $(g)_0 = 0$. We shall extend $\kappa(g)$ to $\kappa(g + 1)$ so that $\{0, k, e\}$ where $k = (g)_1$ and $e = (g)_2$ will be satisfied for all extensions β of $\kappa(g + 1)$. That is, we shall render it impossible, for every such extension, that

$$(c) \quad \beta_k(a) = U(\mu y T_1^{1,2}(\tilde{\alpha}_2(y), \tilde{\beta}^k(y, y), e, a, y))$$

hold for all a . In fact, we shall render this impossible for $a = \nu(g)$.

SUBCASE 0.1: for some extension β of $\kappa(g)$ and some y ,

$$T_1^{1,2}(\tilde{\alpha}_2(y), \tilde{\beta}^k(y, y), e, \nu(g), y).$$

This is equivalent to saying that there exists besides a y also a number $b (= \tilde{\beta}^k(y, y))$ such that $b \neq 0$, $(i)_{i < y}[(b)_i \neq 0]$, $(i)_{i < y}(j)_{j < y}[(b)_{i,j} < 2]$, $(i)_{i < \min(y, \nu(g))}(j)_{j < \min(y, \nu(g) + \text{sg}(\nu(g) + k))}[(b)_{i,j} = (\kappa(g))_{i,j + \text{sg}((j+1)-k)}]$ and $T_1^{1,2}(\tilde{\alpha}_2(y), b, e, \nu(g), y)$. Putting $x = 2^y \cdot 3^b$, this can be written $(Ex)R_0^{\alpha_2}(g, \nu(g), \kappa(g), x)$ with an $R_0^{\alpha_2}$ (primitive) recursive in α_2 . Put

$$(42) \quad X_0 = \mu x R_0^{\alpha_2}(g, \nu(g), \kappa(g), x).$$

Now first we extend (if necessary) the choice of values for β so that $\tilde{\beta}^k((X_0)_0, (X_0)_0) = (X_0)_1$, and secondly we choose for $\beta(\nu(g), k)$ the value $\overline{\text{sg}}(U((X_0)_0))$. After taking²⁰

²⁰ Recall that $k = (g)_1$, and note that for $y = (X_0)_0 > k$ defining $\beta(a, n)$ only for $a, n < y$ would define $\beta^k(a, n)$ only for $a < y$ and $n < y - 1$.

$$(43) \quad \nu(g+1) = \max(\nu(g)+1, (g)_1+1, (X_0)_0+1),$$

thirdly we choose the further values (if any) of $\beta(a, n)$ for $a, n < \nu(g+1)$ to be 0. This makes

$$(44) \quad \begin{aligned} \beta(\nu(g+1), \nu(g+1)) &= \mu t \{ t \neq 0 \text{ & } (i)_{i<\nu(g+1)}[(t)_i \neq 0] \\ &\quad \& (i)_{i<\nu(g)}(j)_{j<\nu(g)}[(t)_{i,j} = (\beta(\nu(g), \nu(g)))_{i,j}] \\ &\quad \& (i)_{i<(X_0)_0}(j)_{j<(X_0)_0}[(t)_{i,j+sg((j+1)-(g)_1)} = (X_0)_{1,i,j}] \\ &\quad \& \& (t)_{\nu(g),(g)_1} = \overline{\text{sg}}(U((X_0)_0)), \end{aligned}$$

whence (using (43) as well as (42))

$$(45) \quad \kappa(g+1) = \chi_0^{\alpha_2}(g, \nu(g), \kappa(g))$$

where $\chi_0^{\alpha_2}(g, u, v)$ is partial recursive in α_2 (being defined exactly when $(Ex)R_0^{\alpha_2}(g, u, v, x)$).

SUBCASE 0.2: otherwise. Similarly to Subcase 0.2 for Theorem 1, we take $\nu(g+1) = \nu(g)+1, \kappa(g+1) = \kappa(g) \cdot p_{\nu(g)}$.

CASE 1: $(g)_0 = 1$. We extend $\kappa(g)$ to $\kappa(g+1)$ so that $\{1, e\}$ where $e = (g)_1$ will be satisfied for all extensions β of $\kappa(g+1)$. The treatment is like Case 2 for Theorem 1, except that instead of two 1-place course-of-values functions $\tilde{\beta}_1$ and $\tilde{\beta}_2$ we have now to deal with one 2-place course-of-values function $\tilde{\beta}$.

CASE 2: $(g)_0 > 1$. Similarly we extend $\kappa(g)$ so that $\{2, e\}$ where $e = (g)_1$ will be satisfied.

3.3. COROLLARIES 1-4 (for \aleph_0) to Theorem 2 can be stated corresponding to Corollaries 1-4 (for finite n) to Theorem 1. The independence in Corollaries 1, 2 and 4 is now of the predicates in sequence rather than of their degrees. The proofs are readily given, paralleling the former proofs, with the help of the remarks at the end of 3.1.

3.4. In this subsection, A shall be a 1-place predicate, B a 2-place predicate $\lambda a k B(a, k)$. Then $B_k = \lambda a B(a, k), B^k = \lambda a n B(a, n + sg(n+1-k)), B_\phi = \lambda a n B(a, \phi(n))$ for any number-theoretic function ϕ , etc. The degrees of A, B, B_k, B^k, B_ϕ are written a, b, b_k, b^k, b_ϕ , respectively.

LEMMA 1. Suppose ϕ and ψ are recursive in A ; and B_k is not recursive in A, B^k ($k = 0, 1, 2, \dots$). Then $a \cup b_\phi \leq a \cup b_\psi$ if and only if the range of ϕ is included in that of ψ .

As a special case for A recursive, the lemma will also hold omitting the references to A and the “ $a \cup$ ”.

PROOF. First, suppose the range of ϕ is included in that of ψ , i.e. $(n)(Em)[\phi(n) = \psi(m)]$. Then $\mu m[\phi(n) = \psi(m)]$ is a function of n recursive in A , and $B(a, \phi(n)) \equiv B(a, \psi(\mu m[\phi(n) = \psi(m)]))$, so $B(a, \phi(n))$ is recursive in $A, B(a, \psi(n))$; i.e. $b_\phi \leq a \cup b_\psi$, whence using (5) and (6), $a \cup b_\phi \leq a \cup b_\psi$. Second, suppose on the other hand the range of ϕ is not included in that of ψ , and let k belong to the range of ϕ but not of ψ . Then (since b_ϕ is the degree of an infinite join having B_k as a member, end 3.1) $b_k \leq b_\phi$. Also

$$B(a, \psi(n)) = \begin{cases} B^k(a, \psi(n)) & \text{if } \psi(n) < k, \\ (B^k(a, \psi(n)) + 1) & \text{otherwise, i.e. if } \psi(n) > k, \end{cases}$$

so $B(a, \psi(n))$ is recursive in A , $B^k(a, n)$ (since ψ is recursive in A); i.e. $b_\psi \leq a \cup b^k$, whence $a \cup b_\psi \leq a \cup b^k$. Now were $a \cup b_\psi \leq a \cup b_\phi$, then $b_k \leq b_\phi \leq a \cup b_\phi \leq a \cup b_\psi \leq a \cup b^k$, contradicting the second hypothesis.

LEMMA 2. Suppose ϕ_1 and ϕ_2 are recursive in A ; B_k is not recursive in A , B^k ($k = 0, 1, 2, \dots$); and B is recursive in A' . Then if the range of ϕ_1 is properly included in that of ϕ_2 , which in turn is properly included in the set of all the natural numbers,

$$a < a \cup b_{\phi_1} < a \cup b_{\phi_2} < a \cup b \leq a'.$$

PROOF. By Lemma 1, $a \cup b_{\phi_1} < a \cup b_{\phi_2} < a \cup b$. By (5), $a \leq a \cup b_{\phi_1}$; but were $a = a \cup b_{\phi_1}$, then $b_{\phi_1(0)} \leq b_{\phi_2} \leq a \cup b_{\phi_1} = a \leq a \cup b^{\phi_1(0)}$, contradicting the second hypothesis. By the third hypothesis with (11) and (6), $a \cup b \leq a'$.

3.5. When we say "rational number" or "real number" we can mean one in the open interval $(0, 1)$; but the results, or the treatment itself, adapt to any other open interval e.g. $(0, \infty)$ or $(-\infty, \infty)$.

It is a simple matter to set up an enumeration without repetitions r_0, r_1, r_2, \dots of the rational numbers such that there are three recursive functions μ, ν, ρ with the property that, if $r_i = m_i/n_i$ (m_i, n_i relatively prime), then $m_i = \mu(i)$, $n_i = \nu(i)$ and $i = \rho(m_i, n_i)$.²¹

For any two such enumerations, using the functions μ_1, ν_1, ρ_1 one can pass recursively from the index i_1 of a rational r in the first to its index i_2 in the second, and inversely using μ_2, ν_2, ρ_2 . So in what follows it will be immaterial which particular enumeration is employed.

Now the order relation among the rationals is *recursive*, in the sense that the predicate of natural numbers $i <_r j \equiv r_i < r_j$ is recursive; for $i <_r j \equiv \mu(i)\nu(j) < \nu(i)\mu(j)$.

We call a non-empty class of rationals *recursively enumerable* (*enumerable recursively in A*, or *A- or a-enumerable*), if the indices of the rationals in the class are recursively enumerable (enumerable recursively in A of degree a), i.e. enumerated by a recursive function ϕ (a function ϕ recursive in A).

Here we are interested particularly in the rationals which form the lower half of a Dedekind cut for a real x , where if x is rational the lower half contains no greatest (IM p. 30); call these lower half Dedekind cuts briefly *Dedekind cuts*. We call a real number *lower recursively generable* (*lower a-generable*), if its Dedekind cut is recursively enumerable (a-enumerable).

In particular, each rational is lower recursively generable, a fortiori lower a-generable for any degree a , since with fixed j the class of i 's such that $i <_r j$ is recursive, a fortiori recursively enumerable (IM p. 307).

²¹ If the rationals are those in $(-\infty, \infty)$, we can use four functions μ, ν, σ, ρ , corresponding to the representation $r_i = (-1)^{s_i} m_i/n_i$ with minimal m_i, n_i, s_i .

COROLLARY 5. To each degree a , there exists a system of degrees between a and a' which are ordered among themselves in the same order as the lower a -generable real numbers (including the rational numbers).

PROOF. Apply Theorem 2 with $m = 1$, $A (= A_1)$ of the degree a . To each lower a -generable real x , let the function ϕ_x of degree $\leq a$ enumerate the (indices of the) rationals in its Dedekind cut, and let $c_x = a \cup b_{\phi_x}$ be the correlated degree. By Lemma 2 these degrees c_x have the desired order properties.²²

By making the application of Theorem 2 as for Corollary 4, we obtain a COROLLARY 6 giving a system of degrees ordered as the lower a_1 -generable reals, lying between a_1 and a'_1 , and each incomparable to every degree c such that $a_2 \leq c \leq a_3$.

A quite differently ordered system of degrees between a and a' is derivable similarly from the Church-Kleene theory of constructive ordinals ([4], [3], [7], [19], [9], [12]). In the system S_3 of notation for ordinal numbers (Kleene [7] p. 155, [9] p. 51, or [12] 20), the notations assigned to ordinal numbers form a class O of natural numbers partially ordered under a relation $<_o$. The set $\hat{y}(y \leq_o x)$ for a given member x of O is linearly ordered by $<_o$ and is recursively enumerable (cf. [9] p. 53 (28), (29) and (32); or [12] 21). Both in the definition of O and $<_o$, and in the proof of this enumerability, 'recursive' can be replaced by 'recursive in A '; let the O and $<_o$ thus relativized be written O^A and $<_o^A$ (cf. [12] 30).

COROLLARY 7. To each predicate A of degree a , there exists a system of degrees between a and a' which are ordered among themselves in the partial order in which the members of O^A are ordered by $<_o^A$.

PROOF. Let ϕ_x of degree $\leq a$ enumerate $\hat{y}(y \leq_o^A x)$, and let $c_x = a \cup b_{\phi_x}$.

A COROLLARY 8 gives a system of degrees, ordered as O^{A_1} under $<_o^{A_1}$, between a_1 and a'_1 , and each incomparable to every c such that $a_2 \leq c \leq a_3$.

²² Since the set of the lower a -generable real numbers has the cardinal number \aleph_0 , contains no greatest or least, and is dense, by a theorem of Cantor [2] p. 504 it has the same order type as the set of the rationals. So on its face the corollary says no more than if it merely stated that there is a system of degrees between a and a' ordered as the rationals. However in terms of the underlying construction we do have somewhat more. The degrees which the construction correlates to the rationals are correlated effectively (see below). Then furthermore the degrees correlated to the other lower a -generable reals can be inserted in their proper order among those. The whole system of the degrees thus obtained including those correlated to the rationals is not given effectively, since there is no effective way to give a class of ϕ_x 's of degrees $\leq a$, one to each lower a -generable real x . But for any sequence x_0, x_1, x_2, \dots of such reals such that some $\phi_{x_i}(n) = \phi(n, i)$ where ϕ is of degree $\leq a$, the degree corresponding to x_i is possessed by the predicate $A \cup \lambda anB(a, \phi(n, i))$, and hence (end 3.1) by $\lambda anD(a, n, i)$ for a predicate $\lambda aniD(a, n, i)$ of degree $\leq a'$, and whenever $x_i \leq x$; then $\lambda anD(a, n, i)$ is recursive in $\lambda anD(a, n, j)$ with Gödel number $\eta(i, j)$ where η is a (primitive) recursive function constructible on the basis of the first part of the proof of Lemma 1 (IM Theorem XVIII p. 330 and XXIII p. 342). In particular, in the case x_0, x_1, x_2, \dots are the rationals in the presupposed enumeration such a $\phi(n, i)$ can be given which is recursive.

3.6. The degree of $A_0 \cup A_1 \cup A_2 \cup \dots$ is not determined by the degrees a_0, a_1, a_2, \dots of its members A_0, A_1, A_2, \dots .

As the first illustration of this, let \mathbf{a} be any degree, let $A(a)$ be a 1-place predicate of this degree, and let $A(a, k) = A(k)$. Then for each k , $A_k = \lambda a A(a, k)$ is recursive, i.e. of degree 0; but $A = \lambda a \lambda k A(a, k) = A_0 \cup A_1 \cup A_2 \cup \dots$ is of degree \mathbf{a} . Thus a 1-place predicate of arbitrary degree can be considered as a 2-place predicate so that it becomes an infinite join of 1-place predicates each of degree 0.²³

Except if the set $\partial A(a)$ or its complement $\partial \bar{A}(a)$ is finite (only for $\mathbf{a} = 0$), a suitable non-recursive permutation of this join will change it to one $A_{k_0} \cup A_{k_1} \cup A_{k_2} \cup \dots$ of any other degree b .

Given an infinite join $B_0 \cup B_1 \cup B_2 \cup \dots$ of degree b , we can change it to one of any greater degree \mathbf{a} without changing the degrees of its members, thus. For the recursive A_k of the first illustration, change B_k to $A_k \cup B_k$, where all these joins are to be formed by the second method of 1.3 (cf. end 3.1). Then $(A_0 \cup B_0) \cup (A_1 \cup B_1) \cup (A_2 \cup B_2) \cup \dots$ is of degree \mathbf{a} .

To illustrate that the recursive independence of each finite subset of A_0, A_1, A_2, \dots does not imply that of the whole sequence, start with a recursively independent sequence B_0, B_1, B_2, \dots , produced by Theorem 2, every finite subset of which are then recursively independent (end 3.1). Let $A_k = \lambda a 2 \mid k$, and form $A_k \cup B_k$ by the second method of 1.3. Then $A_k \cup B_k$ is a predicate $\lambda a C_k(a)$, such that $\hat{k}C_k(0)$ are exactly the even numbers, and which is of the same degree as B_k , so that any finite subset of $A_0 \cup B_0, A_1 \cup B_1, A_2 \cup B_2, \dots$ are recursively independent, which must remain the case after any permutation. By a suitable non-recursive permutation $0, k_1, k_2, \dots$ of the natural numbers leaving 0 fixed, we can make $C_{k_{n+1}}(0) = B_0(a)$, so then $A_0 \cup B_0$, being recursive in B_0 , will be recursive in $\lambda a n C_{k_{n+1}}(a)$, i.e. in $(A_{k_1} \cup B_{k_1}) \cup (A_{k_2} \cup B_{k_2}) \cup (A_{k_3} \cup B_{k_3}) \cup \dots$. Thus $A_0 \cup B_0, A_{k_1} \cup B_{k_1}, A_{k_2} \cup B_{k_2}, \dots$ will not be recursively independent, though every finite subset of them are.

When will A_0, A_1, A_2, \dots and B_0, B_1, B_2, \dots of the same respective degrees have recursive joins of the same degree?

To say that A_k is recursive in B_k uniformly in k means that A_k is recursive in B_k with a Gödel number e independent of k (IM pp. 275, 292).

Let us say A_k is *recursive in B_k recursively in k* , if A_k is recursive in B_k with a Gödel number $\varepsilon(k)$ a recursive function of k .

To say simply that A_k is recursive in B_k for each k means the same except that it is not required that respective Gödel numbers $\varepsilon(k)$ can be chosen so that ε is recursive.

Of the three relationships just enumerated,²⁴ the first obviously implies the second, but not conversely (e.g. when $A_1 \neq A_2$, but $B_1 = B_2$); and the second

²³ By writing instead $A(a, k) = A(a)$, it becomes an infinite join of predicates each of degree \mathbf{a} .

²⁴ Incidentally, each of the three relationships is reflexive and transitive. The transitivity of the second is easily established by methods illustrated in 4.1.

obviously implies the third, but not conversely (e.g. when $A_k(a) \equiv (Ex)T_1(k, k, x)$ and $B_k(a) \equiv a = a$, in view of the next remark²⁵).

A sufficient condition that $A = A_0 \cup A_1 \cup A_2 \cup \dots$ be recursive in $B = B_0 \cup B_1 \cup B_2 \cup \dots$ is that A_k be recursive in B_k recursively in k . For then, letting α and β be the representing functions of A and B ,

$$(d) \quad \alpha(a, k) = U(\mu y T_1^1(\beta(y; k), \varepsilon(k), a, y)).$$

Thus a sufficient condition that $A_0 \cup A_1 \cup A_2 \cup \dots$ and $B_0 \cup B_1 \cup B_2 \cup \dots$ have the same degree is that A_k be recursive in B_k recursively in k and vice versa.

These sufficient conditions are not necessary. Of course we may have A recursive in B without having A_k recursive in B_k for each k (e.g. $A(a, k) \equiv (Ex)T_1(a, a, x)$, $B(a, k) \equiv (Ex)T_1(k, k, x)$). But even with A recursive in B and also A_k recursive in B_k for each k , A_k may not be recursive in B_k recursively in k . For example, let

$$\begin{cases} A(a, 2k) \equiv (Ex)T_1(k, k, x), & B(a, 2k) \equiv a = a, \\ A(a, 2k + 1) \equiv a = a, & B(a, 2k + 1) \equiv (Ex)T_1(k, k, x). \end{cases}$$

Then A is recursive in $\lambda a (Ex)T_1(a, a, x)$, which is recursive in B ; and vice versa. Also A_k and B_k are recursive, a fortiori recursive in each other, for each k . But A_k is not recursive in B_k recursively in k , for (d) would give $\alpha(0, 2k) = U(\mu y T_1^1(\beta(y; 2k), \varepsilon(2k), 0, y)) = U(\mu y T_1^1(1, \varepsilon(2k), 0, y))$, which is impossible for ε recursive since $\lambda k \alpha(0, 2k)$ is the representing function of the non-recursive predicate $\lambda k (Ex)T_1(k, k, x)$; and vice versa.

In the finite case $a \cup b$ was characterized as the l.u.b. of the degrees a and b (1.3). Now the degree of $A_0 \cup A_1 \cup A_2 \cup \dots$ is an upper bound for the degrees a_0, a_1, a_2, \dots of A_0, A_1, A_2, \dots (end 3.1), but not necessarily a l.u.b. (beginning this subsection).

Does each countably infinite set of degrees bounded above possess a l.u.b., or in other words is the upper semi-lattice ω -complete?²⁶ No. For consider the degrees, obtained in Theorem 2 Corollary 5 for any given a , which have the order type of the rational numbers; call them "rational degrees". Since each of them is $< a'$, if a Dedekind cut (3.5) in them has a l.u.b., that l.u.b. is $\leq a'$. If two distinct Dedekind cuts in them both have l.u.b.'s, those l.u.b.'s are distinct, since the set-difference of the two cuts is a set of rational degrees which separate the two l.u.b.'s.²⁷ With 2^{\aleph_0} cuts, corresponding to the reals in the interval in question, but only \aleph_0 degrees $\leq a'$ (end 1.2), there are not enough degrees to supply a l.u.b. to each cut.

²⁵ We are using $(Ex)T_1(a, a, x)$ as a familiar example of a non-recursive predicate; cf. 1.4 or IM p. 283.

²⁶ That it is not complete (in the sense analogous to [1] p. 16 for lattices) follows from end 1.2 or (11) in 1.4.

Let us say A_k is recursive in B recursively in k , if A_k is recursive in B with a Gödel number $\varepsilon(k)$ a recursive function of k . This however is equivalent to saying that $A = A_0 \cup A_1 \cup A_2 \cup \dots$ is recursive in B . For if A_k is recursive in B recursively in k , and e.g. B is a 2-place predicate, then letting α and β be the representing functions of A and B , respectively,

$$(e) \quad \alpha(a, k) = U(\mu y T_1^2(\beta(y, y), \varepsilon(k), a, y));$$

and if A is recursive in B , then A_k is recursive in B recursively in k with Gödel number $S_1^{1,2}(e, k)$ where e is any Gödel number of $\lambda kaA(a, k)$ from B (IM p. 342). Thus the upper semi-lattice is “recursively ω -complete” in the sense that there is a least degree, that of $A_0 \cup A_1 \cup A_2 \cup \dots$, among the degrees of sets etc. B in which A_k is recursive recursively in k .

4. Construction of degrees between all a, a', a'', \dots and $a^{(\omega)}$

4.1. Let $A(a)$ be a 1-place predicate of degree a . We begin by introducing the degree $a^{(\omega)}$. Let $L_0^A(a)$, $L_1^A(a)$, $L_2^A(a)$, \dots be the predicates of the increasing degrees a, a', a'', \dots given by 1.4, when in particular we choose A itself from among the predicates of degree a , and at each successive step choose $(Ex)T_1^P(a, a, x)$ from among the complete P -generable predicates; i.e. let

$$(46) \quad \begin{cases} L_0^A(a) \equiv A(a), \\ L_{k+1}^A(a) \equiv (Ex)T_1^{L_k^A}(a, a, x). \end{cases}$$

It will require proof that the degree of the join $L^A = \lambda akL_k^A(a) = L_0^A \cup L_1^A \cup L_2^A \cup \dots$ depends only on the degree a of A .

As the first step in this direction, we evaluate the Gödel number for the proof of (10) in 1.4. Say A is recursive in B with Gödel number e . Write α and β for the representing functions of A and B , respectively. Now by IM pp. 291, 292, 231,

$$T_1^A(a, a, x) \equiv T_1^1(\tilde{\alpha}(x), a, a, x) \equiv T_1^1(\prod_{i < \omega} p_i \exp U(\mu y T_1^1(\tilde{\beta}(y), e, i, y)), a, a, x).$$

Let f be a uniform²⁸ Gödel number from β of

$$\lambda w a \bar{z} \mu x T_1^1(\prod_{i < \omega} p_i \exp U(\mu y T_1^1(\tilde{\beta}(y), w, i, y)), a, a, x);$$

by IM p. 342, $S_2^{1,1}(f, e)$ is then one from β of $\lambda a \bar{z} \mu x T_1^A(a, a, x)$, and by p. 343, $(Ex)T_1^A(a, a, x) \equiv (Ex)T_1^B(S_1^{1,1}(S_2^{1,1}(f, e), a), S_1^{1,1}(S_2^{1,1}(f, e), a), x)$. Let g be a uniform Gödel number of $\lambda u a P(S_1^{1,1}(S_2^{1,1}(f, u), a))$ from an arbitrary predicate P . Then $S_1^{1,1}(g, e)$ is a Gödel number of $(Ex)T_1^A(a, a, x)$ from $(Ex)T_1^B(a, a, x)$, for any 1-place predicates A and B and number e such that e defines A recursively from B .

Now secondly, suppose A is recursive in B with Gödel number e_0 . By induction on k (using in the induction step the previous result for L_k^A , L_k^B , $\varepsilon(k)$ as the A, B, e), L_k^A is recursive in L_k^B with Gödel number $\varepsilon(k)$ where

²⁸ I.e. a Gödel number independent of β for β arbitrary (cf. IM pp. 275, 292).

$$\begin{cases} \varepsilon(0) = e_0, \\ \varepsilon(k+1) = S_1^{1,1}(g, \varepsilon(k)). \end{cases}$$

Then ε is recursive, so L_k^A is recursive in L_k^B recursively in k ; and hence by 3.6, L^A is recursive in L^B . This is for any A and B with A recursive in B .

Applying this second result first with $A = A_1$ and $B = A_2$, and then vice versa, for any A_1 and A_2 of degree a , we conclude that the degree of L^A depends only on the degree a of A . Call this degree $a^{(\omega)}$. Then also, by that result,

$$(47) \quad a \leq b \rightarrow a^{(\omega)} \leq b^{(\omega)}.$$

Of course by (11) and end of either 3.1 or 3.6,

$$(48) \quad a < a' < a'' < \dots < a^{(\omega)}.$$

We doubt that $a^{(\omega)}$ is a complete degree.²⁹

LEMMA 3.³⁰ *If $\lambda a\psi(a, 0)$ is recursive in A , and $\lambda a k\psi(a, k + 1)$ is recursive uniformly in $\lambda a\psi(a, k)$ and 1-quantifier predicates with scope recursive uniformly in $\lambda a\psi(a, k)$, then $\lambda a\psi(a, k)$ is recursive in L_k^A recursively in k .*

²⁹ We chose to define $a^{(\omega)}$ as the degree of $L^A(a, k)$ in keeping with our emphasis in this paper on the operation $'$. There are other predicates which might also quite naturally have been chosen for the role. One is the predicate $M^A(a, k)$ which is defined by an induction on k of especially simple form in IM pp. 287, 292. Let $M_k^A = \lambda a M^A(a, k)$. Another is $N^A(a, k)$ where $N_k^A = \lambda a N^A(a, k)$ is the k -quantifier predicate in the sequence of predicates $A(a)$, $(Ex)T_1^A(a, a, x)$, $(x)(Ey)T_2^A(a, a, x, y)$, $(Ex)(y)(Ez)T_3^A(a, a, x, y, z)$, ... which after the first were used in the proof of IM Theorem V* Part II (b) pp. 283, 284, 292 as examples of predicates not expressible in the respective dual forms.

These two predicates $M^A(a, k)$ and $N^A(a, k)$ have the same degree $a^{(\omega)}$ as $L^A(a, k)$; in fact (which implies this, by 3.6 including Footnote 24), N_k^A is recursive in M_k^A recursively in k , likewise M_k^A in L_k^A , and likewise L_k^A in N_k^A :

M_k^A IN M_k^A . Using the property of M^A IM p. 287 just following (23) (for the starred version p. 292), $N^A(a, 0) = A(a)$ and $N^A(a, k + 1) = M^A(2^{k+1} \cdot 3^{a+1}, k + 1)$, from which the desired result will follow easily if we can show $A(a)$ recursive in $M^A(a, 0)$. But by p. 287 (23), $T_1^A(a, a, y)$ is recursive in $M^A(a, 0)$. To show $A(a)$ recursive in $T_1^A(a, a, y)$, writing α for the representing function of A , we have for any Gödel number e of $\tilde{\alpha}(a)$ from α , $\tilde{\alpha}(a) = U(\mu y T_1^A(\tilde{\alpha}(y), e, a, y))$. Now let $\gamma(a) = 2^{15} \cdot 3^{N^A(a)} \cdot 5^{N^A(a)}$, which is the Gödel number of the equation $0^{(a)} = 0^{(a)}$ in the formalism of recursive functions (IM pp. 263, 276, 258); and let e_1 be the Gödel number of any particular system E_1 of equations which defines $\tilde{\alpha}(a)$ recursively from α . Then $\gamma(a)*e_1$ is $> a$ and defines $\tilde{\alpha}(a)$ recursively from α . Hence $\alpha(a) = (\tilde{\alpha}(\gamma(a)*e_1))_a = (U(\mu y T_1^A(\tilde{\alpha}(y), \gamma(a)*e_1, \gamma(a)*e_1, y)))_a = (U(\mu y T_1^A(\gamma(a)*e_1, \gamma(a)*e_1, y)))_a$.

M_k^A IN L_k^A . By Lemma 3.

L_k^A IN N_k^A . We see quickly by use of a theorem of Post (IM Theorem XI* pp. 293, 295) with IM p. 343 Example 2 that L_k^A is recursive in N_k^A for each k , i.e. that for each k there exists a Gödel number $\varepsilon(k)$ of L_k^A from N_k^A . It remains only to be shown that these numbers $\varepsilon(k)$ can be chosen so that ε is recursive. Since the proof of the existence for each k of an $\varepsilon(k)$ is constructive, Church's thesis allows us to anticipate the general recursiveness of ε . In fact there is a primitive recursive ε , as will be shown in [11] 9.10.

³⁰ The definition in 3.6 of recursive recursiveness, and the remarks there, apply to functions as well as to predicates.

The present lemma will also hold replacing in the second hypothesis "in $\lambda a\psi(a, k)$ " by "in $A, \lambda a\psi(a, k)$ ", as will follow by using $2^{\alpha(a)} \cdot 3^{\psi(a,k)}$ as the $\psi(a, k)$ of the original lemma. The lemma is generalized further in [11] §7.

In the second hypothesis, by 1-quantifier predicates with scope recursive uniformly in $\lambda a\psi(a, k)$ we mean predicates each of the form

$$(Ex)R^{\lambda a\psi(a, k)}(a_1, \dots, a_n, x) \text{ or } (x)R^{\lambda a\psi(a, k)}(a_1, \dots, a_n, x)$$

where R^θ is general recursive uniformly in θ .³¹ Say these 1-quantifier predicates are Q_1, \dots, Q_l with respective representing functions v_1, \dots, v_l ; the second hypothesis means that there is a partial recursive functional $F(\sigma, \tau_1, \dots, \tau_l; a, k)$ (IM p. 326, and for the notation p. 234) such that, for each a and k ,

$$\psi(a, k + 1) = F(\lambda a\psi(a, k), v_1, \dots, v_l; a, k).$$

PROOF. We can first normalize the second hypothesis. First $\psi(a, k) = \mu t(Ex)[\psi(a, k) = t]$, so putting $Q_0 = \lambda at(Ex)[\psi(a, k) = t]$, $\lambda a\psi(a, k)$ is recursive in Q_0 . Now for those among Q_0, \dots, Q_l which are not 1-place predicates we can substitute 1-place predicates of like forms in which they are respectively recursive (IM p. 291, line 20), and then those with a generality quantifier can likewise be changed to predicates with an existential quantifier instead (IM p. 228 *D), after which the $l + 1$ resulting predicates are all recursive in the one predicate $\lambda a(Ex)T_1^{\lambda a\psi(a, k)}(a, a, x)$ (1.4 or IM p. 343). After this has all been done by the indicated methods, we shall have $\lambda a\psi(a, k + 1)$ recursive uniformly in $\lambda a(Ex)T_1^{\lambda a\psi(a, k)}(a, a, x)$.

Now for an arbitrary predicate $P = \lambda aP(a)$ and natural number e , consider the partial recursive function $\lambda a\{e\}^P(a)$ defined recursively from P by e (for the notation cf. IM p. 341). The predicate $\lambda ea x T_1^{\lambda a\{e\}^P(a)}(a, a, x)$ is only partial recursive uniformly in P ; nevertheless the method of IM p. 337 Example 4 (or indeed of p. 343) produces from it a predicate R^P primitive recursive uniformly in P such that (i) $(Ex)R^P(e, a, x) \equiv (Ex)T_1^{\lambda a\{e\}^P(a)}(a, a, x)$ whenever (for given P, e) $\lambda a\{e\}^P(a)$ is completely defined. Now on the one hand, $\lambda ea(Ex)R^P(e, a, x)$ is primitive recursive uniformly in $\lambda a(Ex)T_1^P(a, a, x)$ (IM pp. 291, 343); on the other, we can introduce e as parameter into the recursive scheme defining $\lambda a\psi(a, k + 1)$ uniformly from $\lambda a(Ex)T_1^{\lambda a\psi(a, k)}(a, a, x)$ (by the lemma on uniform recursiveness IM p. 344), so as to define uniformly recursively from $\lambda ea(Ex)R^P(e, a, x)$, and thence from $\lambda a(Ex)T_1^P(a, a, x)$, a function $\lambda eka\theta(e, k, a)$ such that (ii) if (for given P, k, e_k) $(Ex)R^P(e_k, a, x) \equiv (Ex)T_1^{\lambda a\psi(a, k)}(a, a, x)$, then $\theta(e_k, k, a) = \psi(a, k + 1)$. Say f is a uniform Gödel number of $\lambda eka\theta(e, k, a)$ from $\lambda a(Ex)T_1^P(a, a, x)$; then (iii) $S_1^{2,1}(f, e_k, k)$ is a Gödel number of $\lambda a\theta(e_k, k, a)$ from $\lambda a(Ex)T_1^P(a, a, x)$.

Now let e_0 be a Gödel number of $\lambda a\psi(a, 0)$ from A (using the first hypothesis of the lemma), and put

$$\begin{cases} \varepsilon(0) = e_0, \\ \varepsilon(k + 1) = S_1^{2,1}(f, \varepsilon(k), k). \end{cases}$$

Then $\varepsilon(0)$ defines $\lambda a\psi(a, 0)$ recursively from L_0^A . Assume $\varepsilon(k)$ defines $\lambda a\psi(a, k)$ recursively from L_k^A , and take P in (i)-(iii) to be L_k^A . Then by (i),

³¹ $\lambda a\psi(a, k)$ is to be completely defined. This being the case, allowing R^θ to be partial recursive uniformly in θ would give no more generality, by IM p. 337 with T_{n+1}^1 instead of T_2 .

$(Ex)R^P(\varepsilon(k), a, x) \equiv (Ex)T_1^{\lambda a \psi(a, k)}(a, a, x)$, so by (ii) and (iii) $\varepsilon(k + 1)$ defines $\lambda a \psi(a, k + 1)$ recursively from $\lambda a(Ex)T_1^{L_k^A}(a, a, x)$, i.e. from L_{k+1}^A . Thus by induction on k , $\lambda a \psi(a, k)$ is recursive in L_k^A with Gödel number $\varepsilon(k)$, where ε is recursive.

4.2. The upper semi-lattice of the degrees of recursive unsolvability would be a lattice, if each pair of degrees b_1 and b_2 had a g.l.b., i.e. a degree d such that $d \leq b_1$, $d \leq b_2$ and (c) [$c \leq b_1 \& c \leq b_2 \rightarrow c \leq d$]. We shall now construct a pair of sets B_1 and B_2 the degrees b_1 and b_2 of which possess no g.l.b.

THEOREM 3. Given any set A , sets B_1 and B_2 can be found such that: (A) B_1 and B_2 are recursive in L^A . (B) For $j = 0, 1, 2, \dots$, L_j^A is recursive in B_1 and in B_2 . (C) To each set D which is recursive in B_1 and in B_2 , there is a set C which is recursive in B_1 and in B_2 but not in D .

PROOF. By the normal form theorem, (C) can be separated into an infinity of conditions $\{e_1, e_2\}$ ($e_1, e_2 = 0, 1, 2, \dots$), where $\{e_1, e_2\}$ is the condition: If e_1 and e_2 define the same set D recursively from B_1 and from B_2 , respectively, then there is a set C which is recursive in B_1 and in B_2 but not in that set D .

As before (cf. the proofs of Theorems 1 and 2), we shall define β_1 and β_2 by stages, but now we shall sometimes choose an infinite number of values in proceeding from stage g to stage $g + 1$. However, at each stage g ($g = 0, 1, 2, \dots$) all of the first $\nu(g)$ values will have been chosen, where ν will be such that

$$(49) \quad \nu(g) < \nu(g + 1).$$

We write $\kappa_k(a, g)$ ($k = 1, 2$) for the function which, for a given g , has the values which have been chosen for $\beta_k(a)$ at stage g (each of these values being 0 or 1 according as a is to be in B_k or not), and otherwise (i.e. for each a for which the value of $\beta_k(a)$ has not yet been chosen at stage g) the value 2. Thus as g increases, we "extend" $\lambda a \kappa_k(a, g)$ in the sense of changing some of its values from 2 to 1 or 0 while leaving the values which are already 1 or 0 unchanged. Then

$$(50) \quad \beta_k(a) = \kappa_k(a, g) \text{ if } a < \nu(g),$$

$$(51) \quad \beta_k(a) = \kappa_k(a, a + 1).$$

At stage 0 no values of β_k have been chosen, so

$$(52) \quad \nu(0) = 0,$$

$$(53) \quad \kappa_k(a, 0) = 2.$$

Our choice of new values of β_1 and β_2 in passing from stage g to stage $g + 1$ when $g = 2^{e_1} \cdot 3^{e_2}$ will assist in the demonstration (to be left to the end of the proof) that $\{e_1, e_2\}$ will be satisfied for the completed β_1 and β_2 .

In choosing values for β_k in passing from stage g to stage $g + 1$ we shall observe the following RESTRICTION: New values of $\beta_k(a)$ for $a \geq \nu(g + 1)$ will be chosen only when g is of the form $2^{e_1} \cdot 3^{e_2}$, and then only for arguments of the form $2^{e_1} \cdot 3^{e_2} \cdot 5^b$.

Now we are ready to indicate by two cases how the new values of β_1 and β_2 are to be chosen in passing from stage g to stage $g + 1$, i.e. how $\lambda\alpha_k(a, g)$ is to be "extended" to $\lambda\alpha_k(a, g + 1)$. In the process, at any stage g , any function β_k for which $(a)[\kappa_k(a, g) < 2 \rightarrow \beta_k(a) = \kappa_k(a, g)]$ may be called an "extension of $\lambda\alpha_k(a, g)$ ".

CASE 1: $g = 2^{(a)_0} \cdot 3^{(a)_1}$. Then put $e_1 = (g)_0$, $e_2 = (g)_1$.

SUBCASE 1.1: for some a it is possible to extend β_1 from $\lambda\alpha_k(a, g)$ and β_2 from $\lambda\alpha_k(a, g)$ so that e_1 gives a value for a as argument from β_1 and e_2 from β_2 and those two values are different, i.e. in symbols so that

$$(f) \quad U(\mu y_1 T_1^1(\tilde{\beta}_1(y_1), e_1, a, y_1)) \neq U(\mu y_2 T_1^1(\tilde{\beta}_2(y_2), e_2, a, y_2)).$$

In this subcase we shall adopt such an extension. The subcase hypothesis is equivalent to saying that there exist $a, y_1, y_2, b_1 (= \tilde{\beta}_1(y_1))$ and $b_2 (= \tilde{\beta}_2(y_2))$ such that $b_k \neq 0$, $(i)_{i < y_k}[(b_k)_i < 2]$, $(i)_{i < y_k}[\kappa_k(i, g) < 2 \rightarrow (b_k)_i = \kappa_k(i, g)]$ and $T_1^1(b_k, e_k, a, y_k)$ ($k = 1, 2$), and $U(y_1) \neq U(y_2)$. Putting $x = 2^a \cdot 3^{y_1} \cdot 5^{y_2} \cdot 7^{b_1} \cdot 11^{b_2}$, this takes the form $(Ex)R^{\lambda\alpha_k(a, g), \lambda\alpha_k(a, g)}(g, x)$ where $R^{\lambda\alpha_k(a, g), \lambda\alpha_k(a, g)}$ is (primitive) recursive uniformly in $\lambda\alpha_k(a, g)$, $\lambda\alpha_k(a, g)$. Put

$$(54) \quad X = \mu x R^{\lambda\alpha_k(a, g), \lambda\alpha_k(a, g)}(g, x).$$

First, to guarantee that β_k will be one of the extensions described in the subcase hypothesis, we extend (if necessary) the choice of values for β_k so that $\tilde{\beta}_k((X)_k) = (X)_{k+2}$. After taking

$$(55) \quad \nu(g + 1) = \max(\nu(g) + 1, (X)_1, (X)_2),$$

secondly we choose all the remaining values (if any) of $\beta_k(a)$ for $a < \nu(g + 1)$ to be 0. We shall also choose some values for $a \geq \nu(g + 1)$, which will help in establishing (B) (and (C)). By the restriction, at stage g values of $\beta_1(a)$ and $\beta_2(a)$ for $a \geq \nu(g)$ will have been chosen only for a 's of the form $2^{f_1} \cdot 3^{f_2} \cdot 5^b$ where $2^{f_1} \cdot 3^{f_2} < g$. Accordingly, since $2^{e_1} \cdot 3^{e_2} \cdot 5^{\nu(g+1)+x} \geq \nu(g + 1)$, we are free thirdly to put $\beta_k(2^{e_1} \cdot 3^{e_2} \cdot 5^{\nu(g+1)+x}) = \lambda^A(x, g)$ where λ^A is the representing function of L^A . These three acts of extending the choice of values for β_k are accomplished by taking

$$(56) \quad \begin{aligned} \kappa_k(a, g + 1) &= \mu t \{ [\kappa_k(a, g) < 2 \rightarrow t = \kappa_k(a, g)] \\ &\& [a < (X)_k \rightarrow t = (X)_{k+2,a}] \\ &\& [a = 2^{(a)_0} \cdot 3^{(a)_1} \cdot 5^{(a)_2} \& (a)_2 \geq \nu(g + 1) \rightarrow t = \lambda^A((a)_2 - \nu(g + 1), g)] \\ &\& [a \geq \nu(g + 1) \& (a = 2^{(a)_0} \cdot 3^{(a)_1} \cdot 5^{(a)_2} \vee (a)_2 < \nu(g + 1)) \rightarrow t = \kappa_k(a, g)] \}. \end{aligned}$$

SUBCASE 1.2: otherwise. In this subcase, we take $\nu(g + 1) = \nu(g) + 1$, and carry out only the second and third of the acts, so in (56)

$$\text{"}\& [a < (X)_k \rightarrow t = (X)_{k+2,a}]\text{"}$$

is omitted.

CASE 2: $g \neq 2^{(g)_0} \cdot 3^{(g)_1}$. We take $\nu(g+1) = \nu(g) + 1$, and carry out only the second of the acts, so (56) simplifies to

$$\kappa_k(a, g+1) = \mu t \{ \kappa_k(a, g) < 2 \vee a \geq \nu(g+1) \rightarrow t = \kappa_k(a, g) \}.$$

CONCLUSION. Using (52) and (53) for the basis, and combining the cases and subcases for the induction step, the three functions $\nu(g)$, $\kappa_1(a, g)$ and $\kappa_2(a, g)$ are defined from λ^A simultaneously by induction on g , after which $\beta_1(a)$ and $\beta_2(a)$ are defined by (51).

However to establish (A), we shall instead first combine this definition by induction with the definition by induction of $\lambda^A(a, g)$ as obtained by expressing the definition of L^A (cf. (46)) as the definition of its representing function λ^A . Thus the four functions $\nu(g)$, $\kappa_1(a, g)$, $\kappa_2(a, g)$, $\lambda^A(a, g)$ are defined from A by a simultaneous induction on g . We introduce the joint function

$$(57) \quad \psi(a, g) = 2^{\nu(g)} \cdot 3^{\kappa_1(a, g)} \cdot 5^{\kappa_2(a, g)} \cdot 7^{\lambda^A(a, g)},$$

from which ν , κ_1 , κ_2 , λ^A are definable recursively, so (with (51)) it will suffice for (A) to show that ψ is recursive in L^A .

Accordingly, consider the character of the definition of $\psi(a, g)$ by induction on g . First, $\psi(a, 0)$ is recursive in A . Second, when all the cases in the definitions of $\nu(g+1)$, $\kappa_1(a, g+1)$, $\kappa_2(a, g+1)$, $\lambda^A(a, g+1)$ from $\nu(g)$, $\kappa_1(a, g)$, $\kappa_2(a, g)$, $\lambda^A(a, g)$, and thence from $\psi(a, g)$, are taken into account, we find that, to obtain $\psi(a, g+1)$, first we apply uniformly to $\lambda a \psi(a, g)$ recursive operations, then to predicates thus obtained a single quantification (both in the definition of $\lambda^A(a, g+1)$ from $\lambda^A(a, g)$, and also in the subcase hypotheses for Case 1), and then to the resulting predicates and $\lambda a \psi(a, g)$ again recursive operations. So by Lemma 3, $\lambda a \psi(a, g)$ is recursive in L_g^A recursively in g ; and hence (3.6, but for a function ψ instead of a predicate A) ψ is recursive in L^A .

To establish (B), given any j , let $g_j = \mu g [g \geq j \wedge g = 2^{(g)_0} \cdot 3^{(g)_1}]$. Now L_j^A is recursive in $L_{g_j}^A$ (by (1), (2) and (11), as by 4.1, for every j , L_j^A is of degree $a^{(j)}$ where a is the degree of A); and by the treatment of Case 1,

$$L_{g_j}^A(a) = L^A(a, g_j) = \beta_k(2^{(g_j)_0} \cdot 3^{(g_j)_1} \cdot 5^{\nu(g_j+1)+a}) = 0 \quad (k = 1, 2).$$

This shows that L_j^A is recursive in B_1 and in B_2 , since (for the given j) g_j and $\nu(g_j+1)$ are particular numbers.

To establish (C), suppose D with representing function δ is recursive in B_1 , say with Gödel number e_1 , and in B_2 , with Gödel number e_2 . Now for $g = 2^{e_1} \cdot 3^{e_2}$ it was Case 1 which applied, and moreover Subcase 1.2, since the extensions β_1 of $\lambda a \kappa_1(a, g)$ and β_2 of $\lambda a \kappa_2(a, g)$ which were actually chosen in carrying the definitions of β_1 and β_2 through all the subsequent stages have made e_1 and e_2 give for every a the same number $\delta(a)$ as value from β_1 and from β_2 , respectively; had Subcase 1.1 applied for $g = 2^{e_1} \cdot 3^{e_2}$ the treatment there would have prevented this.

But now we can evaluate $\delta(a)$ as follows. Consider any a . Starting with only the values of β_1 incorporated into $\lambda a \kappa_1(a, g)$, we know that there is an extension β_1 of $\lambda a \kappa_1(a, g)$ which makes e_1 give from β_1 a value $U(\mu y T_1^1(\beta_1(y), e_1, a, y))$ for

a as argument, since the value $\delta(a)$ is given by e_1 from the β_1 actually chosen. Moreover if any extension β_1 of $\lambda a \kappa_1(a, g)$ makes e_1 give a value for a as argument, it can only be this same value $\delta(a)$, because e_2 gives this value for a as argument from the β_2 actually chosen, so if it were a different value, at stage g we would have been in Subcase 1.1 instead.

So, for any given a , we seek any such extension β_1 of $\lambda a \kappa_1(a, g)$, which amounts to finding a y and a b ($= \beta_1(y)$) such that $b \neq 0$, $(i)_{i < y}[(b)_i < 2]$,

$$(i)_{i < y}[\kappa_1(i, g) < 2 \rightarrow (b)_i = \kappa_1(i, g)]$$

and $T_1^A(b, e_1, a, y)$. Putting $x = 2^y \cdot 3^b$, the property of y and b can be expressed in the form $R_1^{\lambda a \kappa_1(a, g)}(a, x)$ where $R_1^{\lambda a \kappa_1(a, g)}$ is recursive in $\lambda a \kappa_1(a, g)$. So putting $X_1 = \mu x R_1^{\lambda a \kappa_1(a, g)}(a, x)$, we shall have $\delta(a) = U((X_1)_0)$. This makes δ recursive in $\lambda a \kappa_1(a, g)$, which in turn is recursive in $\lambda a \psi(a, g)$, which as we saw above from Lemma 3 is recursive in L_g^A . Thus (4.1) δ or D is of degree $\leq a^{(g)}$ for $g = 2^{e_1} \cdot 3^{e_2}$. So L_{g+1}^A , which is of degree $a^{(g+1)}$, is not recursive in D , though by (B) it is recursive in B_1 and in B_2 .

4.3. COROLLARY 1. *To each degree a , there exist degrees b_1 and b_2 possessing no g.l.b., incomparable, $> a^{(j)}$ for every finite j , and $< a^{(\omega)}$.*

PROOF. By (C), the degrees b_1 and b_2 of the sets B_1 and B_2 of the theorem for A of degree a possess no g.l.b. Hence they are incomparable, for were e.g. $b_1 \leq b_2$, then b_1 would be their g.l.b. By (B), $a^{(j)} < b_1, b_2$ for each finite j . By (A), $b_1, b_2 \leq a^{(\omega)}$. But were e.g. $b_1 = a^{(\omega)}$, the incomparability would be contradicted.

DISCUSSION. By end 3.6, L_j^A is thus not recursive in B_1 or in B_2 recursively in j ; our proof for (B) that L_j^A is recursive in B_1 and in B_2 for each j made use of the function ν (which is thus non-recursive).

By the corollary, $a^{(\omega)}$ is not l.u.b. of a, a', a'', \dots . It remains an open question (but doubtful) whether a, a', a'', \dots possess a l.u.b.³

Another question not settled here is whether the upper semi-lattice of the degrees of the arithmetical sets is a lattice.³ Using IM Theorems VII (d) p. 285 and X Corollary (a) p. 292, these are the sets having degrees each \leq some one of $0, 0', 0'', \dots$.

4.4. THEOREM 4. *Given any 1-place predicate A , a 2-place predicate $B(a, k)$ can be found such that, putting $B = \lambda a k B(a, k)$, $B_k = \lambda a B(a, k)$, $B^k = \lambda a n B(a, n + sg((n + 1) \div k))$: (A) B is recursive in L^A . (B) For $j, k = 0, 1, 2, \dots$, L_j^A is recursive in B_k . (C) For $k = 0, 1, 2, \dots$, B_k is not recursive in B^k .*

PROOF. In passing from stage g to stage $g + 1$ we observe the RESTRICTION: New values of $\beta(a, k)$ for $a \geq \nu(g + 1)$ will be chosen only when g is of the form $2^k \cdot 3^e$, and then only for arguments of the form $2^k \cdot 3^e \cdot 5^b$. CASE 1: $g = 2^{(g)_0} \cdot 3^{(g)_1}$. Put $k = (g)_0$, $e = (g)_1$. By the restriction, at stage g the value $\beta(g \cdot 5^{(g)_0}, k)$ will not have been chosen. SUBCASE 1.1: for some extension β of $\lambda a n \kappa(a, n, g)$ and some y , $T_1^A(\beta^k(y, y), e, g \cdot 5^{(g)_0}, y)$. We carry out four acts of extending the choice of values for β , corresponding respectively to the first, second and third acts for Theorem 2 Subcase 0.1, and the third for Theorem 3 Subcase 1.1. SUBCASE 1.2 and CASE 2. Similar to Subcase 1.2 and Case 2 for Theorem 3.

COROLLARY 1. To each degree a , there exists a system of degrees $> a^{(j)}$ for every finite j , $< a^{(\omega)}$, and ordered among themselves in the same order as the lower a -generable real numbers (including the rational numbers).

PROOF. Cf. 3.4, 3.5. To x we now correlate the degree b_{ϕ_x} .

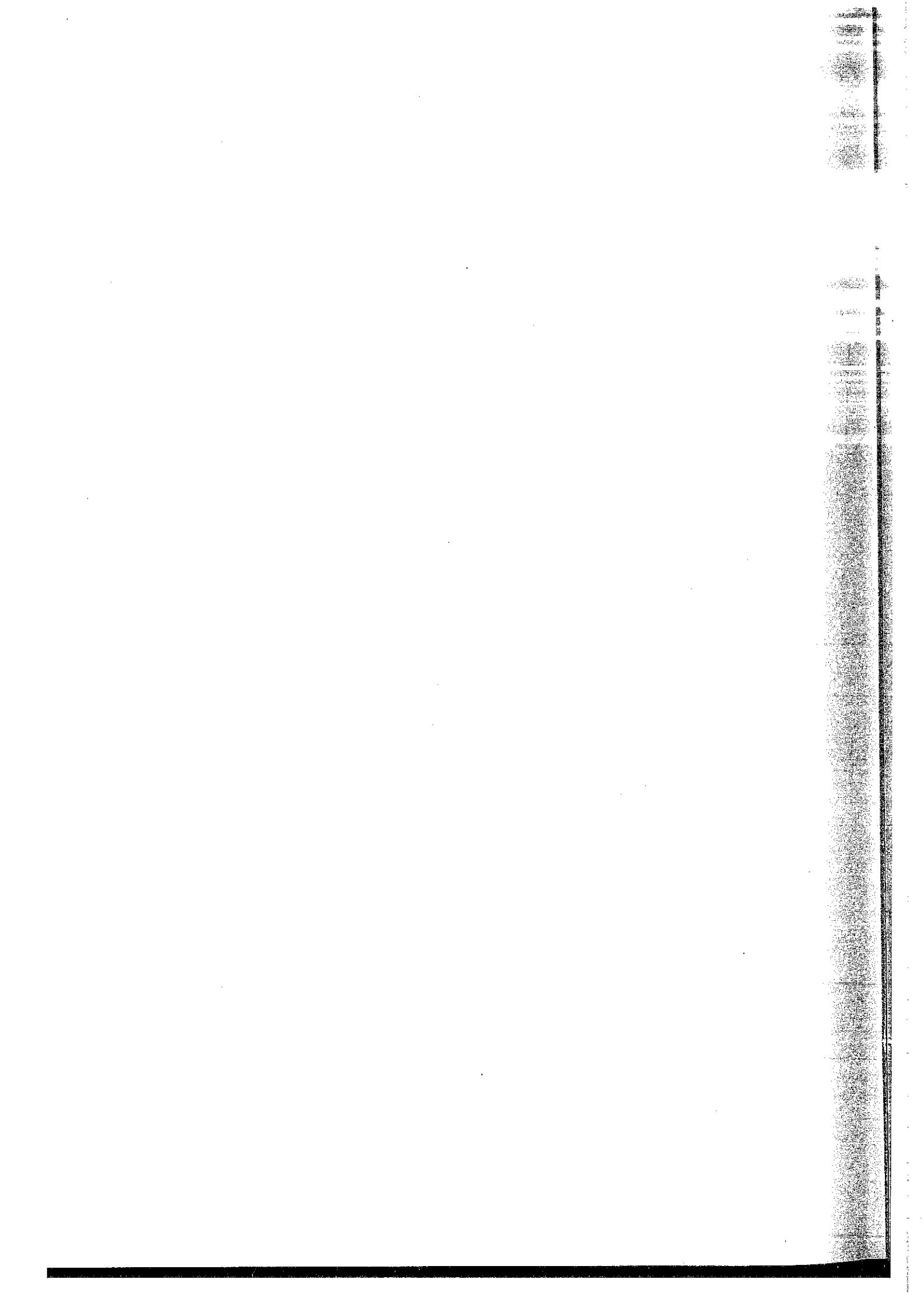
Similarly a COROLLARY 2 corresponds to Theorem 2 Corollary 7.

THE UNIVERSITY OF WISCONSIN
THE CITY COLLEGE, COLLEGE OF THE CITY OF NEW YORK

BIBLIOGRAPHY

- [1] GARRETT BIRKHOFF, Lattice Theory, Amer. Math. Soc. Colloquium Series, vol. 25, revised ed. (1948), xiv + 283 pp.
- [2] GEORG CANTOR, *Beiträge zur Begründung der transfiniten Mengenlehre*, Math. Ann., vol. 46 (1895), pp. 481–512 and vol. 49 (1897), pp. 207–246.
- [3] ALONZO CHURCH, *The constructive second number class*, Bull. Amer. Math. Soc., vol. 44 (1938), pp. 224–232.
- [4] ALONZO CHURCH and S. C. KLEENE, *Formal definitions in the theory of ordinal numbers*, Fund. Math., vol. 28 (1936), pp. 11–21.
- [5] MARTIN DAVIS, *On the theory of recursive unsolvability*, Ph.D. thesis, Princeton University, 1950 (typed).
- [6] S. C. KLEENE, *General recursive functions of natural numbers*, Math. Ann., vol. 112 (1936), pp. 727–742.
- [7] S. C. KLEENE, *On notation for ordinal numbers*, J. Symbolic Logic, vol. 3 (1938), pp. 150–155.
- [8] S. C. KLEENE, *Recursive predicates and quantifiers*, Trans. Amer. Math. Soc., vol. 53 (1943), pp. 41–73.
- [9] S. C. KLEENE, *On the forms of the predicates in the theory of constructive ordinals*, Amer. J. Math., vol. 66 (1944), pp. 41–58. (For a correction, cf. [10] p. 527, [12].)
- [10] S. C. KLEENE, Introduction to Metamathematics, New York (Van Nostrand), Amsterdam (North Holland Pub. Co.) and Groningen (Noordhoff), 1952, x + 550 pp.
- [11] S. C. KLEENE, *Arithmetical predicates and function quantifiers*, in preparation (abstract in J. Symbolic Logic, vol. 18 (1953), p. 190).
- [12] S. C. KLEENE, *On the forms of the predicates in the theory of constructive ordinals (second paper)*, in preparation.
- [13] ANDRZEJ MOSTOWSKI, *On definable sets of positive integers*, Fund. Math., vol. 34 (1947), pp. 81–112.
- [14] ANDRZEJ MOSTOWSKI, *A classification of logical systems*, Studia Philosophica, vol. 4, (1951), pp. 237–274.
- [15] EMIL L. POST, *Formal reductions of the general combinatorial decision problem*, Amer. J. Math., vol. 65 (1943), pp. 197–215.
- [16] EMIL L. POST, *Recursively enumerable sets of positive integers and their decision problems*, Bull. Amer. Math. Soc., vol. 50 (1944), pp. 284–316.
- [17] EMIL L. POST, *Degrees of recursive unsolvability* (Preliminary report), abstract, Bull. Amer. Math. Soc., vol. 54 (1948), pp. 641–642.
- [18] A. M. TURING, *On computable numbers, with an application to the Entscheidungsproblem*, Proc. London Math. Soc., ser. 2, vol. 42 (1936–7), pp. 230–265. A correction, ibid., vol. 43 (1937), pp. 544–546.
- [19] A. M. TURING, *Systems of logic based on ordinals*, Proc. London Math. Soc., ser. 2, vol. 45 (1939), pp. 161–228.

Abstracts



[3] Emil L. Post: *Introduction to a general theory of elementary propositions.*

In this paper Mr. Post studies in its entirety the deductive system which Whitehead and Russell have developed in Part I, Section A, of their Principia Mathematica. Through the concept of the truth table of a truth function, a uniform method is given for telling whether the assertion of a given propositional function of the system can or cannot be derived from the postulates. By means of this result, a number of properties of the system are obtained, among which is the theorem that any propositional function of the system can either be asserted by means of the postulates or else is inconsistent with them.

Two modes of generalizing the system are considered. One consists in generalizing the primitive functions by means of the truth table concept, and connects up with the work of Sheffer and Nicod. The second or postulational method of generalization is shown to introduce new logical systems.

[4] Emil L. Post: *Determination of all closed systems of truth tables.*

Corresponding to each of the 2^n sets of truth values of the arguments of a truth function $f(p_1, p_2, \dots, p_n)$, there is a unique truth value of the function. The relation thus set up may be called the truth table of f . Mr. Post considers the systems of truth tables that can be generated by combining arbitrary primitive truth tables and shows that there are 66 different systems generated by primitive tables with no more than three arguments, and 8 infinite families of systems which require tables of four or more arguments. A formula is given for the tables in each system, and it is shown that they include all closed systems of truth tables. These results are applied to the determination of all the ways in which the logical system of truth functions may be generated by independent primitive functions.

[6] Emil L. Post: *On a simple class of deductive systems.*

In the present paper the author considers a general class of deductive systems involving primitive functions of but one argument, and solves for all such systems the following problem: to find a method for determining in a finite number of steps whether a given enunciation of the system can or cannot be asserted by means of the postulates of the system. The solution is obtained by directly analyzing the way in which assertions are generated from the primitive assertions by the rules of deduction of the system, and gives the beginning of a distinct alternative to the truth-table method which was introduced in its simplest form in a previous paper.

2 Abstracts

[7] Emil L. Post: *Visual intuition in Lobachevsky space.*

According to Klein, the development of non-euclidean geometry has proceeded through three stages, the synthetic stage centering around Lobachevsky, the differential geometry stage initiated by Riemann, and the projective measurement stage developed by Cayley and Klein. The present paper may be said to belong to a fourth stage, already found in the work of Poincaré. It takes an observer brought up in euclidean space, immerses him in Lobachevsky space, and relates what he sees there. In particular, the observer views the Lobachevsky straight line from all positions, and notes that in general it has the appearance of one branch of a hyperbola, always spreading away from him, and varying in size and shape with his distance from the line. He then observes objects at various distances from him, and notes that for a given distance an object appears much smaller in Lobachevsky than in euclidean space. Another interpretation is that Lobachevsky space is much roomier than euclidean space as one goes out from a given fixed center. These two developments of the Lobachevsky intuition are then related by showing that the shortest line between two points as determined by the metric presents the appearances described above.

[8] Emil L. Post: *Visual intuition in spherical and elliptic space: Einstein's finite universe.*

The method of the present paper is almost the same as the one used in the author's paper on Lobachevskian intuition. However, the difficulties involved in the conception of the finiteness of both the spherical and elliptic spaces and the one-sidedness of the latter require further elucidation. This is accomplished by imagining a plentitude of the straight lines of these spaces, uniformly distributed in them, and noting just how they bind the spaces together; or by noting how the spaces can be broken up into a finite number of adjacent but nonoverlapping cells. In both cases the desired intuition is obtained by noting how these straight lines and cells vary with changes in the position of the observer. Since Einstein's finite universe is one of these types, the present paper may serve to dispel some popular misconceptions.

[9] Emil L. Post: *A non-Weierstrassian method of analytic prolongation.*

The present method arose from an attempt to find satisfactory necessary and sufficient conditions that a function of a real variable be an analytic function. It was found that this was so when and only when a certain straight line construction which makes a network of broken lines correspond to a subdivision of a certain portion of the plane into squares gives in the limit a network of curves through whose parameterization a transformation of the plane is defined whose corresponding functions have continuous total differentials. Since at the same time the method gives the

values of the original function over a region of the complex plane, whereas originally it was known only along a segment of the real axis, it appears as a new method of analytic prolongation. By repeated application of the method, the function can be explored throughout its domain of existence. The method is that used by R.G.D. Richardson in 1917. In its particular development it yields a neat formula for the coordinates of an arbitrary point of intersection of the network as a finite series whose terms approach those of a Taylor's expansion.

[10] Emil L. Post: *A new method for generalizing e^x in the complex domain.*

The method in question is simply the one presented in the preceding paper. In the present case, the broken line network in the plane of the dependent variable becomes a set of lines radiating from the origin crossed orthogonally by a spider's web of broken lines. The lengths cut off along these radiating lines are found to be proportional to their distances from the origin, a fact which easily enables us to find that, in the limit, the network of broken lines is replaced by a pencil of lines crossed orthogonally by a series of concentric circles.

[11] Emil L. Post: *A simple geometric proof of the equality of the Brochardt angles of a triangle.*

It is easy to obtain a simple trigonometric proof of the equality of Brochardt angles of a triangle, but as far as the author is aware no simple geometric proof of this fact is extant. The present note supplies a proof.

[12] Emil L. Post: *Theory of generalized differentiation.*

The theory of derivatives of non-integral orders, begun by Liouville and Riemann, may be said to be complete. On the other hand, the theory of more general operators considered as functions of D , the derivative, is fragmentary. In the present paper, the problem is attacked by an entirely new method, and results are obtained that generalize those of Riemann. Although the present investigation was inspired purely by an interest in the bizarre, it has been found to connect with many fields in the most recent branches of analysis. In particular, it has applications in the theory of the Laplace transformation, Volterra's permutable functions and functions of composition (of the closed cycle group), and the Heaviside operational classes.

[13] Emil L. Post: *The m th derivative of a function of a function; calculus of m th derivatives.*

At the October, 1923, meeting of the Society the author presented a paper on a theory of generalized differentiation. The application of this to

4 Abstracts

the Heaviside theory requires a usable knowledge of m th derivatives. The problem of finding the m th derivative of a function of a function, $F(u)$, has always presented great difficulties. Thus, although the Faa di Bruno formula gives $(d^m/dx^m)F(u)$ as a finite series in $F^{(m)}(u)$ with coefficients depending wholly on u and its derivatives, the formula for these coefficients is quite unmanageable. The present paper concerns itself specifically with these coefficients, and sets up formulas for the coefficients of a sum, product, and quotient of two functions, and for a function of a function in terms of the coefficients of these functions. Thus in the last case if v is a function of u and u a function of x ,

$$(m)v/(n)x = \sum_{l=m}^n ((m)v/(l)u)((l)u/(n)x)$$

where $(m)v/(n)x$ is the m th coefficient in the n th derivative of a function of u . These rules in connection with the coefficients for the simplest elementary functions may be said to constitute a calculus of the coefficients in question, and this in turn, added to the previously known formulas, gives a calculus of m th derivatives.

[15] Emil L. Post: *Polyadic groups*. Preliminary report.

In 1928 W. Dörnte generalized the concept of abstract group to allow for an operation involving an arbitrary number of elements m . In the present investigation the concepts of substitution and transformation are also generalized. An identity is seen to be an $(m - 1)$ -ad of elements, and corresponding definitions of inverse and transform result. The author shows that every abstract m -adic group (Dörnte's m -group) can be represented as an m -adic substitution group. There results the following theorem. Every abstract polyadic group can be represented as a coset of an invariant subgroup of an ordinary group (and conversely for finite groups). Further work is largely restricted to finite groups and includes, for abstract groups, a complete generalization of the theory of cyclic groups and of the first and third parts of Sylow's theorem under the condition g/P^B prime to $(m - 1)$. For substitution groups results on odd and even substitutions, sets of intransitivity, and Jordan's theorem for regular groups have been generalized, while for linear groups the elementary theorems, other than for similarity transformations, go over. Still in progress is the determination of the m -adic linear groups in two variables for the case $m = 3$. (Received October 26, 1935).

[16] Emil L. Post: *Finite combinatory process. Formulation 1*.

The present formulation envisages a general problem consisting of a class of specific problems. A symbol space is postulated in which each

specific problem and corresponding answer can be represented, and in which the work leading from symbolized problem to symbolized answer is to be carried out. A fixed set of directions both directs operations in the symbol space and determines the order in which those directions are to be applied. It gives a solution of the general problem if on applying it to each specific problem the process thus set up terminates in the answer to that problem. The present formulation describes a specific symbol space, enumerates the primitive process the problem solver is assumed capable of performing in that symbol space, and gives a definite structure to the set of directions. Its purpose is to present a model for formulations of increasing psychological complexity and a norm for their logical reduction, the whole to yield a theory of the limitations of mathematics along the lines already initiated by Gödel and Church. (Received October 31, 1936.)

[25] Emil L. Post: *Recursive unsolvability of a problem of Thue.*

Thue's problem (Skrifter utgit av Videnskapselskapet i Kristiania 1914. I. Mathematisk-Naturvidenskabelig Klasse, no. 10) may be restated as follows. Given an arbitrary finite set of symbols, with A 's and B 's arbitrarily given strings (zeichenreihen) involving no other symbols than those in the given set, P and Q operational variables, to determine whether B is an assertion in the system with initial assertion A and operations PA_iQ produces PB_iQ , PB_iQ produces PA_iQ , $i = 1, 2, \dots, \mu$. Through the intermediary of the Turing machine, a known recursively unsolvable decision problem is reduced to the decision problem of a system with initial assertion A' and operations PA'_iQ produces PB'_iQ , $i = 1, 2, \dots, \mu'$, having the property that the set of assertions of the system is unchanged when the system is transformed into Thue type by adding the inverse operations PB'_iQ produces PA'_iQ . The recursive unsolvability of the problem of Thue easily follows. (Received September 20, 1946.)

[28] Emil L. Post: *Degrees of recursive unsolvability.* Preliminary report.

The author's canonical sets (Amer. J. Math. vol. 65) are generalized to S -canonical sets by hypothetically adding primitive assertions representing the membership or non-membership of $1, 2, 3, \dots$ in set of positive integers S . Set S_1 of positive integers is proved (Turing) reducible to S (Post, Bull. Amer. Math. Soc. vol. 50) when and only when S_1 and its complement are S -canonical sets. A "complete" S -canonical set S' is set up, and it is proved that each S -canonical set is reducible to S' , while S' is not reducible to S . With S , say, the null set K_0 , a scale of increasing degrees of unsolvability is thus furnished by $K_1, K_2, K_3, \dots, K_{n+1} = K'_n$. The main completed result is a strengthened Kleene Theorem II, Trans. Amer. Math. Soc. vol. 53: each class of sets in Kleene's $(n + 1)$ st column includes a set of higher degree of unsolvability than any set common to both classes. Work is in progress on further equivalence proofs, further applications of the K_n -scale,

6 Abstracts

incomparable degrees of unsolvability related to the K_n -scale, extension of the main theorem to Mostowski, Fund. Math. vol. 34, and extension of the K_n -scale into the constructive transfinite. (Received March 17, 1948.)

[29] Samuel Linial and Emil L. Post: *Recursive unsolvability of the deducibility, Tarski's completeness, and independence of axioms problems of propositional calculus.*

The primitive connectives are $\sim p$ and $(p \vee q)$, corresponding modus ponens and substitution the rules of deduction. Let S be a fixed normal system on a, b with non-null g 's and g' 's whose deducibility problem is recursively unsolvable (see Post, Bull. Amer. Math. Soc. vol. 50 (1944), pp. 286-287, 292; vol. 52 (1946) footnote 3). To a and b are made correspond $\sim\sim(p \vee \sim p)$ and $\sim\sim\sim(p \vee \sim p)$ respectively. With $\sim(\sim p \vee \sim q)$ as connective, correspondents in propositional calculus result for each non-null string B on $a, b, B(p)$ designating a certain particular correspondent. A normal operation is simulated in propositional calculus by (\sim, \vee) implications which have a corresponding effect for certain correspondents of the strings involved, and implications which transform any one correspondent of a string into any other. A finite set $\{A_i\}$ of tautologies result such that if $\{P_i\}$ is any particular complete finite set of axioms (tautologies) for propositional calculus, $c(p)$ the tautology $\sim\sim(\sim p \vee p)$, then B is asserted in S when and only when: $B(p)$, itself a tautology, is deducible from $\{A_i\}$; $\{A_i\}, \{(B(p) \supset T_i)\}$ is a complete set of axioms; $\{A_i\}, c(p), (B(p) \supset c(p))$ is not an independent set of axioms. The title results follow for arbitrary finite sets of tautologies as axioms, hence also the first and third without the tautology restriction. (Received May 24, 1948.)

[30] Emil L. Post: *Note on a relation recursion calculus.*

The primitive assertions and productions of a canonical system (VIII 50(3)) are replaced by a *development* involving a finite number of relations $r_i(x_1, x_2, \dots, x_{l_i})$ between natural numbers with primitive assertions, premises, and conclusions all of this form. The x 's in the former case are specified natural numbers, in the latter cases variables, 'successors of variables' or natural numbers, each variable or successor thereof in a conclusion of a production being present in at least one of these forms in at least one premise thereof. The assertions generated by the development are then all of the form $r_i(n_1, n_2, \dots, n_{l_i}), n_j = 0, 1, 2, \dots$. A relation $r(x_1, x_2, \dots, x_l)$, defined for all natural numbers, is *recursively generable* (recursively enumerable or null) if its true cases are the assertions of that form generated by some development, *recursive* if both it and its negative are recursively generable. A one-valued function $f(x_1, x_2, \dots, x_m)$ of natural numbers is *recursive* if the relation $y = f(x_1, x_2, \dots, x_m)$ is recursively generable (it then being recursive). By use of the Kleene normal form (II 38(1)) it is

proved that every 'general recursive function' is recursive. The procedure is simple, the possibilities, when restrictions are lifted, open to view. Extension to strings as arguments is contemplated.

- [31] Emil L. Post: *Solvability, definability, provability; history of an error.*

The concept general recursive function leads to a thesis (Kleene, Trans. Amer. Math., Soc. (1943) p. 60) having as a consequence the existence of absolutely unsolvable problems. Also in 1943, the author (Amer. J. Math., footnote 6) proposed the problem of absolutely undecidable propositions. The author later realized that a necessary condition for provability would suffice; still later, that to solvability and provability should be added definability. In 1947, in a lost letter to Tarski via Church, the author proposed a formulation of "Primitive Inductive-Reflective Proof" involving, besides the elementary, only mathematical induction and "Gödelization." But representation theory for Gödel's system P (1931) failed to materialize due to P 's Axiom of Reducibility. Indeed, Kleene's Example (Proceedings of the International Congress of Mathematicians, 1950, vol. II, p. 683) offers hope of an impossibility proof. The obvious specificity of Mostowski's "definable set" concept (Fund. Math. (1947)) suggested a like development for the language (set of W. F. F.) of system P , and constructivity being lost, further led to that generalization which constitutes the language of the P^α of the author's present research (see following abstract). Just prior to the last idea it was seen that the correct order of the earlier triplet is, of course, solvability, definability, provability. (Received February 9, 1953.)

- [32] Emil L. Post: *A necessary condition for definability for transfinite von Neumann-Gödel set theory sets, with an application to the problem of the existence of a definable well-ordering of the continuum. Preliminary report.*

In $P^{(\alpha)}$ (see previous abstract), α is an arbitrary ordinal. The language (sic) of $P^{(\alpha)}$ is that of the simple theory of types semantically treated. An " α -formula" is defined as an $x_i^{(\alpha)}(x_j^{(\alpha)})$ range, all sets of type less than α , and any formula obtainable from α -formulas by a negation, disjunction or quantification. The present thesis is that every definable set is given by some α -formula. The well-ordered series of ordinals leads to a well-ordering A of the continuum—represented by a set of positive integers in classic manner. A Gödelian interval $0 \leq \xi < 1$, ξ sequence of definitions proves that A is an " α -set" and, subject to the filling in of five validation lacunae and checking, by a calculus of validation, the other formulas, leads to the theorem (the corresponding metatheorem is obvious): either (the defined) A well-orders the continuum-interval $0 \leq \xi < 1$ or there is no definable well-ordering of the continuum. Further verifications of the thesis separate themselves into theorems: A. That certain (a) generalizations, (b) modifications, of $P^{(\alpha)}$ lead only to α -sets. B. That a certain diagonal set

of a denumerable set under certain (a) general, (b) particular conditions is again an α -set. (Received February 9, 1953.)

[33] S. C. Kleene and Emil L. Post: *The upper semi-lattice of degrees of recursive unsolvability.*

The degrees of recursive unsolvability (Post, Bull. Amer. Math. Soc. Abstract 54-7-269) form an upper semi-lattice under a l.u.b. operation $a \cup b$, but not a lattice. To any degree a (including the degree O of a recursive set), let a' be the degree of the complete A -canonical set (Post, loc. cit.) or equivalently of the predicate $(\exists x)T_1^A(a, a, x)$ (Kleene, *Introduction to metamathematics*, Chapter XI) where A is of degree a ; and let $a^{(\omega)}$ be the degree e.g. of the predicate $M^A(a, k)$ (Kleene, loc. cit.). There exist between a and a' (between all a, a', a'', \dots and $a^{(\omega)}$), infinitely many degrees pairwise incomparable, and also e.g. infinitely many degrees ordered among themselves as the rational numbers. (Received July 17, 1953).

Permissions

Birkhäuser Boston thanks the original publishers of the papers of Emil L Post for granting permission to reprint specific papers in this collection.

- [1] Reprinted from *Annals of Math* 20, © 1918 by Princeton University
- [2] Reprinted from *Am. Math. Monthly* 26, © 1919 by American Mathematical Society
- [3] Reprinted from *Bull. AMS* 26, © 1920 by American Mathematical Society
- [4] Reprinted from *Bull. AMS* 26, © 1920 by American Mathematical Society
- [5] Reprinted from *Am. J. Math.* 43, © 1921 by The Johns Hopkins University Press
- [6] Reprinted from *Bull. AMS* 27, © 1921 by American Mathematical Society
- [7] Reprinted from *Bull. AMS* 29, © 1923 by American Mathematical Society
- [8] Reprinted from *Bull. AMS* 29, © 1923 by American Mathematical Society
- [9] Reprinted from *Bull. AMS* 29, © 1923 by American Mathematical Society
- [10] Reprinted from *Bull. AMS* 29, © 1923 by American Mathematical Society
- [11] Reprinted from *Bull. AMS* 29, © 1923 by American Mathematical Society
- [12] Reprinted from *Bull. AMS* 30, © 1924 by American Mathematical Society
- [13] Reprinted from *Bull. AMS* 30, © 1924 by American Mathematical Society
- [14] Reprinted from *Trans. of the Am. Math. Soc.* 32, © 1930 by the American Mathematical Society
- [15] Reprinted from *Bull. AMS* 41, © 1935 by American Mathematical Society
- [16] Reprinted from *Bull. AMS* 42, © 1936 by American Mathematical Society
- [17] Reprinted from *The Jour. of Symb. Logic* 1, © 1936 by the Association for Symbolic Logic
- [18] Reprinted from *Trans. of the Am. Math. Soc.* 48, © 1940 by the American Mathematical Society

- [19] Reprinted from the *Annals of Math. Studies*, © 1941 by Princeton University Press
- [20] Reprinted from *The Undecidable*, © 1965 by Raven Press
- [21] Reprinted from *Am. J. Math.* **65**, © 1943 by The Johns Hopkins University Press
- [22] Reprinted from *Bull. AMS* **50**, © 1944 by the American Mathematical Society
- [23] Reprinted from *Bull. AMS* **52**, © 1946 by the American Mathematical Society
- [24] Reprinted from the *Jour. of Symb. Logic* **11**, © 1946 by the Association for Symbolic Logic
- [25] Reprinted from *Bull. AMS* **52**, © 1946 by the American Mathematical Society
- [26] Reprinted from the *Jour. of Symb. Logic* **12**, © 1947 by the Association for Symbolic Logic
- [28] Reprinted from *Bull. AMS* **54**, © 1948 by the American Mathematical Society
- [29] Reprinted from *Bull. AMS* **55**, © 1949 by the American Mathematical Society
- [30] Reprinted from the *Jour. of Symb. Logic* **16**, © 1951 by the Association for Symbolic Logic
- [31] Reprinted from *Bull. AMS* **59**, © 1953 by the American Mathematical Society
- [32] Reprinted from *Bull. AMS* **59**, © 1953 by the American Mathematical Society
- [33] Reprinted from *Bull. AMS* **59**, © 1953 by the American Mathematical Society
- [34] Reprinted from *Annals of Math.* **59**, © 1954 by Princeton University Press