Lambda Calculus: Some Models, Some Philosophy

Dana Scott
Oxford University, Oxford, Great Britain

Dedicated to Professor S.C. Kleene on the occasion of his 70th birthday

0. Introduction

In this essay yet another attempt at an exposition of why the λ-calculus has models is made. The λ-calculus was one of the first areas of research of Professor Kleene, an area in which the experience he gained was surely beneficial in his later development of recursive function theory. In what transpires below, the dialogue will be found to involve Professor Curry rather more than Kleene, since the former has written more extensively on the foundational aspects of combinatory logic. Nevertheless the early works of Church, Curry, Kleene, and Rosser were very closely integrated, and the contributions of Kleene were essential. Thus, the topic does not seem inappropriate to the occasion.

Section 1 provides a very short historical summary, and it will be found that there is considerable overlap with CURRY (1979), which is also in this volume. An earlier version of Professor Curry's paper was in any case the incentive to write the present paper, and the reader should consult Curry's contribution for further references and philosophical remarks.

In Section 2 there is a review of the theory of functions and relations as sets leading up to the important notion of a *continuous set mapping*. In Section 3 the problem of the self-application of a function to itself as an argument is discussed from a new angle, and it is shown that—under the reduction of continuous set mappings to multi-relations—a coherent set-theoretical definition is indeed possible. The model (essentially due to PLOTKIN (1972)) of the basic laws of λ-calculus thus results.

Section 4 relates self-application to recursion by the proof of David Park's Theorem to the effect that the least fixed-point operator and the "paradoxical" combinator are the same in a wide class of well-behaved models. The connection thus engendered to recursion theory (r.e. sets) is outlined, and the section concludes with some remarks on recent results about ill-behaved models and on induction principles.

Section 5 returns to the theme of type theory, and a construction of an (η)-model with fewer type distinctions is presented, There is a brief discussion of how to introduce more type distinctions into models *via* equivalence relations, a topic deserving further study. Finally, Section 6 takes up various points of philosophical disagreement with Professor Curry which can be discussed in the light of the construction presented here. There are many questions remaining, some of which are touched upon. An appendix exhibits a moderately strong axiomatic theory suggested by the models that may help the reader see the difference between the originally proposed calculus and the outlook developed by the author.

1. Some historical background

Priority for the invention of the type-free calculus of functions (*pace* Frege) goes to Moses Schönfinkel in a lecture at Göttingen in December 1920. This talk was written up by H. Behmann and published as SCHÖN-FINKEL (1924); a translation with a useful introduction by W.V. Quine is to be found in van HEIJENOORT (1967, pp. 355–366). Sometime around 1926–1927 as a graduate student, Haskell Curry in an analysis of "the process of substitution" independently discovered the combinators; but then in 1927 in "a literature search" he came across the Schönfinkel paper. Full credit is apportioned by him to Schönfinkel in his thesis (CURRY, 1930). According to CURRY (1968), Alonzo Church prepared a manuscript in 1928 on a system with λ-abstraction, and the publication, CHURCH (1932), indicates that it was work done as a National Research Fellow 1928–1929. In this paper on p. 352 and in nearly the same words in CHURCH (1941, pp. 3–4) we find Schönfinkel only mentioned in connection with the reduction of multiple-argument functions to monadic functions. Of course, it is fair to say that Schönfinkel was concerned merely with a kind of *definitional reduction* of primitives, and he proposed no *postulates* from which properties of these general functions could be derived. Such postulates were the contribution of Curry and Church.

Unfortunately Church, to my knowledge, has never explained as fully as Curry has how he was led to his theory. He must have been strongly influenced by Frege (via Russell), and he hoped to solve the para-doxes—not through the theory of types, but by the rejection of the law of the excluded middle. In CHURCH (1932, p. 347), it is stated that such combinations as occur in the Russell Paradox (namely, $(\lambda x \, \mathbf{not} \, (x(x))) \, (\lambda x.$ $\mathbf{not} \, (x(x)))$, which converts to its own negation) simply *fail to have a truth value*. Thus, we do not have here an intuitionistic theory, but a failure of excluded middle because functions are only *partially defined*. Alas, in

KLEENE and ROSSER (1935) it was shown that Church's system (which was employed by Kleene in KLEENE (1934) (written in 1933) and his thesis, KLEENE (1935) (accepted in September, 1933)) is *inconsistent*. The proof was later very much simplified by Curry and can be found in CURRY and FEYS (1958, pp. 258–260). It applies to various systems proposed by Curry, also, but not to his thesis, which is just the "equational" theory of combinators. This is essentially the system of CHURCH (1941), and the "system of symbolic logic" in that monograph is condensed to a very few pages (§21, pp. 68–71). The consistency of these systems is very forcibly demonstrated by the well-known theorem of CHURCH and ROSSER (1936).

The connections between the systems of Curry and Church were spelled out in his thesis by Barkley Rosser in ROSSER (1935) (written in 1933), who emphasized particularly the elimination of variables. This theme was also taken up by Quine in QUINE (1936) and again in QUINE (1971). It seems very strange to me that in his description of the method of Schönfinkel, QUINE (1971) does not mention the problem of consistency. He says (pp. 8–9).

> "Schönfinkel was the first to reduce analysis to algebra. He was the first to analyze the variable, by showing how to translate it contextually into constant terms. But his treatment is less pure than one could wish; it analyzes the variable only in combination with a function theory that is in effect general set theory."

But an inconsistent set theory it is as soon as we try to give axioms of the usual sort! All later attempts have to make strong restrictions to get anything like a workable set theory. None of these systems are, in my own opinion, particularly natural or beautiful; I do not attempt to catalogue them at this point, therefore—though some further comments are given in the last section of the paper. At this stage it seems fair to say that the model theory of these systems is certainly not very well developed. References and less negative discussion can be found in CURRY (1979) and SELDIN (1980).

For the equational systems proposed by Curry and Church the consistency proof *via* the Church–Rosser Theorem is **not** a great comfort to my mind. As with many proof-theoretic arguments, the result is very sensitive to the exact formulation of the rules. Thus, in a note, KLOP (1977) (see also KLOP (1979)), it is shown that if we extend the usual system by, say, surjective pairing functions, then the Church–Rosser Theorem *no longer holds*. Such an extension is so "natural" in the original style of Schönfinkel, that this result looks very unfortunate. Of course, someone may formulate some modified syntactical property of reduction that will imply that not all terms are interconvertible, but I do not think any such argument has yet been published. There is, however, a consistency proof by *models* (cf. SCOTT (1977), which Klop does not seem to know about)

which allows certain extensions: namely, by the properties of anything that exists in the model. And this brings us to the problem of models and their place in the discussion.

Historically my first model for the λ-calculus was discovered in 1969 and details were provided in SCOTT (1972) (written in 1971). What I have called the "graph model" was found by Gorden Plotkin and is given in PLOTKIN (1972). It was rediscovered by me in 1973, and a simplified version—with proper credit to Plotkin—is found with rather full proofs in SCOTT (1976), which also contains many historical remarks and many references. Motivation for these discoveries is, it is hoped, very fully exposed in SCOTT (1977) and SCOTT (1977a). The two kinds of models are not unrelated, but we shall not be able to go into details here. The graph model was in effect introduced *twice* earlier in recursion theory quite explicitly but at the time *not* identified as a model (see the discussion of *enumeration operators* in SCOTT (1976, pp. 575 ff.)). This failure is an interesting case history in the psychology of discovery. On this score we find in CURRY (1979) the following closing passage:

> "The history of combinatory logic shows that progress can result from the interaction of different philosophies. One who, as I do, takes an empirical view of mathematics and logic, in the sense that our intuitions are capable of evolution, and who prefers constructive methods, would never discover the models which Scott proposed. On the other hand, it is doubtful if anyone, with what seems to be Scott's philosophy, would have discovered combinatory logic. Both of these approaches have added to the depth of our understanding, and their interaction has produced more than either would have done."

The last sentence is very kind, but I am afraid I cannot agree with all of what was said before that. In the first place, the models *are* constructive. In the second place, as an intellectual exercise—even though the combinators are some 12 years older than I am (I was born in 1932) and we cannot change the past—I think I can defend "my philosophy" in a way that will show how they and their laws (in the form of the graph model) could have been discovered by the extension of known elementary ideas—ideas certainly easily understandable in 1928 when Curry and Church were at work on their first systems.

As a sidelight on the question of discovery, I only recently noticed this passage in WHITEHEAD and RUSSELL (1910, p.280):

> "Again, let us denote by ' $\times n$' the relation of m to $m \times n$; then if we denote by '*NC*' the class of all cardinal numbers, $\times n$"*NC* will denote all numbers that result from multiplying a cardinal number by n, i.e., all multiples of n. Thus, e.g., $\times 2$"*NC* will be the class of all even numbers."

These remarks were included merely for the sake of illustration of what happens when we take the image of a class under a function; however, if

anyone had been thinking of a general theory of functions, he would have noticed here how a binary function was reduced to one-place functions ($m \times n = \times (n)(m)$). But no one as far as I know *did* think of that rather simple point before Schönfinkel, who was instead motivated by the idea of having something like a Sheffer-stroke operator for predicate calculus. Did he read *Principia*? We cannot hope to know at this distance in time (for a little more information on Schönfinkel's life, see the review KLINE (1951)); and since he did not write his paper himself, we cannot know what references to the literature he would have made.

2. Some thoughts on functions

What functions are there? Often a *representation* helps in seeing in simpler terms why certain functions exist; we do not need to claim, however, that the representation embodies the essence of the function concept—it is just used to establish possibility. For example, power series can show that a function like e^x is well-defined, differentiable, and that it satisfies its differential equation. (Other properties like being one–one onto the positive reals or satisfying $e^{x+y} = e^x \times e^y$, though elementary, are perhaps not instantly recognizable from this definition.) But not all functions have power-series expansions. It may even be the case that the difficulties about convergence led rather directly in the 19th century to a more abstract notion of function and, step by step, to the "logical monsters" deplored by Poincaré.

From particular functions one goes on to functions in general—and to *spaces* of functions. Who first suggested that a function could be regarded as a set of ordered pairs? By 1914 both Wiener and Hausdorff were doing just that (for relations as well as functions and with pairs reduced to sets as well), but I do not find any earlier references in their works or in some other books I consulted. It is not a very important historical question for our discussion, however, since it is certain that by 1914 functions as abstract objects were well understood. I stress this because I wish to argue that even a set-theoretical discovery of λ-calculus would have been not only possible but even well motivated at an early state. That it was not discovered this way is no argument against my thesis, because many things remain undiscovered even long after all the essential ingredients are available.

In the discussion below we employ the common functional notation $y = F(x)$ to indicate that the *point* x is mapped to the *point* y under the function F. For simplicity, we suppose that both x and y belong to the *same* set A. For *subsets* $X, Y \subseteq A$, as also write $Y = F(X)$ to mean that Y is

the *image* of the set X under the mapping F. (See formula (1) below.) In the *Principia* the notation was $F'x$ and $F''X$, respectively, because the image of a set was regarded as the *plural* of the image of a point.

Formally we define set images by:

(1) $F(X) = \{ y \mid \exists x \in X . y = F(x) \}$.

It may seem unrigorous not to distinguish notationally between $F(x)$ and $F(X)$, but since we are not at this point worrying about a *theory* of sets, we simply imagine A as fixed for the time being. Further, without loss of generality, we could stipulate that A and its power set are *disjoint*; hence, there need be no ambiguity between $x \in A$ and $X \subseteq A$.

Eq. (1) is a particular case of the *image under a relation*. Thus, if R is a binary relation between elements of A, we can define:

(2) $R(X) = \{ y \mid \exists x \in X . yRx \}$,

where, as always, the variables have their obvious ranges: $x, y \in A$ and $X \subseteq A$. (This is meant to correspond to *Principia*, Definition *37.10; and it is in any case standard.) If we like, we can think of R as a *multi-valued* function; that is, even when $X = \{x\}$, the image $R(\{x\})$ can be "plural". For single-valued functions, it is obvious that

$$y = F(x) \quad \text{iff} \quad \{y\} = F(\{x\}).$$

Now even for relations R, the correspondence $X \mapsto R(X)$ *is* a function on the power set of A (with values in the same set of sets). We should ask: "What kind of a function is this?" The answer is well-known (though I could not find it in the nearly endless list of formulae in the *Principia*). Such functions are *distributive* in the sense that they distribute over all unions of sets. Since every set is the union of its singleton subsets, this comes down to:

(3) $R(X) = \bigcup \{ R(\{x\}) \mid x \in X \}$.

Any correspondence satisfying (3) is distributive. If a **mapping** on sets satisfies (3), then we can define a **relation** between points by:

(4) $yRx \quad \text{iff} \quad y \in R(\{x\})$.

Then in view of (3), the *defined* meaning of $R(X)$ in (2) is the same as the *given* meaning of $R(X)$ as a set mapping. We can summarize:

Proposition. *There is a one–one correspondence of interdefinability between binary (point) relations and distributive set mappings provided by formulae (2) and (4).*

Even if this simple theorem is not in *Principia*, the authors (as well as Wiener, Jourdain, Hardy, etc.) would have certainly understood it at once.

It is time now for a *new* idea. Let us recall first how relations are reduced to sets of ordered pairs. We use the notation $\langle x,y \rangle$ for pairs and $A^2 = A \times A$ for the cartesian product. A relation is just a subset $R \subseteq A^2$ and yRx means the same as $\langle y,x \rangle \in R$. The question is: "What should we say about *n*-ary relations?" Using the notation $\langle x_0, x_1, \ldots, x_{n-1} \rangle$ for *n*-tuples and A^n for the *n*-fold Cartesian product, we define the set of *all* finite sequences by:

$$(5) \qquad A^* = \bigcup_{n=0}^{\infty} A^n.$$

(The case $n = 0$ has $A^0 = \{\langle \rangle\}$, where $\langle \rangle$ is the *empty tuple*.)

An *n-ary relation* is a subset $S \subseteq A^n$. It seems reasonable to call a subset $P \subseteq A^*$ a *multi-relation*. The new idea has to do with how to generalize formula (2) for the image. This is one such way:

$$(6) \qquad P(X) = \{ y \mid \exists n \exists x_1, \ldots, x_n \in X. \ \ \langle y, x_1, \ldots, x_n \rangle \in P \}.$$

In case $P \subseteq A^2$ (and we agree that the various A^n are pairwise disjoint), then (6) reduces to (2) exactly—provided we substitute P for R in (2). In case $P \subseteq A^{n+1}$, then P is an $(n+1)$-ary relation; formula (6) could be said to define the image of X^n, if we regard P as a multi-valued mapping where $\langle x_1, \ldots, x_n \rangle \mapsto y$. We might wish to define more generally the image of $X_1 \times X_2 \times \cdots \times X_n$, where the n sets are allowed to be distinct, but it is useful enough to stick to the one-place set mapping in (6) for many purposes.

What is achieved by this new definition? In the case of binary relations, we had $y \in R(X)$ just in case there is *one* element $x \in X$ to which y is R-related. In the case of multi-relations, $y \in P(X)$ requires a *finite subset* $\{x_1, \ldots, x_n\} \subseteq X$ with y P-related to $\langle x_1, \ldots, x_n \rangle$. Clearly then the mapping $X \mapsto P(X)$ need *not* be distributive, because we might have $y \in P(\{x_1, \ldots, x_n\})$ but $y \notin P(\{x_i\})$ for $1 \leqslant i \leqslant n$. But nevertheless the mapping does have a special property which generalizes (3):

$$(7) \qquad P(X) = \bigcup \{ P(E) \mid E \subseteq X, E \text{ finite} \}.$$

We call such functions from sets to sets *continuous*, because a "finite approximation" $E^1 \subseteq P(X)$ is already determined as being true by a finite approximation $E \subseteq X$ where $E^1 \subseteq P(E)$. In fact, the biconditional

$$\forall X \forall E^1 \big[E^1 \subseteq P(X) \leftrightarrow \exists E \subseteq X. \ E^1 \subseteq P(E) \big]$$

is equivalent to the validity of (7) for all $X \subseteq A$, provided we restrict the ranges of the variables E^1 and E to *finite* subsets of A. As another characterization of continuous set mappings, we might mention that they are exactly the ones that distribute over *directed* unions of sets (equivalently: over unions of *chains* of sets). But (7) as a definition will be sufficient for our purposes.

Just as with binary relations, if we have a *given* continuous set mapping P, then we can define a multi-relation P by:

(8) $\langle y, x_1, \ldots, x_n \rangle \in P$ iff $y \in P(\{x_1, \ldots, x_n\})$.

As before, in view of (7), the $P(X)$ defined by (6) will be equal to the given $P(X)$.

A small point of difference: every $P \subseteq A^*$ determines a continuous set mapping, and every continuous map is obtained in this way. However, *different* P's may determine the *same* map. (Thus, whether $\langle \rangle \in P$ is of no moment as empty sequences do not figure in (6); for greater regularity and for a later application we will assume that (8) *allows* $\langle \rangle \in P$.) There is not, therefore, a one–one correspondence between continuous set mappings and subsets $P \subseteq A^*$. In order to regain the uniqueness we had with binary relations, we need to require that the subset $P \subseteq A^*$ satisfy a certain "fullness" condition:

(9) $\langle \rangle \in P$ and whenever $\langle y, x_1, \ldots, x_n \rangle \in P$

              and $\{x_1, \ldots, x_n\} \subseteq \{y_1, \ldots, y_k\}$,   then $\langle y, y_1, \ldots, y_k \rangle \in P$.

Condition (9) makes P *maximal* among the subsets of A^* determining the desired continuous set mapping.

There is still the question: "Why are continuous set mappings *interesting*?" The answer is quite pleasant. In TARSKI (1930) (written in 1928), condition (7) is given **exactly** is his Axiom 4 on the abstract properties of the *consequence relation*. Of course, we are *not* assuming here the special properties of a closure operator, where additionally

$$X \subseteq P(X) = P(P(X)),$$

since these are not even true for relational images (unless the relation is special). But closure operators can provide fine examples of continuous mappings. Aside from the set of logical consequences of a set of sentences, other familiar algebraic examples of continuous closure operators would be, say, to let $P(X)$ be the *subgroup* of A generated by X in the case that the set A carries the structure of a group. Then P would satisfy the above extra conditions as a set mapping.

It does not seem at all far fetched to suggest that in 1928 a proposed Master's thesis could have had the topic of investigating the theory of *arbitrary* continuous set mappings. I can see in my mind's eye exactly the kind of paper that would have been submitted to *Fundamenta Mathematicae*. Here is a summary of results and definitions:

   (i)   *A continuous set function is defined as one satisfying formula (7) above; an alternative definition states that the function preserves directed unions of sets.*

(ii)   *Continuity in several variables is defined as continuity in each variable separately.*

(iii)   *As a side remark it is pointed out that* **P***A, the power set of A, has a topology with the sets of the form* $\{X \subseteq A \mid E \subseteq X\}$, *where* $E \subseteq A$ *is finite, as a basis for the open sets; continuity as defined in* (i) *and* (ii) *is proved to be the same as topological continuity on* **P***A and on the product space* $(\mathbf{P}A)^n$. *This remark would not actually be needed in the sequel, but it helps show why the concepts are natural.*

(iv)   *The composition of continuous functions (in any number of variables) is continuous; this of course follows from* (iii), *but a direct argument is very elementary.*

(v)   *The notion of* A^* *and a multi-relation is defined.*

(vi)   *The image of a set under a multi-relation is defined and it is noted that the mapping is continuous.*

(vii)   *It is proved that every continuous function is obtainable from a multi-relation, and in fact there is a maximal such (cf. formula (8)).*

(viii)   *The maximal multi-relations in* (vii) *are proved to be exactly the "full" relations in the sense of formula (9).*

As it stands, up to this point, the paper seems rather thin and even might not have been accepted for publication. In the next section we turn to some additional ideas that could have been added to it to give it somewhat greater depth.

By the way, a simple observation as to why continuity is better than distributivity (the binary-relation case) concerns functions of several variables. Suppose $F(X, Y)$ is distributive separately in X and in Y. Then the composition (with identity functions) resulting in $F(X, X)$ is continuous *but not necessarily distributive*; further compositions like $F(F(X,X), F(X,X))$ remain continuous but are even further from being distributive. The reason is that the condition $z \in F(X, X)$ though equivalent to $\exists x \in X\, \exists y \in X.$ $z \in F(\{x\}, \{y\})$, is not necessarily equivalent to $\exists x \in X.\ z \in F(\{x\}, \{x\})$. By taking the point-wise union of continuous functions we get other continuous functions that are even less distributive. This remark shows that a study of set mappings might have suggested the notion of continuity independently of the well-known examples of "algebraic" closure operations.

3. Some reflections on self-application and combinators

As in Scott (1973) I recall a passage from the introduction to Church (1941):

"In particular it is not excluded that one of the elements of the range of arguments of a function f should be the function f itself. This possibility has been frequently

denied, and indeed, if a function is defined as a correspondence between two previously given ranges, the reason for the denial is clear. Here, however, we regard the operation or rule of correspondence, which constitutes the function, as being first given, and the range of arguments then determined as consisting of the things to which the operation is applicable. This is a departure which is natural in passing from consideration of functions in a special domain to the consideration of functions in general, and it finds support in consistency theorems which will be proved below."

The philosophical stance assumed by Church does not fit so very well to the λ-K-calculus where *every* application has to be taken as meaningful. (An interesting discussion of the differences between the λ-I-calculus and the λ-K-calculus will be found in BARENDREGT (1980).) If we think instead of partial recursive functions, where rules are programs and where programs have code numbers, then, following Kleene, an application $\{e\}(n)$ may or may not be "meaningful" (that is, the Turing machine need not halt). There are many cases, however, where $\{e\}(e)$ *does* converge, and such self-applications are quite meaningful. The drawback (if it is one) is that the method of code numbers takes functions in *intension*. It may well happen that $\{n\} = \{m\}$ as partial recursive functions, but $\{e\}(n) \neq \{e\}(m)$. Many of the calculi considered by Church and Curry take the function concept to be *extensional*, and it is rather more difficult in such a context to be precise about the meaning of "rule of correspondence". It is really not the least bit fair of Church in the above passage to invoke the consistency proof *via* the Church–Rosser Theorem, since this gives no *intuitive* justification of the choice of reduction rules which are given in advance.

There are nevertheless many theories of functions in extension where self-application is no trouble whatsoever. To have such a model $\mathfrak{M}$, just let $\mathfrak{M} = \mathbb{R}$, the real numbers. Define application as *addition*, so that $x(y) = x + y$. This is the model for a "theory" of *translations*. In $\mathfrak{M}$, the axiom that has $\forall z.\ x(z) = y(z)$ always implying $x = y$ is trivially valid. The only difficulty with this model is that so *few* functions from $\mathfrak{M}$ into $\mathfrak{M}$ are represented by elements of $\mathfrak{M}$ ("few", despite the fact that there are a continuum number of distinct translations). For example, even though $x(x)$ is always meaningful, this is a function **not** represented in the model; that is, there does not exist an element $a \in \mathfrak{M}$ with $a(x) = x(x)$ for all $x \in \mathfrak{M}$. But $x(x) = 2x$ is surely a "harmless" function. Why not throw in *all* linear functions $x \mapsto \alpha + \beta x$ and not just the translations $x \mapsto \alpha + x$? The difficulty now is that the linear functions require *two* parameters α and β. There is no "nice" function of the type $[\cdot, \cdot]: \mathfrak{M} \times \mathfrak{M} \to \mathfrak{M}$ where we could then define application so that $[\alpha, \beta](x) = \alpha + \beta x$. The point is that we would want $x(y)$ to be a linear polynomial in each variable separately. But then $x(x)$ is no longer linear but rather is quadratic. (In any case, the project is doomed as soon as $x \mapsto 1 + x$ is allowed; this function has *no* fixed point, but in a full theory of combinators *all* functions have fixed

points.) The rub comes when we try to combine the definability of $x(y)$ along with what Curry calls *combinatory completeness*. This point is perhaps not made clear by Church, and his few words of justification do not seem to me to be sufficient.

Is there, then, a model that combines the definability of application (and self-application) along with a flexible theory of functions? But hold, we are not allowed to ask this question because we have put ourselves back to 1928 (or earlier) in a state ignorant of the work of Curry and Church. It is required that we *discover* the λ-calculus—the combinators *and* their laws—on our own.

So let us consider again the thoughts we had on functions in the last section. We left off after having written a first chapter of a Master's thesis about continuous set functions and their representation by multi-relations $P \subseteq A^*$. We defined there in formula (6) an application $P(X)$ for all $X \subseteq A$. This seems fairly powerful. The idea of set- and function-abstraction was certainly in the air at that time (Curry and Church did not start in a vacuum), thus the idea of a *notation* (say, like that of Russell, or Peano, or Frege) for *abstraction* could have occurred to our Master's candidate. I will not try to construct an "original" notation, but will use Church's:

(1) $\lambda X. \ \tau[X] = \{\langle \rangle\} \cup \{\langle y, x_1, \ldots, x_n\rangle \, | \, y \in \tau[\{x_1, \ldots, x_n\}]\}$

Here, $\tau[X]$ is an expression with possibly a free set variable X and $\tau[\{x_1, \ldots, x_n\}]$ is short for the substitution of the set-term $\{x_1, \ldots, x_n\}$ for the variable X in τ. Bound variables and substitution were well understood in 1928 even if Curry is right that the substitution process is messy. Formula (1) is a set-theoretical construction (of subsets of A^*), so there is no unhistorical "anticipation" of Church.

It is without difficulty that we imagine the recognition of the following two laws:

(α) $\lambda X. \ \tau[X] = \lambda Y. \ \tau[Y],$

provided of course that Y is not free in $\tau[X]$ and that substitution resolves clashes between free and bound variables. Similarly, there is a principle of extensionality:

(ξ)   If, for all $X \subseteq A$, we have $\tau[X] = \sigma[X]$,

        then $\lambda X. \ \tau[X] = \lambda X. \ \sigma[X].$

Perhaps it is also useful to mention a stronger version of this law which takes special advantage of the fact that abstracts are sets:

(ξ*)   If, for all $X \subseteq A$, we have $\tau[X] \subseteq \sigma[X]$,

        then $\lambda X. \ \tau[X] \subseteq \lambda X. \ \sigma[X].$

These laws are justified by the usual behaviour of bound variables and very, very elementary extensionality properties of set formation in (1).

The next law would, admittedly, require a little more imagination. The point is that we have already seen that there is a one–one correspondence between certain multi-relations and continuous set mappings. What would require a grasp of formal manipulation would be the insight of how to *express* this in the above notation. But, since class abstraction was already known, we can hope our hypothetical Master's candidate would write:

$$(\beta) \qquad (\lambda X.\ \tau[X])(Y) = \tau[Y],$$

provided the mapping $x \mapsto \tau[X]$ is a **continuous** set mapping on subsets of A. This law expresses exactly the fact, *already noted*, that if we use (1) above to define a multi-relation from a continuous map, then under the definition of application we achieve *the same* mapping by applying the multi-relation to any suitable argument $Y \subseteq A$.

The aspect to emphasize at once is that all this discussion is almost purely notational: (α), (β), (ξ) are just *notational variants of facts already known*. There is no hint of tricks of self-application yet, and we need to ask: "How could a coherent notion of self-application be based on such set-theoretical definitions?"

Look again at the definition given in (1) above. There is a plain distinction of **type** involved, because while $X \subseteq A$ is intended for *arguments*, the *results* of what is defined are multi-relations

$$\lambda X.\ \tau[X] \subseteq A^*.$$

Would anyone (let alone our Master's candidate) ever be tempted to confuse these types (that is, to confuse subsets of A and subsets of A^*)?

Obviously, I want the answer to be "yes", but can I motivate it? Let us stop to reason that a confusion of the kind requested would require that subsets of A^* be at the same time subsets of A. If the set A were *closed under the formation of finite sequences*, then we would have $A^* \subseteq A$ and, trivially, $X \subseteq A^*$ would imply $X \subseteq A$.

Are there such sets $A^* \subseteq A$? Perhaps Russell and Wiener might have balked at first, because Wiener's definition of pairs would—strictly applied —have resulted in sets of *infinite type*. But Kuratowski, Hausdorff, Zermelo or Von Neumann would have had no trouble with the suggestion. For Russell, it would have been simple enough to *axiomatize* the finite-sequence-forming operation and to point out how many models there are where we can regard $A^* \subseteq A$. It would even be easy to go further and eliminate *all* type distinctions previously found by getting $A^* = A$. (In Zermelo's theory, consider the *least* set closed under the formation of finite sequences—"generated" from $\varnothing$, so to speak.)

Having had the idea of closing up A under sequences, we remark that the self-application $P(P)$ makes perfectly good sense, because now $P \subseteq A^*$ $\subseteq A$ and $P(P)$ is defined. Even better, from eq. (6) of the last section, we see that $X(Y)$ *always* makes sense for all $X, Y \subseteq A$. Note that

$$X(Y) = (X \cap A^*)(Y),$$

but, since $A^* \subseteq A$, we lose no continuous functions this way. Also it is *obvious* from the definition that $X(Y)$ is continuous in *both* X and Y (the operation is even distributive in X). This is the main insight: the uniformly "type-free" definition of $X(Y)$ is permitted for a very elementary set-theoretical reason. But will we be motivated to go further?

Once a person has an operation he naturally tries to *iterate* it. A particularly simple case of iteration is something like $P(X)(Y)(Z)$. Take P fixed. The set P is a multi-relation applied to X. But the result (possibly better regarded as restricted to A^*) is a multi-relation we can apply to Y; and so on. Once we say to ourselves: "Sets determine multi-relations; the values are sets; these values determine multi-relations in turn", there is no reason to stop.

Now $P(X)(Y)(Z)$ is a three-place function, continuous in all its arguments (recall what we already remarked about composition). We should ask: "What three-place continuous functions do we get in this way?" Surely this is a natural question. The answer is "all".

Theorem. *Assume $A^* \subseteq A$. Let $X_0, X_1, \ldots, X_{n-1} \mapsto \tau[X_0, X_1, \ldots, X_{n-1}]$ define a continuous set mapping of n-variables. Let*

$$P = \lambda X_0 \lambda X_1 \cdots \lambda X_{n-1}. \ \tau\big[X_0, \ldots, X_{n-1}\big].$$

Then P is a multi-relation such that

$$P(X_0)(X_1) \cdots (X_{n-1}) = \tau\big[X_0, X_1, \ldots, X_{n-1}\big]$$

for all $X_0, X_1, \ldots, X_{n-1} \subseteq A$.

This result is the *n*-ary generalization of the principle (β) already mentioned. Recall that—given Chapter 1—(β) followed by *definition*. The above theorem would follow at once from *iteration* of principle (β) provided we establish (the true)

Lemma. *If $X_0, X_1, \ldots, X_n \mapsto \tau[X_0, X_1, \ldots, X_n]$ is a continuous set mapping of $(n+1)$ variables, then*

$$X_0, X_1, \ldots, X_{n-1} \mapsto \lambda X_n. \ \tau\big[X_0, X_1, \ldots, X_n\big]$$

is a continuous set mapping of n-variables.

The proof of the lemma is an elementary consequence of formula (1) of this section (where $\tau[X]$ is replaced by $\tau[X_0, X_1, \ldots, X_n]$ and X is replaced by X_n). One only has to take a free set variable, replace it by a directed union, and then bring the union to the outside of the set abstraction by valid set-theoretical principles.

The point of the theorem is that iterated *application* is the "inverse" of iterated *abstraction*—provided we have a lemma to the effect that abstraction preserves continuity.

Now the Master's thesis is taking on more substance; however, at this point self-application has still played no major role. Rather, we have motivated and implemented in the present theory Schönfinkel's idea that *n*-ary functions can be reduced to monadic functions (at least in this context of continuous set functions). The question that remains is: "Where would the combinators have come from?"

It is not fair to answer that Schönfinkel's paper was published before 1928 and that our Master's candidate may have read it. Let us try to get him to discover the combinators himself. Recall that under the assumption $A^* \subseteq A$ we can regard application as a binary set operation:

(2) $X(Y) = \{ y | \exists n \, \exists y_1, \ldots, y_n \in Y. \ \langle y, y_1, \ldots, y_n \rangle \in X \}.$

(This is just (6) of the last section.) Even in the notation of *Principia* it would have been readily seen that this is continuous in both variables— even if we did not cheat and use the three dots. Compositions of continuous functions are continuous; λ-abstractions of continuous functions are continuous; these constructs can be iterated as much as we please; hence, consider the iterated combination

$$\mathbf{B} = \lambda F \lambda G \lambda X. \ F(G(X)).$$

By the Theorem, $\mathbf{B}(F)(G)$ represents the *composition* of F and G; that is, $\mathbf{B}(F)(G)$ is a multi-relation representing the function represented by the composition of the functions represented by the multi-relations F and G. (This talk of representation is clearly becoming too complex.) Therefore, $\mathbf{B}$ itself represents the "general idea" of composition. This is probably the first non-trivial combinator to be discovered. As is common, we shall use in the sequel the more readable infix notation:

$$F \circ G = \mathbf{B}(F)(G).$$

Admittedly, this discussion all seems totally trivial *after* the work of Curry and Church, but there are two remarks I might make. In the first place, even today the combinators are not all that well-known (though studies in Computer Science have given them a new lease on life). Thus

there is a contemporary problem of motivating them when, say, trying to explain the idea to other mathematical but non logically-trained colleagues. The interpretation proposed here is, I feel, a rather direct way of showing that a non-trivial combinatory algebra is possible (and it is definitely easier than the non-extensional algebra of Gödel numbers under $\{e\}(n)$ as application). Having realized this, the second remark may have greater weight: I believe that in 1928 someone could have reasoned: "If I *can* iterate certain continuity-preserving operations, then I *should* iterate them and try to find out what they give." Many mathematical discoveries have been made on less clear grounds. It strikes me as a suitably "empirical" approach.

So ends Chapter 2 of the hypothetical Master's thesis: the assumption $A^* \subseteq A$ has been utilized more for iteration of concepts than for self-application. What has been shown is that a full theory of continuous functions (as exactly represented by multi-relations) is achieved in such a way that a notation of application and abstraction is completely meaningful in any degree of intermixing (iteration). Moreover, certain obvious laws (such as (α), (β), (ξ), (ξ^*)) hold in what has now become a "type-free" manner. These laws are enough to justify many laws of combinators (e.g. the associativity of **B**). The construction has provided a model for a coherent (though not as yet very formal) theory of application and abstraction—a theory which, as we shall see, is open to many extensions.

It should be stressed that the theory of continuous functions for which these laws are valid is one where the notion of function (as a set mapping) is replaced by the notion of a representing multi-relation. The map $X \mapsto \tau[X]$ (if continuous) is represented perfectly by the corresponding multi-relation $\lambda X. \ \tau[X]$ of formula (1): we could call this multi-relation the *graph* of the map. But the reader must heed the fact that only *continuous* set mappings have graphs of this kind.

Looking back at the early writings on λ-calculus such a point of view may seem very limited, and what has to be done is to illustrate the *scope* of such a theory of continuous functions. In any case in what follows we are simply going to identify continuous functions with their graphs and use this idea as our principal notion of function. Note, however, one advantage of our approach over the formal, purely axiomatic induction of λ-calculus by Curry and Church: in this context the combinators are seen as *special* continuous functions—there are many more such functions than merely those defined in some restricted notation. In this way the combinators (in combination with other operators) are given a greater range of applicability over just the "logical" uses intended by their discoverers.

(Returning from 1928 for a moment to 1972, it should be remarked that Plotkin defined the model by forming not A^* but the closure of A under

ordered pairs of finite subsets of the set being constructed. The two closures are of the same order of complexity, but the author feels that the use of A^* makes the definitions look more elementary.)

4. Some notes on iteration and computability

In the previous section self-application (better: confusion of types) was made palatable (or at least: coherent) through the multi-relational representation of continuous set mappings and the closure condition $A^* \subseteq A$. It should be pointed out again, however, that very little advantage was taken of the possibility of self-application, as the "natural" combinators do not emphasize self-application. Since Church was fundamentally concerned with the paradoxes, he considered the question from the start, but his monograph CHURCH (1941) does not on the surface exploit the freedom of self-application very much.

Whether our Master's candidate would have grasped the significance of diagonal applications is not so compelling an hypothesis. But, nevertheless $X(X)$ is at once seen to be a continuous set function; perhaps the step to looking at

$$(\lambda X. \ X(X))(\lambda X. \ X(X))$$

would have required a certain curiosity. Certainly, the paradoxes were widely discussed. So even the kind of formula that enters in their derivations is not all that weird:

$$(\lambda X. \ F(X(X)))(\lambda X. \ F(X(X))).$$

(Here, as we have already remarked, **not** should replace F for the sake of paradox.) If we regard the above as a continuous map $F \mapsto Y(F)$ (that is, Y is λF of the above expression), then—as in the derivation of the Russell Paradox—we have by a use of (β)

$$Y(F) = F(Y(F)).$$

(Surely this is the way Curry discovered the so-called Paradoxical Combinator; note his derivation of the Russell Paradox in CURRY (1930, p. 511).)

Perhaps by now our Master's student is working on his Doctor's thesis. If he were the kind of person to submit papers to *Fundamenta*, he would have surely read KNASTER (1928) written in 1927. This note contains the general Fixed-Point Lemma for Monotone Set Mappings, a joint result of Knaster and Tarski. By a monotone map, we understand a function such that

(1) $X \subseteq Y$ always implies $F(X) \subseteq F(Y)$.

Here $X, Y \subseteq A$ and the values of the function have the same type: $F(X) \subseteq A$. The following theorem specializes from complete lattices the statement of TARSKI (1955), a paper written in 1953 and reporting on results of 1939 and earlier.

Theorem. *Every monotone set mapping $F : PA \rightarrow PA$ has a least fixed point, and indeed the set of all fixed points, $\{X \subseteq A \mid X = F(X)\}$, forms a complete lattice under $\subseteq$.*

Of course, continuous set mappings are monotone (*Hint*: consider the directed family of sets $\{X, Y\}$, where $X \subseteq Y$), and so the theorem applies to that case. We need not detail the whole and quite elementary proof of the theorem here, but a few words about it might be helpful. Let us take the problem of proving the existence of the *least* fixed point. Consider the family of sets "closed" under F, namely $\{X \subseteq A \mid F(X) \subseteq X\}$. Even though F is not a closure operation and is not assumed to be any better than monotone, we can still easily prove that this family of "closed" sets is closed under arbitrary intersections; in particular, the intersection of the *whole* family belongs to the family itself (that this intersection exists comes from the fact that the set A is itself "closed"). Call this intersection X_0. We have $F(X_0) \subseteq X_0$, and then by monotonicity $F(F(X_0)) \subseteq F(X_0)$. This proves that $F(X_0)$ is "closed" also; whence, $X_0 \subseteq F(X_0)$ because X_0 is already known to be the least one in that family. Obviously, then, X_0 is not only a fixed point but also the least one in the family of fixed points (since every fixed point is "closed").

In the continuous case, the least fixed point of F can be "found" by a simple infinite iteration; calling the least fixed-point operator **fix** we can write:

$$(2) \qquad \mathbf{fix}(F) = \bigcup_{n=0}^{\infty} F^n(\varnothing).$$

On the right, $\varnothing$ is the empty set. The proof by continuity that the operator does indeed give the least fixed point is *very* elementary, for one has only to note that on the right the union is directed. (In the non-continuous case, the iteration has to be continued into the transfinite.)

In KLEENE (1952), the nub of the proof is embedded in his proof of the First Recursion Theorem (pp. 348 ff), where the verification that partial recursive *functionals* are continuous is carried back to first principles. But note that Kleene is proving two things: the least fixed point exists *and* it is computable. (Strictly speaking, he is dealing with partial functions not subsets; a discussion of the obvious connection is given in SCOTT (1976), which also relates closely to the considerations of the present paper.) These

facts ought to be separated, since the existence part is more general and more elementary; in fact, we can prove existence long before we have introduced tools to make computability (a kind of definability) precise. Keep in mind, however, that this remark does not eliminate any of the problems of showing that *particular* functions are continuous; all we have at hand so far from the last section is a method for generating a large number of continuous—and quite complex—functions from other continuous functions.

Now look back at the definition of **flx** in (2) above. For each integer n, the map $F \mapsto F^n(\varnothing)$ is continuous by the composition principle. (Now we are, as usual, identifying functions with their graphs as mutli-relations.) Any point-wise union of continuous functions is continuous. *Hence,* the union map $F \mapsto \mathbf{flx}(F)$ is continuous: **flx** must have a graph. What is it?

By now we should probably give up the rhetorical device of arguing that all this could have been done in 1928. I believe I have at least argued that the *discovery* of the combinators and their elementary laws could have been given a set-theoretical (and, for my taste, *natural*) grounding in 1928. On the other hand, whether their applications and uses would have been so quickly recognized is not clear, since the theory of recursive functions took some time and somewhat different motivation in order to get started. (Mainly, Gödel and the incompleteness theorems were required.) Note, however, that the reason that Knaster and Tarski introduced the fixed-point method was to produce the *recursions* needed in set-theoretical arguments like the Schröder–Bernstein Theorem and many generalizations. Perhaps, then, the step to a general theory of recursion could have come forward along these lines—but such conjectures are more or less pointless.

So we return to the question of relating **flx** to the "pure" combinators. Since $\mathbf{flx}(F)$ is always the least fixed point of the mapping $X \mapsto F(X)$, and since $\mathbf{Y}(F)$ is *some* fixed point, then $\mathbf{flx}(F) \subseteq \mathbf{Y}(F)$ for all F. It was a most useful discovery of DAVID PARK (1969) that $\mathbf{flx} = \mathbf{Y}$ was actually possible. (The proof of Park was given for a slightly different kind of interpretation of λ-calculus, but I had no trouble in recasting the idea for a similar proof in SCOTT (1976, p. 569 f).) In order to give the proof here, some analysis of A^* *within* A is required.

Let $\prec$ be a binary relation on A. We say $\prec$ is a *well-founded relation,* if whenever $X \subseteq A$ is non-empty, then $x \in X$ exists with no $y \prec x$ also satisfying $y \in X$. (The element x is "minimal" in X with respect to $\prec$.) We say that A^* is *progressive* in A (with respect to $\prec$) if whenever we have $x_0, x_1, \ldots, x_n \in A$, then $x_i \prec \langle x_0, x_1, \ldots, x_n \rangle$ for $1 \leqslant i \leqslant n$. (Tuples are always worse than their (later) terms.) The essential point of the well-founded, progressive case is that every element of A^* can be well-pictured as a *tree:*

each node, if an element of A^*, has as its immediate descendants the terms of the sequence of *positive* index. If you come to a sequence of length less than two or come to an element of $A \setminus A^*$, then you stop. What the assumption amounts to is that this tree is always *finite*.

Theorem. *If $\prec$ is a well-founded relation on A and if $A^* \subseteq A$ is progressive in A, then the paradoxical combinator $\mathbf{Y}$ is the least fixed-point operator.*

Proof. Let F be a multi-relation representing a continuous set function. Let $B = F(B)$ be *any* fixed point of F. For the moment write $\Delta = \lambda X. \ X(X)$. Let $U = \lambda X. \ F(X(X))$, and recall that $\mathbf{Y}(F) = \Delta(U)$. We know $\Delta(U)$ is a fixed point of F, but we must show that $\Delta(U) \subseteq B$.

Let us write $\langle x_0, x_1, \ldots, x_n \rangle \ll U$ provided $\{x_1, \ldots, x_n\} \subseteq U$. By continuity

$$\Delta(U) = \bigcup \ \{\Delta(\{x_1, \ldots, x_n\}) | \langle x_0, x_1, \ldots, x_{n-1} \rangle \ll U\}.$$

Thus, we need only show that

$$\langle x_0, x_1, \ldots, x_n \rangle \ll U \text{ always implies } \Delta(\{x_1, \ldots, x_n\}) \subseteq B.$$

So suppose *not*. Let S be the set of tuples $\ll U$ for which the inclusion *fails*. Since $\prec$ is well-founded, let $\langle x_0, x_1, \ldots, x_n \rangle$ be a minimal element of S. We seek a contradiction.

To this end, let y_0 be any element of $\Delta(\{x_0, x_1, \ldots, x_n\}) \setminus B$. Because

$$y_0 \in \{x_1, \ldots, x_n\}(\{x_1, \ldots, x_n\}),$$

we see from formula (2) of the last section that there are elements $y_1, \ldots, y_m$ where

$$\{y_1, \ldots, y_m\} \subseteq \{x_1, \ldots, x_n\}, \quad \text{and} \quad \langle y_0, y_1, \ldots, y_m \rangle \in \{x_1, \ldots, x_n\}.$$

But then, by assumption,

$$\langle y_0, y_1, \ldots, y_m \rangle \prec \langle x_0, x_1, \ldots, x_n \rangle.$$

Since $\langle y_0, y_1, \ldots, y_m \rangle \ll U$, we know then that $\Delta(\{y_1, \ldots, y_m\}) \subseteq B$. But

$$\langle y_0, y_1, \ldots, y_m \rangle \in U = \lambda X. \ F(X(X)),$$

so by definition $y_0 \in F(\Delta(\{y_1, \ldots, y_m\}))$. By monotonicity

$$F(\Delta(\{y_1, \ldots, y_m\})) \subseteq F(B) = B,$$

so $y_0 \in B$, which is impossible!

Actually we need not invoke an arbitrary $\prec$ on A for the above result, because there is always a *least* relation for which A^* is progressive in A. We can call the whole structure $A^* \subseteq A$ a *well-founded model for tuples* provided this least relation is well-founded.

But what does this have to do with computability and Kleene's First Recursion Theorem? We need some definitions. A structure $A^* \subseteq A$ will be said to be *computable with respect to an enumeration* $A = \{a_n | n \in \mathbb{N}\}$ (where $\mathbb{N} = \{0, 1, 2, \ldots\}$) provided that we can show uniformly that the relationship

$$\langle a_{n_0}, a_{n_1}, \ldots, a_{n_{k-1}} \rangle = a_{n_k}$$

is recursively enumerable in $n_0, n_1, \ldots, n_k$.

In the same vein we say a *subset* $U \subseteq A$ is *computable* if $\{n \in \mathbb{N} | a_n \in U\}$ is recursively enumerable. A *continuous set function* $X_0, X_1, \ldots,$ $X_{m-1} \mapsto \tau[X_0, X_1, \ldots, X_{m-1}]$ is *computable* provided its graph $\lambda X_0 \lambda X_1 \cdots \lambda X_{m-1} \cdot \tau[X_0, X_1, \ldots, X_{m-1}]$ is computable.

Theorem. (i) *In a computable structure, the computable functions contain all the computable constants and are closed under composition and λ-abstraction;*

(ii) *Application is a computable function (of two arguments), and the computable functions are closed under point-wise application;*

(iii) *If $G(X_0, X_1, \ldots, X_n)$ is computable function, then so is the least solution to the equation*

$$F(X_1, \ldots, X_n) = G(F, X_1, \ldots, X_n);$$

in fact, **fix** *is a computable function;*

(iv) *Provided $A^* \subseteq A$ is well-founded,* **fix** *can be proved computable by explicit definition as the* Y-*combinator.*

Proof. Though the theorem has been stated for *functions* (of several variables)—which we could think of as defined by *terms* (with the appropriate number of free variables) built up from constant symbols for computable subsets of A by application and λ-abstraction—here is a case where the argument is easier by combinators. We reduce all λ-definitions to the well-known combinators **S** and **K**, which we have to prove are computable subsets of A. Then we need only check that if U and V are computable sets, then so is $U(V)$. This will take care of parts (i) and (ii). Recursion in part (ii) still requires that we show **fix** to be computable; however, in part (iv) in the well-founded case, there is nothing to prove, because **Y** is λ-definable by our previous theorem.

Turning now to the details we recall

$$\mathbf{K} = \lambda X \lambda Y. \, X.$$

By the definition of λ-abstraction this comes out:

$$\mathbf{K} = \{\langle \rangle\} \cup \{\langle x_0, x_1, \ldots, x_n \rangle | x_0 \in \lambda Y. \, \{x_1, \ldots, x_n\}\}$$

$$= \{\langle \rangle\} \cup \{\langle \langle \rangle, x_1, \ldots, x_n \rangle | x_1, \ldots, x_n \in A\}$$

$$\cup \{\langle \langle y_0, y_1, \ldots, y_m \rangle, x_1, \ldots, x_n \rangle | y_0 \in \{x_1, \ldots, x_n\}\}.$$

This means that $a \in \mathbf{K}$ is equivalent to

$$a = \langle \rangle \vee \exists n \exists x_1, \ldots, x_n \in A. \ a = \langle \langle \rangle, x_1, \ldots, x_n \rangle \vee$$

$$\exists n, m \exists x_1, \ldots, x_n \in A \exists y_0, \ldots, y_m \in A \exists z \in A. \ a = \langle z, x_1, \ldots, x_n \rangle \wedge z$$

$$= \langle y_0, y_1, \ldots, y_m \rangle \wedge [y_0 = x_1 \vee \cdots \vee y_0 = x_n]].$$

This is wholly existential on r.e. predicates (r.e. with respect to the enumeration), so $\mathbf{K}$ is computable.

In order to be able to state the formula for $\mathbf{S}$ it is convenient to make some definitions:

$$A_{[0]} = \{ \langle \rangle \};$$

$$A_{[n+1]} = \{ \langle u_0, u_1, \ldots, u_m \rangle | u_0 \in A_{[n]} \wedge u_1, \ldots, u_m \in A \}.$$

We recall that

$$\mathbf{S} = \lambda F \lambda G \lambda X. \ F(X)(G(X)).$$

(**S** is of course the combinator for point-wise application of two functions.) Now we can write:

$$\mathbf{S} = A_{[0]} \cup A_{[1]} \cup A_{[2]} \cup$$

$$\{ \langle \langle \langle x_0, x_1, \ldots, x_k \rangle, g_1, \ldots, g_m \rangle, f_1, \ldots, f_n \rangle |$$

$$x_0 \in \{ f_1, \ldots, f_n \} (\{ x_1, \ldots, x_k \})(\{ g_1, \ldots, g_m \}(\{ x_1, \ldots, x_k \})) \}.$$

To show that $a \in \mathbf{S}$ is characterizable by an existential predicate, the main difficulty is in showing

$$x_0 \in \{ f_1, \ldots, f_n \}(\{ x_1, \ldots, x_k \})(\{ g_1, \ldots, g_m \}(\{ x_1, \ldots, x_k \}))$$

is existential. The first step is to make this equivalent with

$$\exists p \exists t_1, \ldots, t_p \in \{ g_1, \ldots, g_m \}(\{ x_1, \ldots, x_k \}).$$

$$\langle x_0, t_1, \ldots, t_p \rangle \in \{ f_1, \ldots, f_n \}(\{ x_1, \ldots, x_k \}).$$

It then takes two more applications of the definition of application to get down to the elements of the finite sets. What is new in this calculation over the one for $\mathbf{K}$ is that the clause of the form $t_0, \ldots, t_p \in U$ is really a *finite conjunction* of variable length. A more formal (and less dotty) notation for finite sequences would give us existential quantifiers mixed with *bounded universal quantifiers* (finite conjunctions) which would be enough to show that the predicate is r.e.

We use the same kind of argument to show the closure of the computable sets under application. The case of **flx** is similar, since $a \in \mathbf{flx}$ means

$$\exists n, m \exists f_0, f_1, \ldots, f_m \in A. \ a = \langle f_0, f_1, \ldots, f_m \rangle \wedge f_0 \in \{ f_1, \ldots, f_m \}^n(\varnothing).$$

The iterated application has to be shown to be uniformly r.e. in the parameters (including n) with respect to the underlying enumeration of A. It should be clear how to do this. The reason why it is sufficient for (iv) merely to consider **fix** is that we can write:

$$F = \textbf{fix}(\lambda H \lambda X_1 \cdots \lambda X_n. \ G(H, X_1, \ldots, X_n));$$

and, since G is computable, everything on the right is also.

An interesting corollary to this theorem is the observation that the computable subsets of any computable structure form a model of the λ-calculus. This is one rather strong reason why this approach to model building can be considered "constructive."

What is the difference between the *Fixed-Point Theorem* and the *First Recursion Theorem*? The former only involves the proof that fixed points *exist* and that the least fixed-point operator is *continuous*. (The second part of this statement is essential for *iterated* recursive definitions: things introduced by fixed points can be employed for further introductions by fixed points—the reader will get the (fixed) point here if he recalls that parameters have to be allowed.) The latter statement further requires a proof that (assuming we start with computable things) the results are computable. The Second Recursion Theorem (due to Kleene) would demonstrate that not only can we effectively Gödel number all these definitions, but in a recursion for a function F the right-hand side can employ the Gödel number of the very function being defined (that is, there is a function whose recursive definition calls for its own Gödel number). We do not discuss the details here for the present model, but a quite neat version was given in SCOTT (1976, Section 3).

We should stop to consider whether suitable structures exist. The minimal one where A equals the closure of $\varnothing$ under finite sequences is computable and well-founded. It does not make any difference here what kind of sequence we consider (what our model is), because *all* theories of sequences must satisfy

$$\langle x_0, \ldots, x_{n-1} \rangle = \langle y_0, \ldots, y_{m-1} \rangle \leftrightarrow n = m \wedge \forall i < n. \ x_i = y_i,$$

and, when we start with "nothing," all theories lead to the same minimal model $A = A^*$ (up to isomorphism). The same would go for a model generated by *atoms* (i.e., elements which are non-sequences). This kind of structure is the minimal solution of

$$A = B \cup A^*,$$

where we require that $B \cap A^* = \varnothing$. All such models are well-founded structures determined up to isomorphism by the cardinality of B. Provided that B is countable, there are many enumerations of A making the

structure computable. Provided also that we strengthen the definition of well-foundedness to include $x_0 \prec \langle x_0, x_1, \ldots, x_n \rangle$, then *all* strongly well-founded structures are of this form.

In an enumeration making $A^* \subseteq A$ computable it is not necessarily the case that $B = A \setminus A^*$ is computable (r.e. with respect to the enumeration). Similarly, the inequality predicate $x \neq y$ need not be computable; perhaps these stronger conditions ought to be imposed in the definition. In a different direction perhaps we should only consider computable those predicates r.e. with respect to *all* enumerations of A making $A^* \subseteq A$ (or whatever) computable.

So much for remarks on "standard" models. A "non-standard" model (but one with only *finite* sequences) is formed by making up a "peculiar" one–one correspondence between A and A^*. Call it $\pi : A^* \leftrightarrow A$. The theory of tuples then uses $\pi(\langle x_0, \ldots, x_{n-1} \rangle) \in A$ in place of $\langle x_0, \ldots, x_{n-1} \rangle$ (This is what is done for $A = \mathbb{N}$ when we use a Gödel numbering of finite sequences.) We could also do a similar theory with atoms, but the extra generality is not needed here. A different way of introducing atoms into A is suggested in the next section.

It is a quite remarkable discovery of BAETEN and BOERBOOM (1978) that such non-standard models completely change the behavior of the resulting λ-calculus. (Earlier, J. Owlett when a graduate student at Oxford had found that the connection between **fix** and **Y** could be ruined by a non-standard tupling.) The result of Baeten and Boerboom (proved for the graph model of SCOTT (1976) but clearly transferable to the present context) can be stated as follows (A is assumed denumerable):

Theorem. (i) *Let $X \subseteq A$ be arbitrary. There is a choice of $\pi : A^* \leftrightarrow A$ so that under this notion of tuple $\Delta(\Delta) = X$ (instead of $\varnothing$ as in the well-founded case).*

(ii) *Given any closed λ-term τ, there is a choice of $\pi : A^* \leftrightarrow A$ so that $\Delta(\Delta) = \tau$.*

The proof of (ii) is by a fairly straight-forward forcing construction, where the forcing conditions are certain finite partial functions $P \subseteq \pi$ (for $\pi : A^* \leftrightarrow A$). It would be worthwhile to check whether the proof works with terms involving **fix** as a primitive combinator and not defined as **Y**. The resulting tupling is non-recursive, by the way.

Having provided a somewhat detailed analysis of iteration in these models and a glimpse of how recursive definitions enter *via* combinators, we ought to conclude this section with some remarks on how *properties* of the recursively defined functions should be proved.

In Section 6 of CURRY (1979) we find a useful, but brief review of combinatory arithmetic and Kleene's early work on the λ-definability of recursive functions (see the Curry paper for the explicit references). Perhaps we should remark that the word "definability" is not quite properly used in this regard in view of later work on the connection between recursion theory and formalized calculi: it would be better to say that Kleene established the *numeralwise representability* of partial recursive functions in the (pure) λ-calculus. The reason for making this verbal distinction is that the "numerals" are each taken *separately* (Curry calls them the combinators Z_n), and there is no predicate in the theory for the *class* of integers. Therefore, even though we can see the results of any *one* calculation, there is no way to formulate—in the theory—a proof by mathematical induction in order to establish *general* facts about the integers. (The variable n is outside the system, for example.) This strikes me as something of a drawback, but of course Curry was striving for the weaker, more basic, more ultimate foundational systems he wanted to see *common* to all formalized theories. Not all theories, obviously, should be as strong as first-order arithmetic. Be that as it may, there is still a question of just how (or where) we are to do our inductions.

Instead of introducing by one scheme or the other the integers into our present system, we will fix attention for the sake of illustration on the combinator **fix** as the embodiment of the iteration concept. For us this is reasonable because we always have continuous set functions at the back of the mind, and the least-fixed-point construction is quite fundamental in this context. We have already seen how to use **fix** in definitions; what remains is to see how it comes into inductions.

In a highly schematic way, we could consider *directed-complete predicates* of subsets of A. Such a predicate $\mathscr{P}(X)$ has the property of being **closed** under directed unions of sets. Now the least fixed point is a directed union; thus, it is certainly valid in our model that the following holds for all directed-complete predicates:

$$(\iota\uparrow) \qquad \mathscr{P}(\varnothing) \wedge \forall X [\, \mathscr{P}(X) \rightarrow \mathscr{P}(F(X)) \,] \rightarrow \mathscr{P}(\mathbf{fix}(F)).$$

The only trouble with $(\iota\uparrow)$, the principle of directed-complete induction, is that it would require some machinery for the introduction of the predicates. If we want a more elementary principle of the nature of (ξ) or of (η), then we need to take a form of the predicate expressible in our previous notation. Here is one example:

$$(\iota^*) \qquad P(\varnothing) \subseteq Q(\varnothing) \wedge \forall X [\, P(X) \subseteq Q(X) \rightarrow P(F(X)) \subseteq Q(F(X)) \,]$$
$$\rightarrow P(\mathbf{fix}(F)) \subseteq Q(\mathbf{fix}(F)).$$

It is simple to verify that $P(X) \subseteq Q(X)$ *is a directed-complete predicate. Quite a lot can be done with this form of induction, though not everything.* An interesting example of something statable by combinators but requiring induction to prove was given in SCOTT (1976, p. 534):

$$\text{flx}(\lambda F \lambda X. \ G(X)(F(X))) = \lambda X. \ \text{flx}(G(X)).$$

This equation can easily be rewritten in combinators: **B(flx(S)) = B(flx)**. Other examples and references can be found in BARENDREGT (1977, pp. 1121 and 1126).

As special cases of (ι^*), we remark that the implication

$$F(A) \subseteq A \rightarrow \text{flx}(F) \subseteq A$$

follows at once by the substitutions $P = \lambda X. \ X$ and $Q = \lambda X. \ A$. In the appendix to this paper we collect together some axioms suggested by our model construction (cf. also SCOTT (1973)). The reader should note two things: in the first place, this is a first-order theory not just an equational theory. (One reason for this is the fact that it seems impossible to regard the *unrestricted* quantifiers as combinators—they are not continuous as operators.) Secondly, what we propose is a very weak theory, because it could be interpreted within the theory of r.e. sets and very likely proved consistent in ordinary first-order arithmetic. In any case it is formulated in as close a form to the original view of combinatory logic as we can come under the plan of modelling functions by continuous set mappings.

5. Some aspects of type theory

In this section of the paper we shall restrict attention to the models where $A^* = A$. As we pointed out before, not only do such sets exist, but this equation eliminates completely the distinction in type between subsets $X \subseteq A^*$ and $X \subseteq A$. Every set in our "universe" is at the same time a set of sequences, and conversely. We saw this was helpful as regards the discussion of continuous set functions; however, this initial elimination of one type distinction hardly eliminates all type distinctions—there is another important one close at hand.

Is there a difference between functions and arguments? Certainly in *use*, but the model presented above has shown that functions may be incorporated as objects among the arguments. That was how we justified self-application. But, with reference to condition (9) of Section 2, it will be seen that (even when $A^* = A$) only *certain* subsets of A^* are used to represent functions; in a precise sense in this model there are "fewer" functions than arguments. Thus, despite our being able to give a useful meaning to $X(Y)$

for all $X, Y \subseteq A$, the distinction between function and argument remains on the level of the object of the model.

We can easily give symbolic form to the distinction by employing in a well-known way the λ-notation. What does $\lambda X. \ P(X)$ represent (in the model) for arbitrary $P \subseteq A$? *Answer*: the arbitrary continuous function. Condition (9) referred to above is equivalent to the satisfaction of this equation:

(η) $\qquad P = \lambda X. \ P(X)$

Often suggested as a universally valid *law* of λ-calculus, it is often wrongly called the axiom of extensionality—law (ξ) is correctly the extensionality principle for the λ-calculus. CURRY (1979) speaks of "strict" extensionality, which is fair enough. The strictness consists of the requirement that every object *uniquely* represents a function. As is well-known, we could replace (the universal generalization of) (η) by the biconditional:

(φ) $\qquad P = Q \leftrightarrow \forall X. \ P(X) = Q(X).$

By a very exact analogy with the axiom of set theory, whereby two sets with the same *elements* are equal, we can read (φ) as saying that two functions with the same *values* are equal. The rub is that in general we do *not* known whether P and Q are always the chosen representatives of functions. In the case of (ξ), the two λ-abstracts are by primary intent the representatives of the functions in question, and so we say they are equal in an extensional theory. (φ) above says too much, for, just as in set theory, we can imagine a universe where some objects are not functions (some objects are not sets—atoms, for instance). I prefer to call (φ) a principle of functionality, meaning that every object is (uniquely) to be regarded as a function. This should not restrict the use of the word "functionality" for other uses—for example functionality *relative* to certain mapping properties of the kind we shall discuss below.

Indeed, law (η) always *fails* in the kind of model constructed above, because, for $P = \varnothing$, it is clear from formula (1) of Section 3 that $\lambda X. \ \varnothing(X) \neq \varnothing$ for a trivial reason. (It would do no good to leave out the element $\langle \rangle \in \lambda X. \ \tau[X]$. Consider $R = \{\langle x, x \rangle | x \in A\}$, then $R(X) = X$ for all X but $R \neq \lambda X. \ X$ according to the actual definition.) What *is* true in all these models—and this is the reason I have carried around a seemingly superfluous empty sequence—is a somewhat weaker law:

(η^-) $\qquad P \subseteq \lambda X. \ P(X).$

Owing to our assumption that $A^* = A$, arbitrary subsets of A do satisfy (η^-). For if $a \in P$, then either $a = \langle \rangle$, in which case the element belongs to $\lambda X. \ P(X)$ by definition of abstraction, or else we have $a = \langle y_0, y_1, \ldots, y_n \rangle$,

in which case $y_0 \in P(\{y_1,\ldots,y_n\})$, by definition of application, and hence again the element $a \in \lambda X.\ P(X)$.

Interesting as this is, it does not at once answer our question about the distinction between argument and function: this model still makes the distinction, but we want to know whether there is *some* (non-trivial) model in which the law (η) holds. I gave an answer in 1969 with my first model construction by a method that has often been given the unfortunate name "Scottery". (An integrated presentation is planned for BARENDREGT (1980) and a very thorough discussion is contained in PLOTKIN and SMYTH (1978), where the process is given a categorical formulation incorporating suggestions of several other people. Another presentation together with the connections with the topological and lattice-theoretical aspects of continuous lattices will appear in GIERZ ET AL. (1980).) A direct construction (without inverse limits) was mentioned in SCOTT (1976, p. 549 ff), but people have not enjoyed very much reading it there; thus, let me explain once more using the models of this paper how easy it is without trying to put the approach in a wider context. Essentially the same proof is given in PLOTKIN (1972), but the details (by "retracts") as presented here are very much simpler.

Principle (η^-) can be stated purely in terms of combinators and inclusions between them. We have, in fact,

$$\lambda X.\ X \subseteq \lambda F \lambda X.\ F(X),$$

and this is just the start of a sequence of such containments. Define recursively:

(1) $\qquad D_0 = \lambda X.\ X$

(2) $\qquad D_{n+1} = \lambda F \lambda X.\ D_n(F(D_n(X)))$

We can prove the:

Lemma. *For all integers n,*
  (i) $D_n \subseteq D_{n+1}$;
  (ii) $D_n \circ D_n = D_n$.

Proof. For $n = 0$, both (i) and (ii) are clear from what we have already said. Thus assume the case of n and pass to $n + 1$.

We can write by (2) above:

$$D_{n+1}(F) = D_n \circ F \circ D_n,$$
$$D_{n+2}(F) = D_{n+1} \circ F \circ D_{n+1}.$$

Hence if $D_n \subseteq D_{n+1}$, then $D_{n+1} \subseteq D_{n+2}$ follows by monotonicity and (ξ).

Also we see

$$D_{n+1} \circ D_{n+1} = \lambda F. \ D_{n+1}(D_{n+1}(F))$$
$$= \lambda F. \ D_n \circ D_n \circ F \circ D_n \circ D_n.$$

Thus if $D_n \circ D_n = D_n$, then $D_{n+1} \circ D_{n+1} = D_{n+1}$.

Now define $D_\infty = \bigcup \{D_n | n = 0, 1, 2, \ldots\}$. We have:

Theorem. $D_\infty = \lambda F. \ D_\infty \circ F \circ D_\infty = D_\infty \circ D_\infty$, *consequently the fixed points of* D_∞ *are closed under application and the following form of* λ-*abstraction*:

$$\lambda_\infty X. \ \tau[X] = \lambda X. \ D_\infty(\tau[D_\infty(X)]).$$

Moreover, as a model, the fixed points of D_∞ *satisfy* (α), (β), (ξ) *and* (η).

Proof. Both of the first two equations follow by continuity in view of (i) of the lemma. The first comes from (2) and the second from (ii) of the lemma. If $X = D_\infty(X)$ and $Y = D_\infty(Y)$, then

$$X(Y) = D_\infty(X)(Y)$$
$$= (D_\infty \circ X \circ D_\infty)(Y)$$
$$= D_\infty(X(D_\infty(Y)))$$
$$= D_\infty(X(Y)).$$

A simple calculation also shows that

$$D_\infty(\lambda_\infty X. \ \tau[X]) = \lambda_\infty X. \ \tau[X],$$

in view of the equations already proved.

If τ is a term built up by application and λ_∞, then we can leave off the first D_∞ in the formula defining λ_∞ *provided* we assume all free variables have values in the model. It is then easy to check (β), again provided free variables are restricted to the model. The reason that (ξ) holds is that if we assume $\tau[X] = \sigma[X]$ holds for all $X = D_\infty(X)$, then $\tau[D_\infty(X)] = \sigma[D_\infty(X)]$ holds for *all* X. We then employ (ξ) (unrestricted) and the definition of λ_∞. Finally (η) in the model is just a restatement of the first equation of the theorem.

The idea of the theorem is this: the first map D_0 does nothing; the second map D_1 turns everything into a function; the third map D_1 turns everything into a functional in the following sense: given F, it is changed into a mapping which takes its argument, turns that into a function, performs F on the result, and finally converts the answer into a function. In general, D_{n+1} makes everything into an $(n+1)$st-order functional by

performing suitable conversions on arguments and values with the help of D_n. How far can this go on? The answer is: indefinitely! The limit functional D_∞ works arbitrarily deeply on arguments and values, but owing to nice continuity properties of the construction it satisfies the neat fixed-point equations of the theorem. Note, however, that this would not work out so well if we did not have the $n = 0$ case of the inclusion (i) of the lemma.

It should also be noted that the method of proof involves a fixed point —but apparently not one that can be stated in pure λ-calculus. Thus D_∞ is the least fixed point of the following equation:

$$D_\infty = \lambda X. \ X \cup \lambda F. \ D_\infty \circ F \circ D_\infty.$$

We must take care that the theorem is not trivial. In the minimal model for sequence theory, where $A^* = A$ and A^* is generated from "nothing," the least fixed point of D_∞ is analyzed as follows. Indeed by the above the least fixed point is just $D_\infty(\varnothing)$ and we can see

$$D_0(\varnothing) = \varnothing$$

and

$$D_{n+1}(\varnothing) = \lambda X_n. \ D_n(\varnothing);$$

because for all Y it is true that $\varnothing(Y) = \varnothing$. This means that

$$D_\infty(\varnothing) = \lambda X_0. \ \varnothing \cup \lambda X_1 \lambda X_0. \ \varnothing \cup \cdots \cup \lambda X_n \cdots \lambda X_1 \lambda X_0. \ \varnothing \cup \cdots.$$

Trivial as this seems, a strict application of our definitions reveals that $D_\infty(\varnothing) = \bigcup \{A_{[n]} | n = 0, 1, 2, \dots\}$. (See in this regard the calculation of the combinators **K** and **S** in the last section.) Note that every time a new λ comes in a new factor of $A_{[n]}$ goes into the union. But in the minimal model the union of all the $A_{[n]}$ is just A^*; so the only fixed point of D_∞ is the maximal one, A itself. (The situation here is different from the proof mentioned in SCOTT (1976).)

To make a repair (and I did not notice this problem until I started to write up the paper) we must find a non-minimal model; it will, however, turn out to be well-founded though not strongly so. Let A be the closure of the *one* element set $\{\star\}$ under finite sequences. Now here we find that $A \backslash A^* = \{\star\}$ because the element $\star$ should be regarded as a *non-sequence*. (There are many ways in set theory to find such elements.) This model is of course well-founded as we have already remarked. We are next going to take a *quotient* by the least equivalence relation $\equiv$ where

$$\star \equiv \langle \star \rangle \equiv \langle \langle \star \rangle \rangle \equiv \langle \langle \langle \star \rangle \rangle \rangle \equiv \langle \langle \langle \langle \star \rangle \rangle \rangle \rangle$$

$$\equiv \langle \langle \langle \langle \langle \star \rangle \rangle \rangle \rangle \rangle \equiv \langle \langle \langle \langle \langle \langle \star \rangle \rangle \rangle \rangle \rangle \rangle \equiv \cdots,$$

and where $\langle x_0, x_1, \dots, x_{n-1} \rangle \equiv \langle y_0, y_1, \dots, y_{n-1} \rangle$ if $x_i \equiv y_i$ for all $i < n$. The

model that results is $A / \equiv$, which we can think of as the same as $A^* / \equiv$ or $(A / \equiv)^*$ by a slight shift of meaning of the *-operator.

This model would *not* be well-founded in the stronger sense where $x_0 \prec \langle x_0, x_1, \ldots, x_{n-1} \rangle$, but it *is* well-founded in the sense used in Section 4 for the proof of the Y-Theorem. (The quotient could be regarded as resulting from a repetitious replacement of $\langle \star \rangle$ by $\star$ in a given sequence until no occurrence of $\langle \star \rangle$ remains.) In this model $\star \notin A_{[n]}$, and so the *least* fixed point of D_∞ is **not** the *greatest* fixed point, which is still $A = \lambda X. A$. Thus, the fixed-point set of D_∞ has at least two elements, and in this way we have found a non-trivial model for (η). In such a model a further type distinction has been eliminated because *all* elements can be regarded as (unique representatives) of functions (continuous functions). But whether it is really *profitable* to eliminate such distinctions is another question. Note that we could have adjoined as many distinct $\star$-elements as we wished. These elements act just like atoms; thus, the (η)-model would contain something as complex as the space of all continuous set mappings on the (infinite) set of atoms.

Let us therefore turn to the opposite question of how—given a nice λ-calculus model—it is possible to *introduce* type distinctions. There is a point in this, because the distinctions allow us to sort out differences between elements according to natural properties. The advantage of starting with a λ-calculus model is that the whole of the discussion can be built on *one* notation for function abstraction. (An ordinary type theory has, strictly speaking, different application and abstraction notions at all types.) The price for *one* notation for functions is *several* notations for equivalence relations for representing the different types, but this is not so bad since the different types are different in any case.

A considerable amount of detail has already been given in SCOTT (1975) and SCOTT (1976, Section 7). Without making the formulation too heavy, we can describe here briefly how the method works; a deeper investigation would require some familiarity with the theory of continuous lattices and their subspaces. Some further very interesting uses of the idea can be found in PLOTKIN (1973).

Types, for many purposes, can be identified with equivalence relations on (subsets of) our model. Indeed, let $\mathcal{E} \subseteq \mathbf{P}A \times \mathbf{P}A$ be a transitive and symmetric relation. The set of self-related elements, $\{ X \subseteq A \,|\, X \mathcal{E} X \}$, may be regarded as the subspace of the model in question, and this is the subspace of which we are interested in the quotient modulo $\mathcal{E}$. (We shall often write $X : \mathcal{E}$ as short for $X \mathcal{E} X$.) Though this is our interest, we shall not actually take the quotient, for it is easier to work with the representatives of the equivalence classes directly.

For example, let $\mathcal{E}$ and $\mathcal{F}$ be two such equivalence relations. We define the equivalence corresponding to the function space, call it $[\mathcal{E} \to \mathcal{F}]$, by the formula:

(3) $\qquad P\,[\,\mathcal{E} \to \mathcal{F}\,]\,Q \quad$ iff whenever $X \mathcal{E} Y$, then $P(X)\,\mathcal{F}\,Q(Y)$.

That $[\mathcal{E} \to \mathcal{F}]$ is an equivalence relation is clear. Note that P is always equivalent to $\lambda X.\ P(X)$, thus we can regard the equivalence classes as consisting of *functions*. Note, too, that the construction can be iterated—in this way we pass to a notion of *higher-type* function. The reason for stressing equivalence relations rather than classes is that our functions are meant to be *extensional*, in the sense that equivalent arguments should get equivalent values. In words it is easy to read (3): two functions are equivalent if they do equivalent things to equivalent arguments. Keep in mind, however, that (3) has further import depending on how demanding the given equivalence relations are. The point is that (3) implies that if an argument lies in the first subspace, then the value *must* lie in the second subspace—the function is well-defined, therefore.

This plan for defining types *via* equivalence relations has many features of a theory of functionality of the kind advocated by Curry; however, our types are not "obs", that is to say elements of the model. The equivalence relations are constructs *over* the model not elements *of* the model. One approach to having obs represent types (better: classes) was taken in SCOTT (1975), but then a transfinite truth definition is needed in seeing which classes the obs define. This may not be a bad thing, but it is less elementary than we care to be at the moment. There would be no trouble, by the way, in having a theory of equivalence relations (rather than classes) done in the form of the 1975 paper.

Thus there are many approaches to the sorting out of the elements, and still many questions about the nature of possible subsets. In particular, the question of which λ-expressions have types and which types completely determine λ-expressions seems rather basic. For instance, the common combinators are very well behaved as regards type:

$$\mathbf{K} : \mathcal{E} \to [\,\mathcal{F} \to \mathcal{E}\,],$$

$$\mathbf{S} : [\,\mathcal{E} \to [\,\mathcal{F} \to \mathcal{G}\,]\,] \to [\,[\,\mathcal{E} \to \mathcal{F}\,] \to [\,\mathcal{E} \to \mathcal{G}\,]\,].$$

$$\mathbf{B} : [\,\mathcal{F} \to \mathcal{G}\,] \to [\,[\,\mathcal{E} \to \mathcal{F}\,] \to [\,\mathcal{E} \to \mathcal{G}\,]\,].$$

Some further details are given in the cited references, but it seems fair to say that the study of this idea has hardly begun. Here, for example, is a question. The combinator **fix** is very important, but does it have a special character as regards functionality? We are tempted to write:

$$\mathbf{Y} : [\,\mathcal{E} \to \mathcal{E}\,] \to \mathcal{E}.$$

This is not true in general, since we did not put any closure conditions on our equivalence relations (say, closure under directed unions). We should then ask: "Which are the best closure conditions?" as well as the previous question: "How do we prove that a combinator has no functionality?"

6. Some conclusions and some questions

We have spoken at great length about functions and their properties in this essay. In Section 11 of CURRY (1979), Professor Curry gives the well-known reasons why sets can be reduced to functions, and he then continues:

> "Thus, it is simpler to define a set in terms of a function than vice versa (for a similar idea cf. the set theory of VON NEUMANN (1928)); but the idea is repugnant to many mathematicians, and probably to Scott. This has been a great handicap and source of misunderstanding."

May I disassociate myself from these sufferers of repugnance? I feel I understand rather well the logical interrelationships between sets and functions. I would be very happy indeed to reduce sets to functions *if there were any good theory to do this in*. In my opinion there **does not** exist at the present time such a theory—owing to our troubles with the paradoxes. The theory of Von Neumann, for example, turned out to be easier to state as a set/class theory rather than a function theory. What is needed for a workable set theory (regardless of what sets *are*) is a strong comprehension axiom. As far as I can see, the Curry programme has not as yet produced a straight-forward theory that is anywhere near as workable as the standard Zermelo–Fraenkel system (or the system augmented by classes). However, there is a rather fundamental point about the contrast between *extensional* and *intensional* theories of sets and functions, which is hardly touched on in the literature on combinatory logic. For an interesting, and very likely workable intensional theory of functions, see FEFERMAN (1980) and the related papers cited therein. As regards the question of which comes first: the function or the set, it is not a question of repugnance or prejudice on my part that causes me to formulate constructions within set theory but a problem of helplessness. And I can pinpoint rather narrowly where I think the trouble lies.

For the sake of argument think of a set as a truth-valued function. (I know this over simplifies Curry's approach, but a more subtle view is not needed for the point I will make.) Instead of $X \in A$ we will write $A(X)$; to *assert* $X \in A$ means to assert $A(X) =$ **true**. On the other hand, to assert $X \notin A$ means to assert $A(X) =$ **false**. The domain of the function A is

"everything"—but now the rub starts. The Russell Paradox shows that the domain cannot really be everything if we were to allow full comprehension. There is no way around Cantor's Theorem that *there are more functions than there are arguments and values*. (And, from what I can understand, the introduction of some concept of "proposition" to replace the two, very separate truth values does not seem to help.)

As is common knowledge to logicians, the way to make $A(X)$ always meaningful is to restrict by some manner the total possible range of functions. The choice of restriction for the Zermelo − Fraenkel set theory is to make the number of X for which $A(X) =$ **true** *very small* compared to the total number of things (at least insofar as these comparisons of number are expressible in the system). The very feature that makes this view of sets easy to grapple with is that we *do not* need regard sets as functions! The "half" of A consisting of the set $\{X|A(X) =$ **true**$\}$ is enough to determine the whole of A; the other, larger "half" $\{X|A(X) =$ **false**$\}$ is completely determined by the first half. The reason why a half loaf is better than a whole is that—in building up sets—we can regard the first, positive half as **FIXED** long before the rest of the elements that would have to enter into the negative half ever come into view. (This idea of "earlier" and "later" sets is made quite precise in the theory of the *rank* of a set.) There is a certain advantage to regarding the universe of sets as being "open ended" (at the top end, at least) even though we have accepted certain laws as pertaining to *all* sets—no matter how "late" these sets come in. The consequence of this view (which, for all I know, may very well be repugnant to Curry) is that the domain of $A(X)$ as a function is not very well determined on the negative side: our usual set theory is not symmetric in its use of **true** and **false**.

Now the system *New Foundations* of QUINE (1944) (see QUINE (1953) or (1963) for the history of his system) was supposed to restore the **true–false** symmetry by a different kind of restriction on the comprehension axiom. One would hope that Quine's theory would give at once a theoretical basis for a theory of combinators. But it does not—at least if one construes the word "function" the way Quine does as a set of ordered pairs. The comprehension terms needed for the combinators are simply not "stratified" (in Quine's well-known terminology). And why not? Because functions are **binary** relations and *New Foundations* is **not** a suitable system for a general "type-free" theory of binary relations.

This is a, perhaps, not much remarked fact, but it is very easy to explain. Quine's theory *looks* type free, but—sadly—this is only an illusion. When we restrict attention to *one-place* predicates determined by stratified formulae, then it is very true that Quine lets us be *ambiguous* as to type. We can give the free variable any type we want and then start counting up and

down from there. ("Negative" types are permitted, if we like.) When, however, we come to *two-place* predicates, then the story is quite different. There are *two* free variables to cope with now. We can slide the type indices up and down the scale, but in general we can never alter the numerical value of the *difference* between the types of the two variables. In other words, though Quine was successful in banning type distinctions for sets (one-place predicates), he still is faced with *infinitely many* type distinctions for binary relations. Thus, for example, the relations = and ∈ of equality and membership are of essentially different type in Quine's theory: the type difference is 0 in the first case, 1 in the second.

The reader has surely remarked by now that the theory of *continuous* functions employed in this paper could have been carried out equally well in *New Foundations*. In fact, there is a definition of ordered pair in *New Foundations* so that the formula $z = \langle x, y \rangle$ is stratified with the three variables all of the *same* type; moreover, all objects of the theory are ordered pairs, i.e. $V = V \times V$. Though I have not carried out the details, I do not think it would be difficult to change the definition a little so that we could say $V = V^*$. With this understanding, all the basic definitions of this paper would go through, since the defining formulae we have used here are stratified; in particular $Z = X(Y)$ is stratified with all variables of the same type. This is nice, but why does it not settle the question of the relationship of set theory and combinatory logic?

The answer lies in the word "continuous." In order to have the function-theoretic comprehension principle (β) by our approach, we had to make the restriction that $X \mapsto \tau[X]$ defined a continuous set mapping. The kind of comprehension terms needed for set theory (particularly those with quantifiers) *are just not continuous*. We seem to have the choice:

<p style="text-align:center">lots of sets but no combinators</p>
<p style="text-align:center">**or**</p>
<p style="text-align:center">lots of combinators but few sets.</p>

By "set" here we mean the characteristic functions of a set represented by a function of the theory. This is no *proof* that there is no good mixture, but there does seem to be some evidence that the two notions of function in the two kinds of theories are not quite the same. Combinators in the model of this paper behave more like the *classes* of the Von Neumann–Bernays–Gödel theory. People have tried to make classes self-applied, but a "canonical" theory has not been found that has won general favour. Just as we could carry out the construction in Quine's system, we could have worked within VBG class theory and spoken about continuous class functions. There is a chance that this might lead to some axiomatic niceties and produce a blend of Curry and Church with Von Neumann–

Bernays–Gödel, but the author is not so certain the effort is worth the trouble. (Such a study might be worth a Master's thesis, however. The candidate should recognize that there are *degrees* of continuity, and that in this paper we have only employed *finitary continuity*. In a full class theory there are transfinite notions of continuity that would probably be more useful.)

Aside from the question of what to do next—if anything—one might ask: "What is special about the combinatory logic of *New Foundations*?" But, as we have no models for Quine's theory, there might not be much to say. (The models of JENSEN (1969) for the theory with atoms could, however, give something new.) A more interesting question might come out of the Von Neumann–Bernays–Gödel set/class theory.

I believe I can now also make clearer my attitude toward *type theory* that Curry discusses in Section 9 of his paper. Professor Curry recalls the harsh tone of SCOTT (1969), written just a few weeks before I discovered the first model construction. The paper was therefore never published, and I recanted on some of my remarks. What I especially reputiated was my feeling (at one time held very strongly) that *combinatory logic did not make good mathematical sense at all*—for instance, in not linking up with the ordinary theory of functions: continuous real-valued functions come to mind. This is not an issue of set-theoretic foundations vs. function-theoretic foundations. It is just a question of having some interesting mathematics. Well, on that score the situation has changed: at least I now know how to embed *every* topological space (along with its function space) into a model of combinatory logic (see SCOTT (1972)).

But, really, nothing has changed in my view since 1969 as regards type theory. I assert: *it is impossible to eliminate from logic and mathematics ALL type distinctions*. As has been illustrated above, *certain* types can be "confused" and then objects of other types can be "forgotten", but no magic so far has ever made a set A and its powerset PA *equal*. Some types are distinct whether anyone chooses to discuss the difference or not. I certainly did not mean to say in 1969 that we need exactly Russell's theory, but I did mean—and still mean—that the kind of type difference that Russell recognized will always be present *somewhere* in a theory of logical objects. Whether the flexibility of combinatory logic will soften the pain of living with these (necessary) distinctions remains to be fully demonstrated in my opinion.

Some comment is also required on Curry's remarks on *conceptualization*. We read (Section 10):

"While it is true that concocting formalisms entirely without regard to interpretation is probably fruitless, yet it is not necessary that there be "conceptualization" in terms of current mathematical intuitions. In fact, mathematical intuition is a result of evolution. Mathematicians depend on their intuitions a great deal; let us hope they always will. But the mathematical intuitions of today are not the same as those of a thousand years ago. Combinatory logic may not have had a conceptualization in what seems Scott's sense; but it did have an interpretation by which it was motivated. The formation of functions from other functions by substitution does form a structure, and this structure it analyzed and formalized. For progress we need the freedom to let our intuitions develop further; this included the possibility of formalizing in new ways."

First "evolution": Though geometry has evolved over 2000 years and the attitude toward the concept of number has radically altered, *still we can sense the continuity of ideas*. The Pythagorean Theorem is still true in Euclidean geometry and the old proofs still stand, even though the Greeks might not have been happy with a proof by analytic geometry. What I always found disturbing about combinatory logic was what seemed to me to be a *complete lack* of conceptual continuity. There were no functions known to anyone else that had the extensive properties of the combinators *and allowed self-application*. I agree that people might wish to have such functions, but very early on the contradiction found by Kleene and Rosser showed there was trouble. What I cannot understand is why there was not more discussion of the question of *how* the notion of function that was supposed to be behind the theory was to be made even midly harmonious with the "classical" notion of function. The literature on combinatory logic seems to me to be somehow silent on this point. Perhaps the reason was that the hope of "solving" the paradoxes remained alive for a long time —and may still be alive. Perhaps the reason was that many people gave up working in the theory. Whatever the reason, I do not think I am reading the record unfairly.

Next "substitution": This is not the place to discuss the well-taken criticisms of the complexity of substitution in the formulation of rules in formal theories, nor do we have time to discuss the pros and cons of real and apparent variables and a logic without variables. The question I have about basic motivation concerns the "structure" Curry mentions in the quote given above. What structure??? I agree that we can regard Group Theory as an analysis of the structure of bijective functions under composition, Boolean Algebra as an analysis of sets under inclusion, Banach Space Theory as an analysis of functions under convergence of infinite series, etc. etc. But Combinatory Logic? It just does not seem to me to be a sound step in *analysis* to say: "We now permit our functions to be self-applied." Just like that. Clearly, after seeing so many *analogous* composition operations of different types, we would dearly wish to put them all

into one big **B**; but the step to **B(B)(B)**, though it may be a small step for Curry, does seem like a big step for the rest of us—especially in the shadow of the paradoxes.

Some acknowledgments

The author is most grateful both to the University of Oxford for granting him a sabbatical year and to the Guggenheim Foundation for a Fellowship. Special thanks are due to the Xerox Palo Alto Research Corporation and the Computer Science Laboratory for their particular generosity in supplying such good working conditions and remarkable text-editing facilities which made the writing of the paper possible. The excellent typing of Melinda Maggiani has been an essential part of putting together the final draft. The many occasions I have had to lecture on the topic of this paper since the Kleene Symposium have also been most useful. I should also like to record here by warm thanks to Professor Kleene for many kindnesses and much help over the years.

Appendix: Some axioms

Throughout the paper we have alluded to various laws of combinators and λ-expressions without being very systematic; thus, it would seem helpful to collect together what is essential by way of formal properties. In the following list, we have tried to follow Curry's notation for the names of laws as closely as possible. However, the theory of this paper takes $\subseteq$ rather than $=$ a primitive and defines the latter. When laws are strengthened by the use of $\subseteq$ we have added an asterisk; in the one case of a weakening we have added a minus to the name.

$$(o^*) \quad \varnothing \subseteq X,$$

$$(\rho^*) \quad X \subseteq X,$$

$$(\tau^*) \quad X \subseteq Y \wedge Y \subseteq Z \rightarrow X \subseteq Z,$$

$$(\sigma^*) \quad X = Y \leftrightarrow X \subseteq Y \wedge Y \subseteq X,$$

$$(\mu^*) \quad X \subseteq Y \rightarrow Z(X) \subseteq Z(Y),$$

$$(\nu^*) \quad X \subseteq Y \rightarrow X(Z) \subseteq Y(Z),$$

$$(\alpha) \quad \lambda X.\ \tau[X] = \lambda Y.\ \tau[Y],$$

(β) $\qquad$ $(\lambda X. \ \tau[X])(Y) = \tau[Y],$

(ξ^*) $\qquad$ $\forall X. \ \tau[X] \subseteq \sigma[X] \rightarrow \lambda X. \ \tau[X] \subseteq \lambda X. \ \sigma[X],$

(η^-) $\qquad$ $P \subseteq \lambda X. \ P(X),$

(π^*) $\qquad$ $F(\mathbf{flx}(F)) \subseteq \mathbf{flx}(F),$

(ι^*) $\qquad$ $P(\varnothing) \subseteq Q(\varnothing) \wedge \forall X[\ P(X) \subseteq Q(X) \rightarrow P(F(X)) \subseteq Q(F(X)) \ ]$

$\qquad\qquad$ $\rightarrow P(\mathbf{flx}(F)) \subseteq Q(\mathbf{flx}(F)).$

(δ^*) $\qquad$ $\mathbf{flx} = \lambda F. \ (\lambda X. \ F(X(X)))(\lambda X. \ F(X(X))).$

The last law has a special character, and the reader might wish to leave it off in view of the large number of models in which it fails. It should also be kept in mind that Curry also formulates his laws as **rules**; we on the other hand in speaking of models have been thinking in terms of first-order theories and the usual notion of truth in models. Nevertheless, our models obviously give interpretations of (some of) Curry's systems. We have also not had time to discuss it here, but the above system is more general than the models of this paper in that we have not formulated principles corresponding to the fact that **P**A is an atomic Boolean algebra—indeed there are any number of interesting models of the above which **do not** form Boolean algebras under $\subseteq$. We have not had time here either to investigate other primitives corresponding to the way in which $A = A^*$ was built from sequences. Finally, we should remark that it is known that the above system is weak, because, with the introduction of **S** and **K** and with a definition of λ, the whole system is *finitely axiomatizable*.

Notes added in Proof (February 1980)

I am much obliged to Professors Church, Curry, and Seldin who wrote me comments and corrections to the original manuscripts. In particular Professor Church wrote briefly to the editors on 2 June 1979 as follows:

> To the best of my recollection I did not become acquainted with Frege in any detail until somewhat later than the period about which Scott is writing, say 1935 or 1936. No guarantee for this, it is just a recollection of something never accurately recorded. But I was attracted to Frege because he does give priority to functions over sets, and his system can be made consistent (presumptively) by imposing a simple type theory. To this I would now add that no doubt such a system can be given as much set-theoretic strength as desired by adjoining strong axioms of infinity.

On 1 May 1979 Professor Curry wrote me a long letter explaining his attitudes toward various of the points I had brought up. I hope to take

account of these remarks in future publications. In the meantime, however, it seems useful to quote two technical remarks from his letter bearing directly on the details:

> Concerning p. 225, my derivation of the inconsistency in CURRY and FEYS (1958), pp. 258-260 (which came originally from JSL 7, pp. 115-117 (1942) is not exactly a simplification of the proof of Kleene and Rosser. I assumed the existence of the combinator **K**, whereas they did not, so that the result is, at least superficially, weaker. In his thesis (1968), pp. 19f, Bunder showed that, if **K** is present, the Kleene–Rosser theorem follows from my assumption. A simplified derivation of the actual Kleene–Rosser theorem has not yet, to my knowledge, been given.
>
> Concerning p. 238, this is not exactly the way I discovered Y. I am not sure just how the discovery was made, or when I adopted the letter "Y" for it. To settle this would require prolonged search in my cellar; which is hardly worth while. I think that the treatment in CURRY and FEYS (1958, §5G) is fairly close to the original approach. This is essentially as follows: If F is the Russell function and N is negation, then
>
> $$Fx = N(xx).$$
>
> But
>
> $$N(xx) = \mathbf{B}Nxx = \mathbf{W}(\mathbf{B}N) = \mathbf{BWB}N.$$
>
> We thus get the paradox by taking
>
> $$F = \mathbf{B}N\mathbf{B}N.$$
>
> To get Y we just express this FF as a function of N. There are various ways of doing this. One way given in CURRY and FEYS (1958, §5G), is
>
> $$Y = \mathbf{WS}(\mathbf{BWB}).$$
>
> However I do not think I used **S** in my earliest work; I did not appreciate its potentialities until later. Another possibility is
>
> $$Y = \mathbf{W}(\mathbf{B}(\mathbf{C}(\mathbf{BWB}))(\mathbf{BWB})).$$
>
> Your treatment may be essentially isomorphic, so to speak, to this, but it seems strange to me. (At a seminar at Harvard about 1926 Whitehead cited Suzanne Langer for the functional (or predicate) form of the Russell paradox; but I think I once saw it in Russell's *Principles of Mathematics*.)

References

BARENDREGT, H. P.

    [1977]   The type-free lambda calculus, *Handbook of Mathematical Logic*, edited by J. Barwise (North-Holland, Amsterdam), pp. 1091–1132.

    [1980]   *The Lambda Calculus: Its Syntax and Semantics* (North -Holland, Amsterdam), to appear.

BAETEN, J. and B. BOERBOOM
  [1978]   Ω can be anything it shouldn't be, Mathematisch Instituut, Utrecht, Preprint no. 84, 9 pp.
CHURCH, A.
  [1932/33] A set of postulates for the foundation of logic. *Ann. Math.*, 33, (1932), 346–366. Second paper, *ibid.*, 34 (1933), 839–864.
  [1941]   *The Calculi of Lambda Conversion* (Princeton University Press) second edition, 1951, reprinted 1963 by University Microfilms, Ann Arbor, MI, 82 pp.
CHURCH, A. and J. B. ROSSER
  [1936]   Some properties of conversion, *Trans. Am. Math. Soc.*, 39, 472–482.
CURRY, H. B.
  [1929]   An analysis of logical substitution. *Am. J. Math.*, 51, 363–384.
  [1930]   Grundlagen der kombinatorischen Logik, *Am. J. Math.*, 52, 509–536 and 789–834.
  [1967]   Logic, combinatory, in: *Encyclopedia of Philosophy, Vol. 4* (Macmillan Co. and The Free Press, New York), pp. 504–509.
  [1968a]   Combinatory logic, in: *Contemporary Philosophy, a Survey*, edited by R. Klibansky (La Nuova Italia, Editrice, Firenze), pp. 295–307.
  [1968b]   Recent advances in combinatory logic, *Bull. Soc. Math. Belg.*, 20, 233–298.
  [1980]   Some philosophical aspects of combinatory logic, in: *The Kleene Symposium*, edited by J. Barwise, H. J. Keisler and K. Künen (North-Holland, Amsterdam).
CURRY, H. B. and R. FEYS
  [1958]   *Combinatory Logic, Vol. I* (North-Holland, Amsterdam).
CURRY, H. B., J. R. HINDLEY and J. P. SELDIN
  [1972]   *Combinatory Logic, Vol. II* (North-Holland, Amsterdam).
FEFERMAN, S.
  [1980]   Constructive theories of functions and classes, in: *Logic Colloquium '78*, edited by D. van Dalen (North-Holland, Amsterdam), pp. 159–224.
GIERZ, G., K. H. HOFMANN, K. KEIMEL, J. D. LAWSON, M. MISLOVE and D. S. SCOTT
  [1980]   *A Compendium of Continuous Lattices* (Springer-Verlag, Berlin), to appear.

JENSEN, R. B.

   [1969]   On the consistency of a (slight?) modification of Quine's new foundations, *Synthese*, **19**, 250–263.

KLEENE, S. C.

   [1934]   Proof by cases in formal logic, *Ann. Math.*, **35**, 524–544.

   [1935]   A theory of positive integers in formal logic, *Am. J. Math.*, **57**, 153–173 and 219–244.

   [1936]   λ-definability and recursiveness, *Duke Math. J.*, **2**, 340–353.

   [1952]   *Introduction to Metamathematics* (North-Holland, Amsterdam, and P. Noordhof, Groningen).

KLEENE, S. C. and J. B. ROSSER

   [1935]   The inconsistency of certain formal logics, *Ann. Math.*, **36**, 630–636.

KLINE, G. L.

   [1951]   Review of: Foundations of mathematics and mathematical logic, by S. A. Anovskaa, *J. Symbolic Logic*, **16**, 46–48.

KLOP, J. W.

   [1977]   A counterexample to the Church–Rosser property of λ-calculus $+DMM \rightarrow M$, Mimeographed note dated November 1977, Mathematisch Instituut, Utrecht, 2 pp.

   [1979]   A counterexample to the Church–Rosser property of λ-calculus with surjective pairing, Mathematisch Instituut, Utrecht, Preprint No. 102, 14 pp.

KNASTER, B.

   [1928]   Un théorème sur les fonctions d' ensembles, *Ann. Soc. Polon. Math.*, **6**, 133–134.

PARK, D. M. R.

   [1970]   The *Y* combinator in Scott's lambda-calculus models, Univ. of Warwick, Unpublished notes.

PLOTKIN, G. D.

   [1972]   A set-theoretical definition of application, Memo. MIP-R-95, School of Artificial Intelligence, Univ. of Edinburgh, 32 pp.

   [1973]   Lambda definability and logical relations, Memo. SAI-RM-4, School of Artificial Intelligence, Univ. of Edinburgh, 20 pp.

PLOTKIN, G. D. and M. B. SMYTH

   [1978]   The category-theoretic solution of recursive domain equations, Memo. DAI-RR-60, Department of Artificial Intelligence, Univ. of Edinburgh, 42 pp.

QUINE, W. V. O.

   [1936]   A reinterpretation of Schönfinkel's logical operators, *Bull.*

Am. Math. Soc., **42**, 87–89.

[1944]   New foundations for mathematical logic, *Am. Math. Monthly*, **44**, 70–80.

[1953]   *From a Logical Point of View* (Harvard Univ. Press, 1953, second ed., 1961, reprinted, Harper and Row, 1963), 184 pp.

[1963]   *Set Theory and Its Logic* (Harvard Univ. Press, 1963, second ed. 1969), 361 pp.

[1971]   Algebraic logic and predicate functors, in: *Logic and Art: Essays in Honor of Nelson Goodman*, edited by R. Rudner and Scheffler (Boobs Merrill), pp. 214–238.

ROSSER, J. B.

[1935]   A mathematical logic without variables, *Ann. Math.*, **36**, 127–150, and *Duke Math. J.*, **1**, 328–355.

SCHÖNFINKEL, M.

[1924]   Über die Bausteine der mathematischen Logik, *Math. Ann.*, **92**, 305–316.

SCOTT, D.

[1969]   A type-theoretical alternative to CUCH, ISWIM, OWHY, unpublished.

[1972]   Continuous lattices, in: *Proc. 1971 Dalhousie Conference, Lecture Notes in Mathematics, Vol. 274* (Springer-Verlag, New York), pp. 97–136.

[1973]   Models for various type-free calculi, in: *Proc. IVth Int. Congr. for Logic, Methodology and the Philosophy of Science, Bucharest*, edited by P. Suppes et al. (North-Holland, Amsterdam), pp. 157–187.

[1975]   Combinators and classes, in: *λ-Calculus and Computer Science*, edited by C. Böhm, Lecture Notes in Computer Science, vol. 37 (Springer-Verlag, Berlin), pp. 1–26.

[1976]   Data types as lattices, *SIAM J. Comput.* **5**, 522–587.

[1977]   Logic and programming languages, *Comm. ACM*, **20**, 634–641.

[1977a] An appreciation of Christopher Strachey and his work, in: *Foreward to Denotational Semantics* by J. E. Stoy (MIT Press), pp. XV–XXX.

SELDIN, J. P.

[1980]   Curry's program, to appear.

TARSKI, A.

[1930]   On some fundamental concepts of metamathematics, in translation by J. H. Woodger, in: *Logic, Semantics, Metamathematics* (Cambridge Univ. Press, 1956), pp. 30–37.

[1955]   A lattice-theoretical fixpoint theorem and its applications, *Pac. J. Math.*, **5**, 285–309.

VAN HEIJENOORT, J.

[1967]   *From Frege to Gödel: A Source Book* (Harvard Univ. Press).

VON NEUMANN, J.

[1928]   Die Axiomatisierung der Mengenlehre, *Math. Zeits.*, **27**, 669–752.

WHITEHEAD, A. N. and B. RUSSELL

[1910]   *Principia Mathematica, 3 Vols.* (University Press, Cambridge, 1910–1913; second edition, 1925–1927).