J Autom Reasoning (2006) 37:277–322
DOI 10.1007/s10817-006-9062-x

The Calculus of Relations as a Foundation for Mathematics

Steven Givant

Received: 1 October 2004 / Accepted: 1 July 2005 /
Published online: 2 February 2007

Abstract A variable-free, equational logic $\mathcal{L}^{\times}$ based on the calculus of relations (a theory of binary relations developed by De Morgan, Peirce, and Schröder during the period 1864–1895) is shown to provide an adequate framework for the development of all of mathematics. The expressive and deductive powers of $\mathcal{L}^{\times}$ are equivalent to those of a system of first-order logic with just three variables. Therefore, three-variable first-order logic also provides an adequate framework for mathematics. Finally, it is shown that a variant of $\mathcal{L}^{\times}$ may be viewed as a subsystem of sentential logic. Hence, there are subsystems of sentential logic that are adequate to the task of formalizing mathematics.

Key words calculus of relations • authomated reasoning • algebraic logic • set theory • mathematical foundation

1 Introduction

The calculus of relations is an algebraic theory of binary relations. The intended universe of discourse is the collection of all binary relations on some set. The basic operations include the Boolean operations of forming unions, intersections, and complements of relations, and the relation-specific, or Peircean, operations of forming the inverse of a relation, the composition of two relations, and the dual of composition, the relational sum of two relations. There are also distinguished

This paper is an expanded version of a talk given by the author at the Special Session on Automated Reasoning in Mathematics and Logic, held March 8–10, 2002, at the Georgia Institute of Technology, during the Joint Southeastern Section MAA/Southeast Regional AMS Meeting. The session was organized by Johan G. F. Belinfante.

S. Givant (✉)
Department of Mathematics and Computer Science, Mills College,
5000 MacArthur Boulevard, Oakland, CA 94613, USA
e-mail: givant@mills.edu

relations: the empty relation, the universal relation, the identity relation, and its complement, the diversity relation.

The first paper on the subject [6], was written by Augustus De Morgan in 1864, not long after the pioneering work of George Boole. A proper foundation for the theory was laid by Charles Saunders Peirce in the 1883 paper [31], after a number of earlier investigations. (In connection with its discussion of the operation of relational composition, this paper contains one of the first discussions of the notion of a quantifier as something distinct from Boolean connectives – see [28].) A systematic development of the calculus was carried out by Ernst Schröder, culminating in the 1895 monograph [33]. In 1903, Bertrand Russell [32] could write:

> The subject of symbolic logic is formed by three parts: the calculus of propositions, the calculus of classes, and the calculus of relations.

In fact, the famous paper [19] by Leopold Löwenheim (in which the first version of the downward Löwenheim-Skolem theorem is proved) is really a paper about the calculus of relations.

During the succeeding years, the calculus of relations went into a gradual eclipse. Löwenheim, in particular, lamented this neglect. In his opinion, the "Peirce-Schröder" calculus provided a more suitable framework for mathematics than the "Peano-Russell" system. He wrote a manuscript in which he developed these ideas. Alfred Tarski obtained a copy of the manuscript around 1938, and through his efforts it was eventually published as [20].

Tarski appears to have been dissatisfied with the approach taken by Löwenheim and his predecessors to the calculus of relations. In 1941, he published a paper [34] in which he formulated the calculus as an abstract quasi-equational theory with finitely many axioms, similar in spirit to the abstract equational formulation of the theory of Boolean algebras that had been developed in the preceding decades. By 1942 or 1943 he had worked out an elegant equational formulation and had succeeded in showing that virtually all versions of first-order set theory and number theory – and hence all of mathematics – can be developed in it. The work of Tarski and his students and colleagues eventually led to a revival of the calculus of relations and to a general development of what today is called "algebraic logic."

The purpose of this paper is to give a brief presentation of the calculus of relations as a foundation for mathematics and to discuss the logical strength of the calculus in comparison to other, well-known systems of logic. A detailed exposition of these results may be found in Chapters 1–5 of [36]. It is our hope that this paper will make some of the results of [36] accessible to a broad audience, including computer scientists (in particular, researchers in automated reasoning), logicians, mathematicians, and philosophers interested in logic and the foundations of mathematics.

Tarski's formulation of the calculus of relations as a logic $\mathcal{L}^{\times}$ suitable for developing mathematics is quite natural and easy to describe. There are no quantifiers, sentential connectives, or (in the present context) variables. There is one logical binary predicate denoting the identity relation and one non-logical binary predicate denoting a membership, or epsilon, relation from set theory. More complex predicates are built up from these basic ones by an inductive procedure, using the Boolean and Peircean operators. Every assertion of the language has the form of an equation between predicates. There are 10 simple and natural logical axiom schemata and one rule of inference: the basic high school rule of replacing equals by equals.

Despite its simplicity, the logic $\mathcal{L}^{\times}$ has considerable deductive power. It is not equivalent to first-order logic, but it is equivalent to a fragment thereof. In fact, its expressive and deductive powers are equivalent to those of a version of first-order logic with just three variables. Logic $\mathcal{L}^{\times}$ *does* become equivalent to all of first-order logic under the assumption of a certain type of equation Q_{AB}. The validity of such an equation is derivable in practically all systems of set theory. Consequently, almost all systems of set theory – and hence all of classical mathematics – can be developed in $\mathcal{L}^{\times}$. At the end of the paper, it will be shown that $\mathcal{L}^{\times}$ is also equivalent to a certain fragment of sentential logic. Thus, a fragment of sentential logic provides an adequate basis for the development of all of classical mathematics.

2 First-Order Logic

We wish to show that the calculus of relations provides an adequate framework for the formalization of mathematics. From a foundational point of view, mathematics is most often developed in a system of set theory. The language usually employed for this purpose is a first-order logic $\mathcal{L}$. It has one logical binary predicate intended to denote the relation of equality. There is also a non-logical binary predicate intended to denote the membership, or *epsilon*, relation from set theory – the relation of one set being a member of another. These predicates will be denoted in this paper by

$$\overset{\circ}{1} \qquad \text{and} \qquad E$$

respectively. It suffices to assume that $\mathcal{L}$ has just two sentential connectives, implication and negation, and one quantifier, the universal quantifier. We denote them by

$$\rightarrow \qquad , \qquad \neg \qquad , \qquad \forall$$

respectively. There is also an infinite supply of variables $v_0, v_1, v_2, \ldots$.

The atomic formulas of $\mathcal{L}$ are expressions of the form

$$v_i \overset{\circ}{1} v_j \qquad \text{and} \qquad v_i E v_j .$$

Compound formulas are built up from atomic ones in a recursive fashion: if X and Y are formulas of $\mathcal{L}$, then so are

$$X \rightarrow Y \qquad , \qquad \neg X \qquad , \qquad \forall v_i X$$

(and every formula of $\mathcal{L}$ is built up from atomic formulas in this fashion). Sentences are formulas without free variables. Other sentential connectives such as conjunction (denoted by $\wedge$), disjunction (denoted by $\vee$), and equivalence (denoted by $\leftrightarrow$), and the existential quantifier (denoted by $\exists$), are defined in the usual way. For instance,

$$X \vee Y \quad \text{abbreviates} \quad \neg X \rightarrow Y \qquad \text{and} \qquad \exists v_i X \quad \text{abbreviates} \quad \neg \forall v_i \neg X .$$

The set of all sentences of $\mathcal{L}$ will be denoted by Σ.

First-order logic has a finite number of logical axiom schemata and rules of inference. The particular choice of schemata and rules is not of paramount importance. However, various proofs are substantially simplified if one chooses a set of schemata that does not involve the notion of substitution and that permits *modus ponens* to be used as the only rule of inference. The following axiomatization is taken from [35]. It consists of nine schemata. The metavariables X, Y, and Z range over arbitrary formulas of $\mathcal{L}$, and the notation $[\![X]\!]$ denotes the universal closure of a formula X (with the free variables of X universally quantified in order of increasing size of the indices).

$$[\![(X \to Y) \to ((Y \to Z) \to (X \to Z))]\!]. \tag{A1}$$

$$[\![(\neg X \to X) \to X]\!]. \tag{A2}$$

$$[\![X \to (\neg X \to Y)]\!]. \tag{A3}$$

$$[\![\forall v_i \forall v_j X \to \forall v_j \forall v_i X]\!]. \tag{A4}$$

$$[\![\forall v_i (X \to Y) \to (\forall v_i X \to \forall v_i Y)]\!]. \tag{A5}$$

$$[\![\forall v_i X \to X]\!]. \tag{A6}$$

$$[\![X \to \forall v_i X]\!] \quad \text{where } v_i \text{ does not occur free in } X. \tag{A7}$$

$$[\![\neg \forall v_i \neg (v_i \mathring{1} v_j)]\!] \quad \text{where } i \neq j. \tag{A8}$$

$$[\![v_i \mathring{1} v_j \to (X \to Y)]\!] \quad \text{where } X \text{ is any atomic formula in which } v_i \text{ occurs, and } Y \text{ is obtained from } X \text{ by replacing a single occurrence of } v_i \text{ with } v_j. \tag{A9}$$

The only rule of inference needed with this axiomatization (in order to prove the Completeness Theorem for $\mathcal{L}$) is *modus ponens*, or *detachment*: from two sentences

$$X \qquad \text{and} \qquad X \to Y,$$

infer the sentence Y. A *derivation* of a sentence X from a set of sentences Ψ in $\mathcal{L}$ is a finite sequence of sentences such that the last sentence in the sequence is X and such that each sentence in the sequence is either (a) an instance of a logical axiom schema, or (b) a sentence in Ψ, or else (c) a consequence of two earlier sentences in the sequence, obtained by an application of the rule of detachment. The notation

$$\Psi \vdash X$$

is used to express the derivability of X from Ψ in $\mathcal{L}$. When Ψ is empty, we write

$$\vdash X$$

and say that X is *logically derivable* in $\mathcal{L}$.

A system of set theory $\mathcal{S}$ formalized in $\mathcal{L}$ is completely determined by specifying its set of (non-logical) axioms. For instance, Zermelo-Fraenkel set theory is the system $\mathcal{ZF}$ formalized in $\mathcal{L}$ that has the following axioms: extensionality, empty set, pairing, union, power set, infinity, choice, regularity (or well-foundedness), subset (or *Aussonderung*), and replacement – see, for instance [8]. (The last two are axiom schemata.)

3 Semantics of $\mathcal{L}$

The *models* of the first-order logic $\mathcal{L}$ are structures

$$\mathfrak{U} = (U\,,\,\epsilon),$$

where U is a nonempty set – the universe of discourse – and ϵ a binary relation on U. The well-known definition (due to Tarski) of truth, for sentences in a model, is based on the notion of satisfaction. The latter is a relation between infinite sequences

$$x = (x_0, x_1, x_2, \ldots)$$

of elements from a model $\mathfrak{U}$ and formulas of $\mathcal{L}$. It is defined by recursion on the definition of formulas. A sequence x *satisfies* an atomic formula $v_i A v_j$ (in the model $\mathfrak{U}$) just in case $x_i = x_j$, respectively, $x_i \epsilon x_j$, where A is the identity predicate, respectively, the membership predicate. The sequence satisfies $\neg X$ just in case it does not satisfy X, and it satisfies $X \to Y$ just in case it satisfies Y whenever it satisfies X. Finally, the sequence x satisfies $\forall v_i X$ just in case the formula X is satisfied by all sequences obtained from x by replacing x_i with an arbitrary element of $\mathfrak{U}$. Formulas have only finitely many variables, so the satisfiability of a formula X by a sequence x depends only on a finite subsequence of x, namely, the subsequence corresponding to the sequence of free variables that occur in X.

A formula is *true in a model* just in case it is satisfied by every infinite sequence of elements from the model. In particular, a sentence is true in a model just in case it is satisfied by some (and hence all) sequences. A sentence is *true* if it is true in all models. A sentence X is a *semantical consequence* of a set of sentences Ψ if X is true in a model whenever each sentence of Ψ is true in that model. The Completeness Theorem for first-order logic (due to Kurt Gödel) says that the semantic notion of consequence and the syntactic (proof-theoretic) notion of derivability coincide: X is a semantical consequence of Ψ just in case X is derivable from Ψ in $\mathcal{L}$. The implication from right to left – that derivability implies semantical consequence – is often referred to as *soundness*.

4 The Logic of the Calculus of Relations

The logic $\mathcal{L}^\times$ of the calculus of relations has no quantifiers or sentential connectives. In the context of formalizing set theory, variables are also unnecessary. (There are more general versions of the calculus with variables ranging over a universe of binary relations – see Chapter 8 of [36].) There are two atomic (that is, basic) predicates, the (logical) identity predicate and the (non-logical) epsilon predicate from $\mathcal{L}$. Compound predicates (or predicate terms) are built up from the atomic predicates by using the operators

$$+ \qquad , \qquad - \qquad , \qquad \odot \qquad , \qquad \smile \qquad ,$$

which are intended to denote, respectively, the operations of union, complementation, relational composition, and conversion on binary relations over the universe of discourse. In more detail,

$$\mathring{1} \qquad \text{and} \qquad E$$

are predicates of $\mathcal{L}^{\times}$. If A and B are predicates of $\mathcal{L}^{\times}$, then so are

$$A + B \qquad , \qquad A^{-} \qquad , \qquad A \odot B \qquad , \qquad A^{\smile}$$

(and every predicate of $\mathcal{L}^{\times}$ is built up from the two atomic predicates in this fashion).

There is one more symbol in the language, an equality symbol

$$\overset{\circ}{=}$$

intended to denote the relation of equality between binary relations over the universe of discourse. (The symbol is not to be confused with the atomic identity predicate, which is intended to denote the identity relation between *individuals* of the universe of discourse. Nor should it be confused with the normal equality symbol "$=$", which is used in the metalanguage to denote the identity of two objects.) The sentences of the language are just the equations

$$A \overset{\circ}{=} B,$$

where A and B are predicates. The set of equations will be denoted by $\Sigma^{\times}$.

Other operators on predicates may be defined in terms of the basic operators. For instance, the binary operators

$$\cdot \qquad \text{and} \qquad \oplus,$$

which are intended to denote Boolean multiplication and relational addition, may be defined as follows:

$$A \cdot B = (A^{-} + B^{-})^{-} \qquad \text{and} \qquad A \oplus B = (A^{-} \odot B^{-})^{-}.$$

(Notice that the equality being employed here is that of the metalanguage, not the equality symbol of $\mathcal{L}^{\times}$.) Similarly, constant predicates intended to denote the empty relation, the universal relation, and the diversity relation may be defined in terms of the atomic predicates:

$$0 = \mathring{1} \cdot \mathring{1}^{-} \qquad , \qquad 1 = \mathring{1} + \mathring{1}^{-} \qquad , \qquad \mathring{0} = \mathring{1}^{-},$$

respectively. Finally, an inequality

$$\leq$$

between predicates, intended to denote the relation of inclusion between binary relations, is defined by the stipulation

$$A \leq B \qquad \text{means} \qquad A + B \overset{\circ}{=} B.$$

This definition justifies referring to an inequality $A \leq B$ as an equation.

A word about the omission of parentheses might be helpful. In the absence of parentheses, it is assumed that unary operators (complementation and conversion) have precedence over binary operators and that multiplication operators (relational and Boolean multiplication) have precedence over addition operators (relational and Boolean addition).

5 Logical Axioms and Derivations in $\mathcal{L}^{\times}$

In Tarski's axiomatization of the calculus of relations there are ten logical axiom schemata. They conveniently fall into five groups, four of which are very well known. The schemata of the first group express the fact that interpretations of the addition and complementation operators are the addition and complementation operations of a Boolean algebra:

$$A + B \stackrel{\circ}{=} B + A, \tag{B1}$$

$$A + (B + C) \stackrel{\circ}{=} (A + B) + C, \tag{B2}$$

$$(A^{-} + B)^{-} + (A^{-} + B^{-})^{-} \stackrel{\circ}{=} A. \tag{B3}$$

The schemata of the second group say that an interpretation of the relational composition operator is a semigroup operation and the interpretation of the identity predicate is an identity element with respect to the semigroup operation:

$$A \odot (B \odot C) \stackrel{\circ}{=} (A \odot B) \odot C, \tag{B4}$$

$$A \odot \stackrel{\circ}{1} \stackrel{\circ}{=} A. \tag{B5}$$

The schemata of the third group say that an interpretation of the converse operator is an involution with respect to the semigroup operation:

$$A^{\smile\smile} \stackrel{\circ}{=} A, \tag{B6}$$

$$(A \odot B)^{\smile} \stackrel{\circ}{=} B^{\smile} \odot A^{\smile}. \tag{B7}$$

The schemata of the fourth group say that interpretations of the relational composition and converse operators distribute across addition:

$$(A + B) \odot C \stackrel{\circ}{=} (A \odot C) + (B \odot C), \tag{B8}$$

$$(A + B)^{\smile} \stackrel{\circ}{=} A^{\smile} + B^{\smile}. \tag{B9}$$

The final axiom schema is Tarski's simplified, equational form of a set of equivalences originally discovered by De Morgan:

$$[A^{\smile} \odot (A \odot B)^{-}] + B^{-} \stackrel{\circ}{=} B^{-}. \tag{B10}$$

(The formulation by De Morgan says that the three equations

$$A \odot B \leq C^{-} \qquad , \qquad A^{\smile} \odot C \leq B^{-} \qquad , \qquad C \odot B^{\smile} \leq A^{-}$$

are mutually equivalent in the sense that any one of them implies the other two.)

The only rule of inference needed with this axiomatization is the high-school rule of replacing equals by equals: from two equations

$$A \stackrel{\circ}{=} B \qquad \text{and} \qquad C \stackrel{\circ}{=} D$$

infer the equation that is obtained by replacing some occurrence of the predicate C in the equation $A \stackrel{\circ}{=} B$ with the predicate D. A *derivation* in $\mathcal{L}^{\times}$ of an equation X from a set of equations (hypotheses) Ψ is a sequence of equations such that the final equation in the sequence is X and such that each equation in the sequence is either (a) an instance of a logical axiom schema, or (b) an equation in Ψ, or else (c) a consequence of two earlier equations in the sequence, obtained by an application of the rule of replacement. The notations

$$\Psi \vdash^{\times} X \qquad \text{and} \qquad \vdash^{\times} X$$

are used to express the derivability of X from Ψ and the derivability of X from the empty set, in $\mathcal{L}^{\times}$. In the latter case we say that X is *logically derivable* in $\mathcal{L}^{\times}$.

To give the reader a sense of what derivations in $\mathcal{L}^{\times}$ look like, we present an example. It will be shown that

$$\vdash^{\times} \mathring{1}^{\smile} \mathrel{\mathring{=}} \mathring{1} \qquad \text{and} \qquad \vdash^{\times} \mathring{1} \odot A \mathrel{\mathring{=}} A$$

for all predicates A. Start with an arbitrary predicate A.

$$(A^{\smile} \odot \mathring{1})^{\smile} \mathrel{\mathring{=}} \mathring{1}^{\smile} \odot A^{\smile\smile}. \tag{1*}$$

$$A^{\smile} \odot \mathring{1} \mathrel{\mathring{=}} A^{\smile}. \tag{2*}$$

$$A^{\smile\smile} \mathrel{\mathring{=}} A. \tag{3*}$$

$$A^{\smile\smile} \mathrel{\mathring{=}} \mathring{1}^{\smile} \odot A^{\smile\smile}. \tag{4*}$$

$$A \mathrel{\mathring{=}} \mathring{1}^{\smile} \odot A^{\smile\smile}. \tag{5*}$$

$$A \mathrel{\mathring{=}} \mathring{1}^{\smile} \odot A. \tag{6*}$$

$$A \mathrel{\mathring{=}} A. \tag{7*}$$

$$\mathring{1}^{\smile} \odot A \mathrel{\mathring{=}} A. \tag{8*}$$

Equations (1*)–(3*) are instances of axiom schemata (B7), (B5), and (B6), respectively. The rule of replacement is used to obtain equation (4*) from (1*) and (2*), equation (5*) from (4*) and (3*), equation (6*) from (5*) and (3*), equation (7*) from (3*) and (3*), and equation (8*) from (7*) and (6*).

A derivation of the equation $\mathring{1}^{\smile} \mathrel{\mathring{=}} \mathring{1}$ is the sequence of the following 10 equations. The first eight are (1*)–(8*) with the identity predicate in place of A. Thus, (8*) takes the form

$$\mathring{1}^{\smile} \odot \mathring{1} \mathrel{\mathring{=}} \mathring{1}.$$

The last two equations in the sequence are

$$\mathring{1}^{\smile} \odot \mathring{1} \mathrel{\mathring{=}} \mathring{1}^{\smile}, \tag{9*}$$

$$\mathring{1}^{\smile} \mathrel{\mathring{=}} \mathring{1} \tag{10*}$$

Equation (9*) is an instance of axiom schema (B5), while (10*) follows from (8*) and (9*) by replacement.

A derivation of the equation $\mathring{1} \odot A \mathrel{\mathring{=}} A$ is the sequence of the following 19 equations. The first 10 are (1*)–(8*), with A replaced by the identity predicate, followed by (9*) and (10*). The next eight equations – call them (11*)–(18*) – are (1*)–(8*) as written. The last equation is

$$\mathring{1} \odot A \mathrel{\mathring{=}} A. \tag{19*}$$

It follows from (18*) (equation (8*)) and (10*) by replacement.

All of the standard Boolean laws are logically derivable from (B1)–(B3) in $\mathcal{L}^\times$. (See [17] and [18].) In particular, each equation $C \mathrel{\mathring{=}} D$ is provably equivalent to the equation

$$C \cdot D + C^- \cdot D^- \mathrel{\mathring{=}} 1$$

(in the sense that each can be derived from the other in $\mathcal{L}^\times$). Also, a finite set of equations

$$\{A_0 \mathrel{\mathring{=}} 1, \dots, A_{n-1} \mathrel{\mathring{=}} 1\}$$

is provably equivalent to the single equation

$$(A_0 \cdot \ \dots \ \cdot A_{n-1}) \mathrel{\mathring{=}} 1$$

in $\mathcal{L}^\times$.

Somewhat surprisingly, a form of the standard deduction theorem from first-order logic is valid for the logic $\mathcal{L}^\times$. The deduction theorem says that a sentence Y is derivable from a set of sentences Ψ and finitely many additional hypotheses $X_0, \dots, X_{n-1}$ just in case the single sentence

$$(X_0 \wedge \dots \wedge X_{n-1}) \to Y$$

is derivable from Ψ alone (without additional hypotheses). There are several versions of this theorem that are valid for $\mathcal{L}^\times$ (see Section 3.3 in [36]), but the following one is most suited to our purposes.

Theorem 1 (Deduction Theorem) *Suppose $A_0, \dots, A_{n-1}, B$ are predicates and Ψ is a set of equations of $\mathcal{L}^\times$. Then*

$$\Psi \cup \{A_0 \mathrel{\mathring{=}} 1, \dots, A_{n-1} \mathrel{\mathring{=}} 1\} \vdash^\times B \mathrel{\mathring{=}} 1$$

if and only if

$$\Psi \vdash^\times [1 \odot (A_0 \cdot \ \dots \ \cdot A_{n-1})^- \odot 1] + B \mathrel{\mathring{=}} 1.$$

Since a finite set of equations is provably equivalent to a single equation, it suffices to prove the theorem for the case when $n = 1$. The proof of the implication from right to left in this case is rather easy, using the rule of replacement. The proof of the reverse implication proceeds by induction on the definition of derivability in $\mathcal{L}^\times$. One shows, for every equation $C \mathrel{\mathring{=}} D$ in a derivation from the set of premises

$$\Psi \cup \{A_0 \mathrel{\mathring{=}} 1\},$$

that the equation

$$1 \odot {A_0}^- \odot 1 + (C \cdot D + C^- \cdot D^-) \mathrel{\mathring{=}} 1$$

is derivable from Ψ. More details may be found in [36].

6 Semantics of $\mathcal{L}^{\times}$

The *models* of the logic $\mathcal{L}^{\times}$ are the same as those of $\mathcal{L}$: they are structures $\mathfrak{U}$ consisting of a non-empty universe U and a binary relation ϵ on U. The definition of truth, for equations of $\mathcal{L}^{\times}$ in a model, is based on the notion of the denotation of a predicate in a model. The definition of the latter proceeds by recursion on the definition of a predicate. The identity predicate and E denote (in $\mathfrak{U}$) the identity relation on U and the relation ϵ, respectively. If predicates A and B denote binary relations R and S on U, then the predicate $A + B$ denotes the union of R and S,

$$R \cup S = \{(x, y) : (x, y) \text{ is in } R \text{ or in } S\},$$

the predicate A^{-} denotes the complement of R,

$$\sim R = \{(x, y) : x, y \text{ are in } U \text{ and } (x, y) \text{ is not in } R\},$$

the predicate $A \odot B$ denotes the relational composition of R and S,

$$R \,|\, S = \{(x, y) : \text{ there exists a } z \text{ such that } (x, z) \text{ is in } R \text{ and } (z, y) \text{ is in } S\},$$

and the predicate $A^{\smile}$ denotes the converse of R,

$$R^{-1} = \{(y, x) : (x, y) \text{ is in } R\}.$$

For future reference, notice that the predicate $A \oplus B$ denotes the relational sum of R and S,

$$R \dagger S = \{(x, y) : \text{ for all } z, \text{ either } (x, z) \text{ is in } R \text{ or } (z, y) \text{ is in } S\}.$$

An equation $A \mathrel{\mathring{=}} B$ is said to be *true*, or *valid*, in a model if the predicates A and B denote the same binary relation in the model. The equation is said to be *true* if it is true in all models. An equation X is a *semantical consequence* of a set of equations Ψ if X is true in a model whenever the equations of Ψ are true in that model. It is easy to show, by induction on the definition of derivability, that the logic $\mathcal{L}^{\times}$ is *sound*: an equation X is a semantical consequence of a set of equations Ψ whenever it is derivable from Ψ in $\mathcal{L}^{\times}$. The reverse implication fails, as we shall see in a moment.

7 A Common Extension of $\mathcal{L}$ and $\mathcal{L}^{\times}$

At first glance, the logics $\mathcal{L}$ and $\mathcal{L}^{\times}$ seem to have nothing in common. They have very different notions of formula, different rules of inference, and different notions of derivation. So it might seem hard to compare the two. Yet, there is a sense in which $\mathcal{L}^{\times}$ may be viewed as a fragment of $\mathcal{L}$. To make this statement precise, it is helpful to introduce an auxiliary logic $\mathcal{L}^{+}$ that proves to be a common extension of $\mathcal{L}$ and $\mathcal{L}^{\times}$. It turns out that $\mathcal{L}^{+}$ is a kind of definitional extension of $\mathcal{L}$ and therefore equivalent to it in means of expression and proof, while $\mathcal{L}^{\times}$ is a sublogic of $\mathcal{L}^{+}$.

The set of symbols of $\mathcal{L}^{+}$ is the union of the sets of symbols of $\mathcal{L}$ and $\mathcal{L}^{\times}$. Predicates are defined in $\mathcal{L}^{+}$ just as they are in $\mathcal{L}^{\times}$, in terms of the atomic identity and epsilon predicates and the operators on predicates. Atomic formulas of $\mathcal{L}^+$ are expressions of the form

$$v_i A v_j \qquad \text{and} \qquad A \overset{\circ}{=} B,$$

where A and B are arbitrary predicates of $\mathcal{L}^+$ (and hence of $\mathcal{L}^\times$). Compound formulas of $\mathcal{L}^+$ are defined in terms of atomic formulas using negation, implication, and universal quantification, just as in $\mathcal{L}$. The set of sentences of $\mathcal{L}^+$ will be denoted by Σ^+.

The language $\mathcal{L}^+$ has 14 axiom schemata. The first nine of these are identical to schemata (A1)–(A9), except that the metavariables X, Y, and Z range over formulas of $\mathcal{L}^+$. The last five schemata provide definitions of the symbols of $\mathcal{L}^+$ that do not occur in $\mathcal{L}$: the predicate operators and the equality symbol between predicates. They are based on the definitions of the operations of union, complementation, relational composition, and conversion and on the definition of equality between binary relations.

$$\forall v_0 \forall v_1 [v_0 (A + B) v_1 \leftrightarrow (v_0 A v_1 \vee v_0 B v_1)]. \tag{D1}$$

$$\forall v_0 \forall v_1 [v_0 A^- v_1 \leftrightarrow \neg (v_0 A v_1)]. \tag{D2}$$

$$\forall v_0 \forall v_1 [v_0 (A \odot B) v_1 \leftrightarrow \exists v_2 (v_0 A v_2 \wedge v_2 B v_1)]. \tag{D3}$$

$$\forall v_0 \forall v_1 [v_0 A^\smile v_1 \leftrightarrow v_1 A v_0]. \tag{D4}$$

$$A \overset{\circ}{=} B \leftrightarrow \forall v_0 \forall v_1 [v_0 A v_1 \leftrightarrow v_0 B v_1]. \tag{D5}$$

The notion of derivability in $\mathcal{L}^+$ is defined exactly as it is in $\mathcal{L}$. In particular, detachment is used as the only rule of inference. The notations

$$\Psi \vdash^+ X \qquad \text{and} \qquad \vdash^+ X$$

will be employed to express that a sentence X is derivable from a set of premises Ψ, and from the empty set of premises, respectively, in $\mathcal{L}^+$. In the second case, we say that X is *logically derivable* in $\mathcal{L}^+$.

From axiom schemata (D1)–(D3) it is possible to derive two further schemata that elucidate the meaning of the relational addition operator $\oplus$. These schemata will be needed later.

$$\forall v_0 \forall v_1 [v_0 (A \oplus B) v_1 \leftrightarrow \forall v_2 (v_0 A v_2 \vee v_2 B v_1)]. \tag{D6}$$

$$\forall v_0 \forall v_1 [v_0 (A^- \oplus B) v_1 \leftrightarrow \forall v_2 (v_0 A v_2 \rightarrow v_2 B v_1)]. \tag{D7}$$

The models of the logic $\mathcal{L}^+$ are the same as the models of the other two logics: they consist of a nonempty universe together with a binary relation on the universe.

The logic $\mathcal{L}^+$ is a kind of definitional extension of $\mathcal{L}$. At any rate, it is clear that $\mathcal{L}$ is a *sublogic* of $\mathcal{L}^+$: the formulas of $\mathcal{L}$ are all formulas of $\mathcal{L}^+$, and every sentence derivable from a set of sentences in $\mathcal{L}$ is derivable from the same set of sentences in $\mathcal{L}^+$. Indeed, a derivation (from a set of hypotheses) in $\mathcal{L}$ is also a derivation in $\mathcal{L}^+$.

It is perhaps not immediately evident that $\mathcal{L}^\times$ is a sublogic of $\mathcal{L}^+$. Certainly, its formulas (equations between predicates) are formulas of $\mathcal{L}^+$. However, it is not obvious at first glance that every equation derivable from a set of equations in $\mathcal{L}^\times$

is derivable from the same set of equations in $\mathcal{L}^+$, since derivations in $\mathcal{L}^\times$ are not derivations in $\mathcal{L}^+$. To prove that the deductive power of $\mathcal{L}^+$ extends that of $\mathcal{L}^\times$, one can invoke the Completeness Theorem for first-order logic. Since $\mathcal{L}^+$ is essentially a definitional extension of $\mathcal{L}$, the Completeness Theorem holds for it as well (see the next section). Suppose that, in $\mathcal{L}^\times$, an equation X is derivable from a set of equations Ψ. Then X is a semantic consequence of Ψ (by the soundness of the logic $\mathcal{L}^\times$). Logics $\mathcal{L}^\times$ and $\mathcal{L}^+$ have the same models, so the sentence X is a semantic consequence of Ψ with respect to $\mathcal{L}^+$. Invoke the Completeness Theorem for $\mathcal{L}^+$ to conclude that X is derivable from Ψ in $\mathcal{L}^+$.

It has been shown that both $\mathcal{L}$ and $\mathcal{L}^\times$ are sublogics of $\mathcal{L}^+$. To show that $\mathcal{L}^\times$ is essentially a fragment of $\mathcal{L}$, we now prove that $\mathcal{L}$ and $\mathcal{L}^+$ are equivalent as logics.

8 Equivalence of $\mathcal{L}$ and $\mathcal{L}^+$

What does it mean to say that $\mathcal{L}$ and $\mathcal{L}^+$ are equivalent as logics? Intuitively, it means that they have the same expressive and deductive powers. We have already seen that $\mathcal{L}$ is a sublogic of $\mathcal{L}^+$. We want to show, conversely, that every formula of $\mathcal{L}^+$ can be translated into an equivalent formula of $\mathcal{L}$, and every derivation (from a set of hypotheses) in $\mathcal{L}^+$ can be translated into an equivalent derivation in $\mathcal{L}$.

Both tasks are accomplished by defining a translation function G mapping formulas of $\mathcal{L}^+$ to formulas of $\mathcal{L}$, with two properties. First of all, every formula X of $\mathcal{L}^+$ is provably equivalent in $\mathcal{L}^+$ to its translation $G(X)$. Second, a sentence X is derivable in $\mathcal{L}^+$ from a set of sentences Ψ if and only if the translation of X is derivable in $\mathcal{L}$ from the set of translations of sentences of Ψ.

The definition of G proceeds by recursion on predicates and formulas. The key idea is that definitional schemata (D1)–(D5) are used to eliminate the "defined notions". The notation

$$X \quad \stackrel{G}{\longmapsto} \quad Y$$

will be employed to express that G maps a formula X to a formula Y. The first six clauses of the definition specify (recursively) the value of G on atomic formulas.

$$
\begin{aligned}
v_i \mathring{1} v_j &\quad \stackrel{G}{\longmapsto} \quad v_i \mathring{1} v_j. \\
v_i E v_j &\quad \stackrel{G}{\longmapsto} \quad v_i E v_j. \\
v_i(A + B)v_j &\quad \stackrel{G}{\longmapsto} \quad G(v_i A v_j) \vee G(v_i B v_j). \\
v_i A^- v_j &\quad \stackrel{G}{\longmapsto} \quad \neg G(v_i A v_j). \\
v_i(A \odot B)v_j &\quad \stackrel{G}{\longmapsto} \quad \exists v_k[G(v_i A v_k) \wedge G(v_k B v_j)]. \\
v_i A^\smile v_j &\quad \stackrel{G}{\longmapsto} \quad G(v_j A v_i). \\
A \mathrel{\mathring{=}} B &\quad \stackrel{G}{\longmapsto} \quad \forall v_0 \forall v_1[G(v_0 A v_1) \leftrightarrow G(v_0 B v_1)].
\end{aligned}
$$

In the fifth clause, v_k is the variable with the smallest index different from i and j. The next three clauses of the definition specify the value of G on compound formulas.

$$\neg Y \quad \stackrel{G}{\longmapsto} \quad \neg G(Y).$$

$$Y \to Z \quad \stackrel{G}{\longmapsto} \quad G(Y) \to G(Z).$$

$$\forall v_i Y \quad \stackrel{G}{\longmapsto} \quad \forall v_i G(Y).$$

It is not difficult to establish the following theorem.

Theorem 2 *For every formula X of $\mathcal{L}^+$, its translation $G(X)$ is a formula of $\mathcal{L}$, and*

$$\vdash^+ [\![X \leftrightarrow G(X)]\!]. \tag{i}$$

For instance, to establish (i) for atomic formulas X, one proceeds by induction on the definition of predicates to show that

$$\vdash^+ [\![v_i C v_j \leftrightarrow G(v_i C v_j)]\!] \tag{1}$$

for all predicates C. When C is the identity or epsilon predicate, this is immediate, since the left- and right-hand sides of the equivalence (1) are exactly the same, by definition of G. To treat one other case, suppose, as the induction hypothesis, that (1) holds when C is predicate A or predicate B. Assume now that C is the predicate $A + B$. Then the universal closure of each of the following equivalences is provable in $\mathcal{L}^+$:

$$\begin{aligned} v_i(A+B)v_j &\leftrightarrow (v_i A v_j) \vee (v_i B v_j) \\ &\leftrightarrow G(v_i A v_j) \vee G(v_i B v_j) \\ &\leftrightarrow G(v_i(A+B)v_j). \end{aligned}$$

The first equivalence uses schema (D1), the second uses the induction hypothesis, and the third uses the third clause in the definition of G. Once (1) has been established, a simple proof by induction on the definition of formulas of $\mathcal{L}^+$, using the final four clauses in the definition of G, yields (i).

Theorem 2 says that $\mathcal{L}$ and $\mathcal{L}^+$ have the same expressive power: every formula of the latter logic is provably equivalent to a formula of the former. (The converse is trivial, since $\mathcal{L}$ is a sublogic of $\mathcal{L}^+$.) To prove that the two logics have the same deductive power, one must prove the following mapping theorem.

Theorem 3 (Mapping Theorem for $\mathcal{L}$ and $\mathcal{L}^+$) *Let X be any sentence and Ψ any set of sentences of $\mathcal{L}^+$. Then*

$$\Psi \vdash^+ X \qquad \textit{if and only if} \qquad \{G(Y) : Y \textit{ in } \Psi\} \vdash G(X).$$

To prove the implication from right to left, suppose $G(X)$ is derivable from

$$\{G(Y) : Y \text{ in } \Psi\} \tag{2}$$

in $\mathcal{L}$. Then $G(X)$ is derivable from (2) in $\mathcal{L}^+$, since $\mathcal{L}$ is a sublogic of $\mathcal{L}^+$. By Theorem 2, sentence X is provably equivalent to its translation $G(X)$, and each sentence Y in

Ψ is provably equivalent to $G(Y)$, in $\mathcal{L}^+$. Combine these observations to conclude that X is derivable from Ψ in $\mathcal{L}^+$.

The reverse implication of the theorem is proved by induction on the definition of derivability in $\mathcal{L}^+$. Suppose a sentence X is derivable from a set of sentences Ψ in $\mathcal{L}^+$. Then there is a sequence of sentences whose last term is X and such that each sentence in the sequence is either an instance of (A1)–(A9), or an instance of (D1)–(D5), or a sentence in Ψ, or a consequence of two earlier sentences in the sequence, obtained by an application of *modus ponens*. The translation by G of each instance of (A1)–(A9) is an instance of the same axiom (in $\mathcal{L}$), by the last three clauses in the definition of G. The translation of an instance of (D1)–(D5) is the universal closure of a tautology, and is therefore derivable in $\mathcal{L}$. For instance, the translation of an instance of schema (D4) is the sentence

$$\forall v_0 \forall v_1 [G(v_0 A^{\smile} v_1) \leftrightarrow G(v_1 A v_0)], \tag{3}$$

by the last three clauses in the definition of G. But

$$G(v_0 A^{\smile} v_1) \qquad \text{and} \qquad G(v_1 A v_0)$$

are equal, by the fifth clause in the definition of G. Consequently, (3) is the universal closure of the tautology

$$G(v_1 A v_0) \leftrightarrow G(v_1 A v_0)$$

and is therefore derivable in $\mathcal{L}$. The translation of a sentence in Ψ is obviously a sentence in (2). Finally, if sentence Z is obtained from sentences Y and $Y \rightarrow Z$ by an application of the rule of detachment, then $G(Z)$ can be obtained from $G(Y)$ and $G(Y \rightarrow Z)$ by an application of detachment, by the penultimate clause in the definition of G.

This completes the proof of Theorem 3, and hence the proof that $\mathcal{L}$ and $\mathcal{L}^+$ are equivalent in means of expression and proof. Because the Completeness Theorem holds for $\mathcal{L}$, it follows easily from Theorems 2 and 3 that the Completeness Theorem holds for the definitional extension $\mathcal{L}^+$. Indeed, suppose a sentence X is a semantical consequence of a set of premises Ψ in $\mathcal{L}^+$. Then $G(X)$ is a semantical consequence of the set of premises (2) in $\mathcal{L}^+$ and in $\mathcal{L}$, by Theorem 2. Consequently, $G(X)$ is derivable from (2) in $\mathcal{L}$, by the Completeness Theorem for $\mathcal{L}$. It follows from Theorem 3 that X is derivable from Ψ in $\mathcal{L}^+$.

The equivalence of $\mathcal{L}$ and $\mathcal{L}^+$ implies that the two logics have isomorphic lattices of theories. A *theory* in a logic is a deductively closed set of sentences. In the case of $\mathcal{L}$, for instance, a theory is a set of sentences Θ with the property that a sentence X is in Θ whenever $\Theta \vdash X$. The theories of a logic form a lattice under the partial ordering of (set-theoretic) inclusion. The meet (greatest lower bound) of two theories Θ and Δ is their intersection $\Theta \cap \Delta$, and the join (least upper bound) is the intersection of all theories that include $\Theta \cup \Delta$. The lattice has a unit and a zero: the set of all sentences (the inconsistent theory) and the set of logically derivable sentences.

A theory is said to be *consistent* if it does not coincide with the set of all sentences, *complete* if it is a maximal consistent theory in the lattice of theories, *finitely axiomatizable* if it is the set of consequences of some finite set of sentences, and *undecidable* if there is no mechanical (recursive) procedure for determining, for each sentence X of the logic, whether X is, or is not, in the theory. A theory Θ is *essentially undecidable* if each consistent theory of the logic that includes

Θ is undecidable, and *hereditarily undecidable* if each subtheory of Θ in the logic is undecidable. A property of a theory Θ is *preserved* by a mapping of theories to theories if the image of Θ under the mapping also has the property.

Corollary 4 *The correspondence*

$$\Theta \longmapsto \Theta^+ = \{X : \Theta \vdash^+ X\} \tag{ii}$$

is an isomorphism from the lattice of theories of $\mathcal{L}$ to the lattice of theories of $\mathcal{L}^+$. Its inverse is the correspondence

$$\Delta \longmapsto \Delta \cap \Sigma. \tag{iii}$$

The isomorphisms preserve the properties of consistency, *completeness*, *finite axiomatizability*, *undecidability*, *essential undecidability*, *and hereditary undecidability*.

The proof of the corollary is not difficult. First, the composition of (ii) with (iii) is the identity correspondence on theories of $\mathcal{L}$. Checking this reduces to verifying, for each sentence X and theory Θ of $\mathcal{L}$, that

$$\Theta \vdash^+ X \tag{4}$$

implies X is in Θ. By Mapping Theorem 3, the assumption of (4) implies

$$\{G(Y) : Y \text{ in } \Theta\} \vdash G(X). \tag{5}$$

Since G is the identity function on sentences of $\mathcal{L}$, it follows from (5) that X is derivable from Θ in $\mathcal{L}$, and hence (since Θ is a theory) that X is in Θ.

Second, the composition of (iii) with (ii) is the identity correspondence on theories of $\mathcal{L}^+$. Checking this reduces to verifying, for every theory Δ of $\mathcal{L}^+$, that each sentence in Δ is derivable from $\Delta \cap \Sigma$ (in $\mathcal{L}^+$). Indeed, a sentence X of $\mathcal{L}^+$ is provably equivalent to the sentence $G(X)$ of $\mathcal{L}$. If X is in Δ, then so is $G(X)$, and X is derivable from $G(X)$, by Theorem 2.

Since the composition of (ii) with (iii) is the identity correspondence, and vice versa, the two correspondences are bijections and inverses of one another. Clearly, they preserve inclusion, so they are lattice isomorphisms. It follows that they preserve the properties of consistency and completeness (these properties are defined in lattice-theoretic terms). Any set of axioms for a theory Θ of $\mathcal{L}$ is obviously a set of axioms for Θ^+. Conversely, if Ψ is a set of axioms for Θ^+, then

$$\{G(Y) : Y \text{ in } \Psi\}$$

is a set of axioms for Θ. Hence, the isomorphisms preserve the property of finite axiomatizability. Because the mapping G and the sets of sentences Σ and Σ^+ are recursive, any decision procedure for a theory Θ of $\mathcal{L}$ would lead automatically to a decision procedure for the theory Θ^+, and conversely. Hence, the isomorphisms preserve the property of undecidability. Since they preserve inclusion and undecidability, they must preserve essential and hereditary undecidability.

9 Inequivalence of $\mathcal{L}^\times$ with $\mathcal{L}$ and $\mathcal{L}^+$

It is natural to inquire whether $\mathcal{L}^\times$ is equivalent to the first-order logic $\mathcal{L}$ in means of expression and proof. In view of the results of the preceding section, this is tantamount to asking whether $\mathcal{L}^\times$ and $\mathcal{L}^+$ are equivalent. If they were, it would substantially simplify the task of formalizing mathematics in $\mathcal{L}^\times$. Unfortunately, they are not: $\mathcal{L}^\times$ is essentially poorer than $\mathcal{L}^+$ in means of both expression and proof. Indeed, it was shown by Korselt early in the last century that the first-order sentence expressing the existence of four distinct elements is not semantically equivalent to any equation of $\mathcal{L}^\times$ (see [19]). Stronger versions of Korselt's negative result were later obtained by Tarski (see [34] and Sections 3.4 and 3.5 in [36]).

The deductive inequivalence of the two logics follows from the fact that the Completeness Theorem holds for $\mathcal{L}^+$ but fails to hold for $\mathcal{L}^\times$: there are equations that are semantically true – true in all models – and therefore derivable in $\mathcal{L}^+$, but that are not derivable in $\mathcal{L}^\times$. An example is the following equation:

$$1 \odot [(E \oplus E) + (A \cdot E^-) + (E \odot E^\smile)^-] \odot 1 \mathrel{\dot{=}} 1, \tag{6}$$

where A is the predicate

$$(E \odot E) + \mathring{1} + ([(E + E^\smile)^- \odot (E + E^\smile)^-] \cdot E^{\smile -}).$$

To grasp the intuition behind this equation, consider the four equations

$$E \cdot E^\smile = \mathring{1} \quad , \quad E \odot E \leq E \quad , \quad E \odot E^\smile = 1, \tag{7}$$
$$(E + E^\smile)^- \odot (E + E^\smile)^- = E + E^\smile.$$

The first two equations express, respectively, that a relation denoted by E is reflexive, antisymmetric, and transitive; in other words it is a partial order. The third equation expresses that (in the partial order) any two elements have an upper bound, and the fourth equation expresses that two elements are comparable (in the partial order) if and only if there exists a third element to which they are both incomparable.

The conjunction of these equations is false in every model. To see this, assume for contradiction that a binary relation R satisfies all four equations. The third equation implies that R is not empty, so by the fourth equation there are elements x and y that are incomparable (in the sense that the pair (x, y) is neither in R nor in R^{-1}.) The elements x and y have a common upper bound z, by the third equation. The comparability of x and z implies the existence of an element u to which x and z are both incomparable, by the last equation. Similarly, the comparability of y and z implies the existence of an element v to which y and z are both incomparable. Since u is incomparable to x, and x to y, the elements u and y must be comparable, by the fourth equation. The pair (u, y) cannot be in R; if it were, then the (u, z) would be in R (since (y, z) is in R), contradicting the incomparability of u and z. Consequently, (y, u) is in R. A completely analogous argument shows that (x, v) is in R. Indeed, v is incomparable to y, and y to x, so the elements v and x must be comparable; the pair (v, x) cannot be in R, for then (v, z) would be in R (since (x, z) is in R), contradicting the incomparability of v and z. Finally, the elements u and v are both incomparable to z, so they must be comparable to each other, by the fourth equation. If (u, v) is in R, then so is (y, v) (since (y, u) is in R), and this contradicts the incomparability of y and v. If (v, u) is in R, then so is the pair (x, u) (since (x, v) is in R), and this contradicts the incomparability of x and u.

The preceding argument shows that the disjunction of the negations of the four equations in (7) is true in every model; (6) is just an equational version of this disjunction, so it too is valid in all models.

To prove that (6) is not derivable in $\mathcal{L}^{\times}$, one must give nonstandard interpretations of the predicate operators of $\mathcal{L}^{\times}$ in some structure so that all instances of axiom schemata (B1)–(B10) are valid in the structure, but (6) fails. Essentially, this amounts to constructing a (nonrepresentable) abstract relation algebra in which (6) fails to be true. Such an algebra was given by McKenzie in his doctoral dissertation [24] (see also [25]). The universe of the structure consists of all subsets of the four element set

$$\{a, b, c, d\}.$$

The predicate operators of addition and complementation are interpreted as the operations of set-theoretic union and complementation on the universe. The identity and epsilon predicates are interpreted as the subsets $\{a\}$ and $\{a, b\}$, respectively. The predicate operators for converse and relational composition are interpreted as operations on the universe that distribute across unions. On atoms (singletons) they are determined by the following tables.

$\smile$	
$\{a\}$	$\{a\}$
$\{b\}$	$\{c\}$
$\{c\}$	$\{b\}$
$\{d\}$	$\{d\}$

$\odot$	$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$
$\{a\}$	$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$
$\{b\}$	$\{b\}$	$\{b\}$	$\{a, b, c, d\}$	$\{b, d\}$
$\{c\}$	$\{c\}$	$\{a, b, c, d\}$	$\{c\}$	$\{c, d\}$
$\{d\}$	$\{d\}$	$\{b, d\}$	$\{c, d\}$	$\{a, b, c\}$

Straightforward computations show that the denotations of the predicates

$$E^{\smile} \quad , \quad E^{-} \quad , \quad E^{\smile -} \quad , \quad E + E^{\smile} \quad , \quad E \odot E,$$

$$E \odot E^{\smile} \quad , \quad E^{-} \odot E^{-} \quad , \quad E \oplus E \quad , \quad (E + E^{\smile})^{-},$$

$$(E \odot E^{\smile})^{-} \quad , \quad (E + E^{\smile})^{-} \odot (E + E^{\smile})^{-}$$

under this interpretation are the subsets,

$$\{a, c\} \quad , \quad \{c, d\} \quad , \quad \{b, d\} \quad , \quad \{a, b, c\} \quad , \quad \{a, b\},$$

$$\{a, b, c, d\} \quad , \quad \{a, b, c, d\} \quad , \quad \varnothing \quad , \quad \{d\},$$

$$\varnothing \quad , \quad \{a, b, c\}$$

respectively. With the help of these computations it is easy to see that predicate

$$(E \oplus E) + (A \cdot E^{-}) + (E \odot E^{\smile})^{-}$$

denotes the empty set under the interpretation, so (6) fails to be true in the structure. (Equivalently, all four equations in (7) are true in the structure.) On the other hand, similar computations show that every instance of schemata (B1)–(B10) does hold in the structure. Therefore (6) is not derivable in $\mathcal{L}^{\times}$.

It is natural to ask whether the incompleteness of $\mathcal{L}^{\times}$ is due to a deficient selection of logical axiom schemata. In other words, is it possible to add finitely many true schemata (schemata whose instances are true in all models) to the axiomatization of $\mathcal{L}^{\times}$ so that the Completeness Theorem holds for the resulting logic? Roger Maddux

[22], building on earlier work of Donald Monk [26], [27], has shown that the answer to this question is negative: in every extension of $\mathcal{L}^\times$ obtained by the addition of finitely many (true) axiom schemata there will be true equations that are not logically derivable.

10 Formalizing Set Theory in $\mathcal{L}^\times$

At first blush, the inequivalence of $\mathcal{L}$ and $\mathcal{L}^\times$ might seem an insuperable obstacle to the goal of formalizing mathematics in $\mathcal{L}^\times$. However, mathematics is almost always developed within the framework of some system of set theory (formalized in the first-order logic $\mathcal{L}$), and in set-theory there are other, non-logical axioms that may be of help in achieving our goal.

Below, we shall specify an equational property Q_{AB} that two predicates A and B may satisfy in a system of set theory. More precisely, the translation $G(Q_{AB})$ of Q_{AB} into first-order logic may, for specific predicates A and B, be derivable from the axioms of set theory. It turns out that, under the assumption of Q_{AB} or its first-order translation, the logic $\mathcal{L}^\times$ *is* equivalent to $\mathcal{L}$ and $\mathcal{L}^+$ in means of expression and proof. In other words, there is a recursive translation function K (defined in terms of the predicates A and B) that maps the sentences of $\mathcal{L}^+$ to equations of $\mathcal{L}^\times$ and possesses the following properties. First of all, each sentence X of $\mathcal{L}^+$ is provably equivalent in $\mathcal{L}^+$ to its translation $K(X)$ under the hypothesis Q_{AB}. In other words, $\mathcal{L}^\times$ and $\mathcal{L}^+$ have the same expressive power relative to the sentence Q_{AB}. Second, in $\mathcal{L}^+$ a sentence X is derivable from a set of sentences Ψ under the assumption of hypothesis Q_{AB} just in case, in $\mathcal{L}^\times$, the translation $K(X)$ is derivable from the set of translations

$$\{K(Y) : Y \text{ in } \Psi\}$$

under the assumption of Q_{AB}. In other words, $\mathcal{L}^\times$ and $\mathcal{L}^+$ have the same deductive power relative to Q_{AB}.

Tarski's original approach to formalizing set theory in $\mathcal{L}^\times$ proceeded along the lines described above. The definition of his translation function may be found in Section 4.3 of [36]. It seems that later, he developed a somewhat simpler approach in which one defines a recursive translation function M mapping sentences of $\mathcal{L}^+$ to *predicates* of $\mathcal{L}^\times$. (See the historical remark on p. 113 of [36].) The function M has the following properties. First, each sentence X and the corresponding predicate $M(X)$ have the same semantical meaning in all models of Q_{AB}. In particular, X is true in a model of Q_{AB} just in case $M(X)$ denotes the universal relation on the universe of the model, that is, just in case the equation

$$M(X) \stackrel{\circ}{=} 1 \tag{8}$$

is valid in the model. Second, a sentence X is derivable from a set of sentences Ψ under the assumption of Q_{AB} just in case (8) is derivable from the set of equations

$$\{M(Y) \stackrel{\circ}{=} 1 : Y \text{ in } \Psi\}$$

under the assumption of Q_{AB}. Once such a function M is constructed, the translation mapping K can be defined by taking $K(X)$ to be just the equation (8).

Tarski apparently never worked out any details of this second approach. However, a complete proof along these lines *was* worked out by his student Donald Monk around 1960, and a set of mimeographed notes of this proof still exists. The notes were never published and were eventually forgotten. In the mid-1970s, Roger Maddux rediscovered a version of this approach, and his version was used in [36]. It is Maddux's version that will be presented below.

11 Quasi-Projection Predicates

The equation Q_{AB} expresses two properties of predicates A and B. First of all, in each model they denote functions, say R and S. More precisely, for any elements x and y of the model, if there is an element z such that the pairs (z, x) and (z, y) are in R or in S, then x coincides with y. This property is expressed by the two equations

$$A^{\smile} \odot A \leq \mathring{1} \qquad \text{and} \qquad B^{\smile} \odot B \leq \mathring{1},$$

or, equivalently, by the equations

$$(A^{\smile} \odot A)^{-} + \mathring{1} \stackrel{\circ}{=} 1 \qquad \text{and} \qquad (B^{\smile} \odot B)^{-} + \mathring{1} \stackrel{\circ}{=} 1.$$

Second, in each model the functions R and S denoted by the predicates are left and right *quasi-projections*, respectively. More precisely, for any pair of elements x and y from the model, there is an element z such that R maps z to x, and S maps z to y. This property is expressed by the equation

$$A^{\smile} \odot B \stackrel{\circ}{=} 1.$$

The idea is that the element z "represents" the ordered pair (x, y). The mathematical content of the three equations may be expressed by the single equation

$$[(A^{\smile} \odot A)^{-} + \mathring{1}] \cdot [(B^{\smile} \odot B)^{-} + \mathring{1}] \cdot [A^{\smile} \odot B] \stackrel{\circ}{=} 1,$$

which we denote by Q_{AB}. One of the key tasks before us will be to show that, in systems of set theory formalized in $\mathcal{L}$, predicates A and B satisfying the equation Q_{AB} are in fact constructible.

A sequence of auxiliary predicates $P_0, P_1, P_2, \ldots$ may be constructed in terms of predicates A and B. The predicate P_k is defined to be

$$\underbrace{A \odot A \odot \ldots \odot A}_{k \text{ times}} \odot B. \tag{9}$$

(When $k = 0$, the predicate in (9) is to be interpreted as B.) In a model of Q_{AB}, predicates

$$P_0, \ldots, P_{n-1} \tag{10}$$

represent quasi-projection functions $R_0, \ldots, R_{n-1}$ for sequences of n elements from the universe. More precisely, each relation R_k is a function, and for every n-termed sequence $(x_0, \ldots, x_{n-1})$ of elements from the model, there is an element z that represents the sequence in the sense that R_k maps z to x_k for each index $k < n$:

$$z \xmapsto{R_0} x_0 \quad , \quad z \xmapsto{R_1} x_1 \quad , \quad \ldots \quad , \quad z \xmapsto{R_{n-1}} x_{n-1}.$$

Thus, the predicates (10) allow one to code up, or talk about, n variables using just a single variable.

The auxiliary predicates P_k just constructed were used by Tarski in his original approach to formalizing set theories in $\mathcal{L}^{\times}$. In the approach adopted here and in [36], a slightly more complicated construction is used in order to obtain predicates that denote quasi-projection functions having the entire universe of the model as their domain. Specifically, the predicate A is replaced everywhere in (9) by the predicate

$$A + [\mathring{1} \cdot (A \odot 1)^{-}], \tag{11}$$

and analogously for the predicate B. If A denotes a function R in a model of Q_{AB}, then (11) denotes the unique function obtained by adjoining to R the pairs (x, x) for elements x of model that are not in the domain of R. In what follows, P_k will be this modified predicate.

12 Translation Mapping *M*

The auxiliary predicates are used in the definition of the function M. The key idea is that the predicates allow one to code up, or talk about, any finite sequence of variables using just a single variable, and a single variable can be expressed in the calculus of relations using the predicate operator for relational composition, or its dual. The function is defined on all formulas of $\mathcal{L}^{+}$ by recursion on the definition of formulas. The notation

$$X \quad \xmapsto{M} \quad A$$

is used to express that M maps the formula X to the predicate A.

$$\begin{aligned}
v_i C v_j \quad &\xmapsto{M} \quad [\mathring{1} \cdot (P_i \odot C \odot P_j{}^{\smile})] \odot 1.\\
C \mathrel{\mathring{=}} D \quad &\xmapsto{M} \quad 0 \oplus [(\mathring{1} \cdot [P_0 \odot (C \cdot D + C^{-} \cdot D^{-}) \odot P_1{}^{\smile}]) \odot 1].\\
\neg Y \quad &\xmapsto{M} \quad M(Y)^{-}.\\
Y \rightarrow Z \quad &\xmapsto{M} \quad M(Y)^{-} + M(Z).\\
\forall v_j Y \quad &\xmapsto{M} \quad [(P_{i_0} \odot P_{i_0}{}^{\smile}) \cdot \ldots \cdot (P_{i_{n-1}} \odot P_{i_{n-1}}{}^{\smile})]^{-} \oplus M(Y).
\end{aligned}$$

Here, C and D are predicates, and Y and Z formulas, of $\mathcal{L}^{+}$, while $i_0, \ldots, i_{n-1}$ are the indices, in order of increasing size, of the free variables of the formula $\forall v_j Y$. (When $n = 0$, that is, when Y has no free variables, the right-hand side of the final clause is to be interpreted as $M(Y)$.)

To explain the intuition underlying the definition of M, fix a model in which the equation Q_{AB} is valid. Suppose the predicate P_k denotes the relation R_k in the model. On each element x of the model, the quasi-projection R_k assumes a value, which we denote by x_k. (Recall that the domain of R_k is the entire universe of the model.) For each formula X of $\mathcal{L}^+$, a pair (x, y) of elements from the model will be in the relation denoted by the predicate $M(X)$ just in case the infinite sequence

$$(x_0, x_1, x_2, \ldots) \tag{12}$$

satisfies the formula X in the model. In other words, the definition of M mimics the definition of satisfaction.

To see this more clearly, consider the case when X is an atomic formula $v_i C v_j$. A pair (x, y) is in the relation denoted by the predicate

$$[\mathring{1} \cdot (P_i \odot C \odot P_j{}^{\smile})] \odot 1$$

just in case there is an element z in the model such that the pair (x, z) is in the relation denoted by the predicate

$$\mathring{1} \cdot (P_i \odot C \odot P_j{}^{\smile}).$$

Of course this happens just in case the pair (x, x) is in the relation denoted by the predicate

$$P_i \odot C \odot P_j{}^{\smile},$$

which is, in turn, equivalent to the pair (x_i, x_j) being in the relation denoted by C. But this is just what it means for the sequence (12) to satisfy the atomic formula X in the model.

To treat one more case, suppose X is the formula $\forall v_j Y$ and the free variables of X are, in order of increasing size of the indices,

$$v_{i_0}, v_{i_1}, \ldots, v_{i_{n-1}}.$$

(Thus, Y has at most one more free variable than X, namely, v_j.) A pair (x, y) is in the relation denoted by the predicate

$$[(P_{i_0} \odot P_{i_0}{}^{\smile}) \cdot \ldots \cdot (P_{i_{n-1}} \odot P_{i_{n-1}}{}^{\smile})]^{-} \oplus M(Y)$$

just in case, for all elements z of the model, if the pair (x, z) is in the relation denoted by the predicate

$$(P_{i_0} \odot P_{i_0}{}^{\smile}) \cdot \ldots \cdot (P_{i_{n-1}} \odot P_{i_{n-1}}{}^{\smile}), \tag{13}$$

then the pair (z, y) is in the relation denoted by the predicate $M(Y)$ – see schema (D7). To say that (x, z) is in the relation denoted by (13) is to say that the function R_k maps the elements x and z to the same element for $k = i_0, i_1, \ldots, i_{n-1}$. In other words x_k and z_k are equal for these values of k. To say that the pair (z, y) is in the relation denoted by $M(Y)$ is, by induction, equivalent to saying that the sequence

$$(z_0, z_1, z_2, \ldots) \tag{14}$$

satisfies the formula Y in the model. Summarizing, the pair (x, y) is in the relation denoted by $M(X)$ just in case, for any element z of the model, if the sequences (14) and (12) agree on terms with indices corresponding to free variables of X, then (14) satisfies the formula Y in the model. By definition of the notion of satisfaction, this is just what it means for the sequence (12) to satisfy the formula X in the model.

It follows from the preceding argument that, for a sentence X of $\mathcal{L}^+$, the predicate $M(X)$ denotes the universal relation on the model just in case the sequence (12) satisfies X in the model for every element x of the model. This is just what it means for X to be true in the model. In other words, the sentence X and the equation $M(X) \mathrel{\mathring{=}} 1$ are semantically equivalent. By the Completeness Theorem for $\mathcal{L}^+$, they are also provably equivalent under the hypothesis Q_{AB}.

Theorem 5 *For each sentence X of $\mathcal{L}^+$,*

$$Q_{AB} \vdash^+ X \leftrightarrow (M(X) \mathrel{\mathring{=}} 1).$$

One small point may bother the reader: is every infinite sequence of elements from the model represented by a single element x? In general, the answer is negative. However, formulas have only finitely many variables, so it is really important only that each finite sequence

$$(y_0, y_1, \ldots, y_{n-1})$$

of elements from the model be represented by a single element x. Of course, this is the case, because of the basic quasi-projection property of the relations $R_0, R_1, \ldots, R_{n-1}$.

When X is an equation of $\mathcal{L}^\times$, the preceding theorem can be strengthened: X and its translation $M(X) \mathrel{\mathring{=}} 1$ can be derived from one another in $\mathcal{L}^\times$, under the hypothesis Q_{AB}.

Theorem 6 *Each equation X of $\mathcal{L}^\times$ is provably equivalent in $\mathcal{L}^\times$ to its translation $M(X) \mathrel{\mathring{=}} 1$, under the hypothesis Q_{AB}.*

13 Relative Equivalence of $\mathcal{L}^\times$ and $\mathcal{L}^+$

It follows from Theorem 5 that, under the assumption of Q_{AB}, the expressive power of $\mathcal{L}^\times$ is equivalent to that of $\mathcal{L}^+$. The next theorem asserts that, under the same assumption, the two logics also have equivalent deductive powers.

Theorem 7 (Mapping Theorem for $\mathcal{L}^\times$ and $\mathcal{L}^+$) *Let X be any sentence and Ψ any set of sentences of $\mathcal{L}^+$. Then*

$$\Psi \cup \{Q_{AB}\} \vdash^+ X$$

if and only if

$$\{M(Y) \mathrel{\mathring{=}} 1 : Y \in \Psi\} \cup \{Q_{AB}\} \vdash^\times M(X) \mathrel{\mathring{=}} 1.$$

The proof of the implication from right to left is not difficult. Assume the equation

$$M(X) \mathrel{\mathring{=}} 1 \tag{15}$$

is derivable from the set of equations

$$\{M(Y) \mathrel{\mathring{=}} 1 : Y \text{ in } \Psi\} \cup \{Q_{AB}\} \tag{16}$$

in the logic $\mathcal{L}^\times$. Then (15) is also derivable from (16) in the logic $\mathcal{L}^+$, since the former is a sublogic of the latter. Each sentence Y of Ψ is provably equivalent to its translation $M(Y) \mathrel{\mathring{=}} 1$, and the sentence X is provably equivalent to its translation (15), under the hypothesis Q_{AB}, by Theorem 5. The derivability of X from

$$\Psi \cup \{Q_{AB}\} \tag{17}$$

in $\mathcal{L}^+$ now follows from the derivability of (15) from (16) in $\mathcal{L}^+$.

The proof of the reverse implication of the theorem is more involved. One first establishes a lemma for each axiom schema of $\mathcal{L}^+$ to the effect that, for each instance Z of the schema, the equation

$$M(Z) \mathrel{\mathring{=}} 1 \tag{18}$$

is derivable in $\mathcal{L}^\times$ from Q_{AB}. In some instances, the proofs of the lemmas depend upon a substantial development of the proof-theoretic machinery of $\mathcal{L}^\times$. Details may be found in Section 4.4 of [36]. (It may be mentioned in passing that, in the case of Tarski's original translation function, the proofs of the corresponding lemmas tend to be more complicated still.)

The proof of the implication of the theorem from left to right now proceeds by induction on the definition of a derivation in $\mathcal{L}^+$. A derivation of a sentence X from (17) is a sequence of sentences, the last term of which is X, and such that each sentence of the sequence is either an instance of one of the logical axiom schemata (A1)–(A9) and (D1)–(D5), or else a sentence in (17), or else a consequence of two earlier sentences in the derivation, obtained by an application of *modus ponens*. One must show, for each sentence Z in such a derivation, that (18) is derivable from (16). It then follows that (15) is derivable from (16).

If Z is an instance of an axiom schema, then (18) is derivable from Q_{AB}, by the corresponding lemma for the schema. Suppose now that Z is a sentence in (17). If Z is in Ψ, then obviously (18) is derivable from (16). When Z is the equation Q_{AB} itself, invoke Theorem 6 to conclude that (18) is derivable from Q_{AB} in $\mathcal{L}^\times$.

Finally, suppose Z is obtained from two earlier sentences Y and $Y \to Z$ using the rule of detachment. By the induction hypothesis, the equations

$$M(Y) \mathrel{\mathring{=}} 1 \qquad \text{and} \qquad M(Y \to Z) \mathrel{\mathring{=}} 1 \tag{19}$$

are derivable in $\mathcal{L}^\times$ from (16). The predicate $M(Y \to Z)$ is just $M(Y)^- + M(Z)$, by the fourth clause in the definition of M. Therefore, the second equation in (19) assumes the form

$$M(Y)^- + M(Z) \mathrel{\mathring{=}} 1. \tag{20}$$

It is obvious that (18) is derivable from (20) and the first equation in (19) using Boolean algebraic laws of $\mathcal{L}^\times$ and the rule of replacement. This completes our sketch of the proof of Theorem 7.

The proof just outlined is syntactic in character. An algebraic proof that does not depend on the proof-theoretic machinery of $\mathcal{L}^\times$ is also known. (See footnotes 1* and 3* on pages 242–244 of [36].) One advantage of the syntactic proof over an algebraic one, from the point of view of this paper, is that it presents a concrete algorithm for translating proofs in first-order logic to proofs in $\mathcal{L}^\times$ (under the hypothesis of Q_{AB}). This facilitates a comparison of the first-order proofs with their translations.

The equivalence of $\mathcal{L}^+$ and $\mathcal{L}^\times$ relative to a given equation Q_{AB} implies that the lattices of theories of the two logics have isomorphic sublattices, namely, the sublattices of theories that contain the equation Q_{AB}.

Corollary 8 *The correspondence*

$$\Theta \longmapsto \{X : \Theta \vdash^{+} X\}$$

is an isomorphism from the sublattice of theories of $\mathcal{L}^{\times}$ that contain Q_{AB} to the sublattice of theories of $\mathcal{L}^{+}$ that contain Q_{AB}. Its inverse is the correspondence

$$\Delta \longmapsto \Delta \cap \Sigma^{\times}.$$

The isomorphisms preserve the properties of consistency, completeness, finite axiomatizability, undecidability, essential undecidability, and hereditary undecidability relative to Q_{AB}.

(Recall that $\Sigma^{\times}$ is the set of all equations of $\mathcal{L}^{\times}$. A theory Θ is *hereditarily undecidable relative to Q_{AB}* if every subtheory of Θ that contains Q_{AB} is undecidable.) The proof of the corollary is almost the same as the proof of Corollary 4. In a certain part of the proof of the latter, the fact that translation function G is the identity mapping on sentences of $\mathcal{L}$ is used. The translation mapping from $\mathcal{L}^{+}$ to $\mathcal{L}^{\times}$ is not the identity on equations of $\mathcal{L}^{\times}$. However, one can use Theorem 6 instead.

14 Computational Questions

The translation function M is clearly recursive: one can write a simple computer program to calculate $M(X)$. So it is natural to ask about the practical computability of $M(X)$. For instance, how does the number of bits needed to encode $M(X)$ relate to the number of bits needed to encode X? Suppose each of the basic symbols of $\mathcal{L}^{+}$, except for variables, takes the same number of bits to encode, say c bits. The variable v_n takes roughly $c + \lg n$ bits to encode, where $\lg n$ denotes the logarithm of n in base two. Therefore, the atomic formula

$$v_m E v_n$$

takes roughly $3c + \lg m + \lg n$ bits to encode. On the other hand, the number of bits needed to encode its translation

$$[\mathring{1} \cdot (P_m \odot E \odot P_n^{\smile})] \odot 1 = [\mathring{1}^{-} + (P_m \odot E \odot P_n^{\smile})^{-}]^{-} \odot (\mathring{1} + \mathring{1}^{-})$$

is approximately

$$(14 + (m + n)(|A| + 1) + 2|B|) \cdot c,$$

where $|A|$ and $|B|$ denote the number of symbols occurring in the (compound) predicates A and B, respectively (see (9)). Thus, the number of bits needed to encode the predicate $M(X)$ is, roughly, an exponential function of the number of bits needed to encode X.

From a computational point of view, this obstacle can be overcome by treating the predicates P_n as primitive symbols, just as the variables v_n are treated as primitive symbols of first-order logic. Under this assumption, a sentence that requires n bits for its encoding is translated to a predicate that requires, in the worst case, order $o(n^2)$ bits for its encoding. (It is the condition for handling quantifiers that introduces the quadratic complexity.)

A similar situation exists with respect to derivations. If a sentence X is derivable in $\mathcal{L}^+$ from a set of sentences Ψ (possibly with the help of the equation Q_{AB}), then the translation $M(X) \mathrel{\mathring{=}} 1$ is derivable in $\mathcal{L}^\times$ from the set of equations

$$\{M(Y) \mathrel{\mathring{=}} 1 : Y \text{ in } \Psi\} \cup \{Q_{AB}\},$$

by Mapping Theorem 7. However, the length of the equational derivation is not a function of the form (or shape) of X. By passing to alphabetic variants X' of X (obtained by renaming bound variables), one may increase arbitrarily the lengths of the derivations of the equations $M(X') \mathrel{\mathring{=}} 1$, at least if one uses the derivations in [36]. (See, in particular, the proof of 4.2(vi).)

This unpleasant situation may be remedied by adding a finite number of additional schemata to the list of axiom schemata for $\mathcal{L}^\times$. Specifically, one may add the equation $M(Y) \mathrel{\mathring{=}} 1$ for each instance Y of a logical axiom schema of $\mathcal{L}^+$. It is also possible to add a new rule of inference to $\mathcal{L}^\times$: from

$$M(Y) \mathrel{\mathring{=}} 1 \qquad \text{and} \qquad M(Y \to Z) \mathrel{\mathring{=}} 1$$

infer $M(Z) \mathrel{\mathring{=}} 1$. In this modified system, the translation of a derivation in $\mathcal{L}^+$ becomes a derivation in $\mathcal{L}^\times$ (under the hypothesis Q_{AB}). Therefore, the length of the derivation of $M(X) \mathrel{\mathring{=}} 1$ is the same in the modified version of $\mathcal{L}^\times$ as the length of the derivation of X in $\mathcal{L}^+$.

15 Set Theory

Suppose a system $\mathcal{S}$ of set theory is formalized in the first-order logic $\mathcal{L}$. How does one obtain a variable-free equational formalization $\mathcal{S}^\times$ of $\mathcal{S}$ in $\mathcal{L}^\times$? The first step is to construct predicates A and B so that the sentence $G(Q_{AB})$ is a consequence of the axioms of $\mathcal{S}$. Once this is done, one can take the axioms of $\mathcal{S}^\times$ to be the translations of the (non-logical) axioms of $\mathcal{S}$, together with the sentence Q_{AB}. In other words, the set of non-logical axioms of $\mathcal{S}^\times$ is just

$$\{M(Y) \mathrel{\mathring{=}} 1 : Y \text{ is an axiom of } \mathcal{S}\} \cup \{Q_{AB}\}\,.$$

It follows from Theorems 5 and 2 that every sentence X of $\mathcal{L}$ is equivalent to the equation $M(X) \mathrel{\mathring{=}} 1$ and every equation X of $\mathcal{L}^\times$ is equivalent to the first-order sentence $G(X)$, in all models of the set theory under discussion. Furthermore, it follows from Mapping Theorems 7 and 3 that a sentence X is derivable from a set of sentences Ψ in the set theory $\mathcal{S}$ just in case the equation $M(X) \mathrel{\mathring{=}} 1$ is derivable from the set of equations

$$\{M(Y) \mathrel{\mathring{=}} 1 : Y \text{ in } \Psi\}$$

in the set theory $\mathcal{S}^\times$, and, conversely, an equation X is derivable from a set of equations Ψ in $\mathcal{S}^\times$ just in case the first-order sentence $G(X)$ is derivable from the set of first-order sentences

$$\{G(Y) : Y \text{ in } \Psi\}$$

in $\mathcal{S}$. In other words, the two set theories are equivalent in both their expressive and deductive powers. Thus, $\mathcal{S}^\times$ may justifiably be regarded as a formalization of $\mathcal{S}$ in the logic $\mathcal{L}^\times$.

The actual construction of predicates A and B such that the sentence $G(Q_{AB})$ is derivable depends on the particular set theory under consideration. To take just one example, suppose the set theory is $\mathcal{ZF}$, or any other system in which the pair axiom

$$\forall v_0 \forall v_1 \exists v_2 \forall v_3 [v_3 E v_2 \leftrightarrow (v_3 \mathring{1} v_1) \vee (v_3 \mathring{1} v_2)]$$

is valid. Put

$$D = E^{\smile} \odot [E^{\smile} \cdot (E^{\smile -} \oplus \mathring{1})] \quad , \quad F = E^{\smile} \odot E^{\smile} ,$$

$$A = D \cdot (D^{-} \oplus \mathring{1}) \quad , \quad B = F \cdot [(F^{-} + A) \oplus \mathring{1}].$$

The intuition underlying the construction of predicates A and B is Kuratowski's construction of the ordered pair (x, y) as the set consisting of $\{x\}$ and $\{x, y\}$. To see this, suppose D, F, A, and B denote relations $\hat{D}$, $\hat{F}$, $\hat{A}$, and $\hat{B}$ in a model of $\mathcal{L}^{\times}$. The form of the predicate A and the definition of denotation imply that a pair (z, x) is in $\hat{A}$ just in case it is in $\hat{D}$, and whenever a pair (z, w) is also in $\hat{D}$, then x and w are equal. (See schema (D7).) In other words, $\hat{A}$ is the functional part of $\hat{D}$. Similarly, a pair (z, y) is in $\hat{B}$ just in case it is in $\hat{F}$, and whenever a pair (z, w) is in $\hat{F}$ but not it $\hat{A}$, then y and w are equal. A simple argument shows that $\hat{B}$ is also a function. Indeed, suppose pairs (z, y) and (z, w) are in $\hat{B}$. If both of them are in $\hat{A}$, then y and w coincide, by the functionality of $\hat{A}$. If one of them is not in $\hat{A}$, then y and w coincide by the explication of when a pair is in $\hat{B}$. The preceding argument can actually be carried out in $\mathcal{L}^{\times}$ to yield

$$\vdash^{\times} A^{\smile} \odot A \leq \mathring{1} \qquad \text{and} \qquad \vdash^{\times} B^{\smile} \odot B \leq \mathring{1} .$$

In other words, the only part of Q_{AB} that does not hold automatically (without any set-theoretic assumptions) for the particular predicates A and B constructed above is the equation $A^{\smile} \odot B \mathrel{\mathring{=}} 1$.

Theorem 9 *The pair axiom is equivalent in $\mathcal{L}^{+}$ to the equation*

$$A^{\smile} \odot B \mathrel{\mathring{=}} 1. \tag{iv}$$

The proof of the theorem requires a further explication of the meaning of the predicates A and B, using the definition of denotation. The form of D implies that a pair (z, x) is in $\hat{D}$ just in case there is a w such that w is an element of z (in the sense that the pair (w, z) is in the epsilon relation of the model), and x is an element of w, and whenever u is an element of w, then u equals x – in other words, w is a singleton of x in the model. (The extensionality axiom is not assumed to hold, so one cannot speak of *the* singleton of x.) Thus, a pair (z, x) is in $\hat{D}$ just in case a singleton of x is an element of z. Combine this with the observations of the preceding paragraph to conclude that a pair (z, x) is in $\hat{A}$ just in case a singleton of x is the unique singleton that is an element of z.

The form of predicate F implies that a pair (z, y) is in $\hat{F}$ just in case y is an element of some element of z. Therefore, the pair (z, y) is in $\hat{B}$ just in case y is an element of an element of z, and whenever w is an element of an element of z, and a singleton of w is not the unique singleton in z, then w and y are equal.

Summary: a pair (x, y) is in the relation denoted by $A^{\smile} \odot B$ just in case there a z such that (a) a singleton of x is the unique singleton in z, (b) y is an element of an element of z – there is a u in z such that y is in u, (c) whenever w is an element of an element of z, and a singleton of w does not coincide with the unique singleton of x in z, then w equals y.

Under the assumption that the pair axiom is true in the model, such an element z can be shown to exist for any x and y, using Kuratowski's ordered pair construction. Thus, (iv) is true in the model. Suppose now that equation (iv) is true in the model. Then for any x and y, there is a set z satisfying conditions (a)–(c). When x and y are equal, condition (a) ensures that an unordered pair of x and x – a singleton of x – exists. When x and y are distinct, the element u from condition (b) cannot be a singleton: z has only one singleton, namely that of x. Any element w in u is an element of an element of z and therefore, by clause (c), is equal to either x or y. Since u is not a singleton, it follows that u consists of precisely two elements: x and y. Thus, an unordered pair of x and y exists in this case as well. Consequently, the pair axiom is true in the model.

16 Undecidable Equational Theories in $\mathcal{L}^{\times}$

It turns out that the set of true equations of $\mathcal{L}^{\times}$ – the set of equations true in all models – is hereditarily undecidable, and the set of logically derivable equations of $\mathcal{L}^{\times}$ – the set of equations derivable from the empty set of hypotheses – is undecidable. The proof of this assertion uses Corollary 8 and the results of the preceding section.

As is well known, there are finitely axiomatizable systems of set theory that are essentially and hereditarily undecidable. One of the simplest of these is a theory Θ with just two axioms: the sentence X_0 asserting the existence of an empty set, and the sentence X_1 asserting the existence, for any sets x and y, of a set whose elements are precisely the element y itself and the elements of x. (See [38], p. 21.) It is quite easy to check that the pair axiom is derivable from these two axioms. Invoke X_1 twice: use an empty set and x to obtain a singleton of x, and then use the singleton of x together with y to obtain a pair consisting of x and y.

The theory Θ^{+} of sentences derivable from Θ in $\mathcal{L}^{+}$ is also axiomatized by X_0 and X_1. It is essentially and hereditarily undecidable, by Corollary 4. The equation Q_{AB} constructed in the previous section is in Θ^{+}, by the observations of the preceding paragraph and by Theorem 9. Consequently, this equation is also in the theory

$$\Theta^{\times} = \Theta^{+} \cap \Sigma^{\times}$$

of $\mathcal{L}^{\times}$. It follows from Corollary 8 that $\Theta^{\times}$ is a finitely axiomatizable and essentially undecidable theory in $\mathcal{L}^{\times}$. In fact, it is axiomatized by the three sentences

$$M(X_0) \mathrel{\mathring{=}} 1 \qquad , \qquad M(X_1) \mathrel{\mathring{=}} 1 \qquad , \qquad Q_{AB}.$$

Recall that a finitely axiomatizable and undecidable theory formalized in a logic with a deduction theorem is hereditarily undecidable (see [37]). Since a deduction theorem holds for $\mathcal{L}^{\times}$, by Theorem 1, we may conclude that the theory $\Theta^{\times}$ is hereditarily undecidable.

The set of true equations of $\mathcal{L}^{\times}$ is a subtheory of $\Theta^{\times}$. Indeed, the true equations are logically derivable in $\mathcal{L}^{+}$, by the Completeness Theorem for $\mathcal{L}^{+}$. Consequently, they are all in Θ^+, and therefore also in $\Theta^\times$, by definition of the latter. It follows from the hereditary undecidability of $\Theta^\times$ that the set of true equations is hereditarily undecidable. For the same reason, the set of logically derivable equations is undecidable. (This last theory has no proper subtheories in $\mathcal{L}^\times$, so it is silly to speak of hereditary undecidability in this case.)

Theorem 10 *The set of true equations of $\mathcal{L}^\times$ is hereditarily undecidable, and the set of logically derivable equations of $\mathcal{L}^\times$ is undecidable.*

17 Logical Strength of $\mathcal{L}^\times$: First-order Logic with Three Variables

As was observed earlier, the logic $\mathcal{L}^\times$ is strictly weaker than first-order logic $\mathcal{L}$ or its definitional extension $\mathcal{L}^+$. It is relative only to one of the sentences Q_{AB} that $\mathcal{L}^\times$ becomes equivalent with them. This leads to the interesting question whether there is a natural fragment of first-order logic that is equivalent to $\mathcal{L}^\times$, without making any additional (non-logical) assumptions. In other words, what is the (first-order) logical strength of $\mathcal{L}^\times$? The following answer to this question was obtained jointly by Tarski and the author.

A closer look at the translation function G provides the key. The function maps each equation of $\mathcal{L}^\times$ (which is, of course, also an equation of $\mathcal{L}^+$) to a sentence of $\mathcal{L}$ with at most three variables, namely, the three variables

$$v_0 \quad , \quad v_1 \quad , \quad v_2 . \tag{21}$$

Thus, every equation of $\mathcal{L}^\times$ is equivalent (in $\mathcal{L}^+$) to a sentence of first-order logic with at most three variables. This suggests that $\mathcal{L}^\times$ may be equivalent to a three-variable fragment $\mathcal{L}_3$ of $\mathcal{L}$. The task is to determine precisely what that fragment is.

It is natural to specify the language of $\mathcal{L}_3$ to be the language of $\mathcal{L}$, with all variables deleted except those in (21). In particular, the notion of a formula for $\mathcal{L}_3$ is defined just as for $\mathcal{L}$, except that the only variables allowed are those of (21). The set of sentences of $\mathcal{L}_3$ is denoted by Σ_3.

The more challenging task is to determine an appropriate set of axioms. At first glance, it might seem as if one could take all instances of axiom schemata (A1)–(A9) that contain only the three variables (21). However, this axiomatization proves to be inadequate: the resulting logic is strictly weaker than $\mathcal{L}^\times$. To obtain an adequate axiomatization, it is helpful to introduce a notion of strong substitutions that simultaneously change bound variables. Several possibilities are known. (See Section 3.7 in [36].) The simplest one involves transpositions. For each formula X of $\mathcal{L}_3$, let X_{ij} be the formula obtained from X by transposing v_i and v_j, that is, by simultaneously replacing every occurrence of v_i (whether free or bound) with v_j, and every occurrence of v_j with v_i.

The set of logical axioms of $\mathcal{L}_3$ consists of all instances in $\mathcal{L}_3$ of axiom schemata (A1)–(A8), together with all instances of the following two schemata. The first is a strengthened version of axiom schema (A9), while the second is a first-order version of schema (B4), the associative law for relational composition. (More precisely, the second is a version of the schema obtained from (B4) by replacing the predicate B everywhere with the predicate $B^{\smile}$.)

$$[\![v_i \overset{\circ}{1} v_j \to (X \to X_{ij})]\!], \tag{A9$'$}$$

where X is any formula, and v_i, v_j any variables, of $\mathcal{L}_3$.

$$[\![\exists v_2(X_{12} \wedge \exists v_0(Y_{12} \wedge Z)) \leftrightarrow (\exists v_2(\exists v_1(X \wedge Y_{02}) \wedge Z_{02})]\!], \tag{A10}$$

where X, Y, Z are any formulas of $\mathcal{L}_3$, each with just two free variables: v_0 and v_1.

A derivation in $\mathcal{L}_3$ of a sentence X from a set of sentences Ψ is a finite sequence of sentences such that the final sentence is X and such that each sentence in the sequence is either (a) an instance of one of the logical axiom schemata (A1)–(A8), (A9′), (A10), or (b) a sentence in Ψ, or else (c) a consequence of two earlier sentences in the sequence, obtained by an application of the rule of detachment. The notations

$$\Psi \vdash_3 X \qquad \text{and} \qquad \vdash_3 X$$

are used to express the derivability of a sentence X from a set of sentences Ψ and the derivability of X from the empty set, in $\mathcal{L}_3$. In the latter case we say that X is *logically derivable* in $\mathcal{L}_3$.

The logic $\mathcal{L}_3$ is strong enough to prove the standard metatheorems of first-order logic, including the Theorem on Alphabetic Variants and the Replacement Theorem (and this can be done without using schema (A10)). The Theorem on Alphabetic Variants says that a formula X is provably equivalent to any alphabetic variant – any formula obtained from X by renaming bound variables in an appropriate fashion. The Replacement Theorem says that if a formula X' is provably equivalent to a formula X, and if a formula Y' is obtained from a formula Y by replacing some occurrences of X with X', then Y' is provably equivalent to Y. It may be remarked, parenthetically, that the three-variable version of $\mathcal{L}$ with (A1)–(A9) as its axiom schemata is *not* strong enough to prove the Theorem on Alphabetic Variants. (See Footnote 8* on p. 66 of [36].)

Some parenthetical remarks about schema (A10) may also be helpful to the reader. The schema is needed in $\mathcal{L}_3$ in order to derive all instances of schema (B4) (see the next section); a restricted three-variable logic based solely on schemata (A1)–(A8) and (A9′) is not adequate to this task (see p. 68 of [36]). Maddux showed (in [21] and [23]) that this restricted three-variable logic is actually equivalent to a weakened version of $\mathcal{L}^{\times}$ in which only some instances of schema (B4) are admitted as axioms, namely those of the form

$$A \odot (B \odot 1) \overset{\circ}{=} (A \odot B) \odot 1.$$

He also showed (in the same work) that the presence of (A10) among the axiom schemata of $\mathcal{L}_3$ is syntactically equivalent to allowing the use of an extra (fourth) variable in the derivations of sentences in the restricted three-variable logic.

The models of $\mathcal{L}_3$ are the same as the models of the first-order logic $\mathcal{L}$: structures with a non-empty universe and a binary relation. However, in contrast to the situation for $\mathcal{L}$, the Completeness Theorem does *not* hold for $\mathcal{L}_3$, as will become apparent below.

18 A Common Extension of $\mathcal{L}^\times$ and $\mathcal{L}_3$

To compare the expressive and deductive powers of $\mathcal{L}^\times$ and $\mathcal{L}_3$, it is very helpful to construct an auxiliary logic $\mathcal{L}_3^+$ that is a common extension of the two. The construction of $\mathcal{L}_3^+$ is carried out in complete analogy with the construction of $\mathcal{L}^+$. Its set of symbols is the union of the sets of symbols of $\mathcal{L}^\times$ and $\mathcal{L}_3$. The notion of a predicate is defined in $\mathcal{L}_3^+$ exactly as it is in $\mathcal{L}^\times$. The formulas of $\mathcal{L}_3^+$ are just the formulas of $\mathcal{L}^+$ in which only the variables of (21) occur.

There are 15 logical axiom schemata for $\mathcal{L}_3^+$. The first ten are identical to schemata (A1)–(A8), (A9′), and (A10), except that the metavariables X, Y, and Z range over formulas of $\mathcal{L}_3^+$ instead of formulas of $\mathcal{L}_3$. The last five are the definitional schemata (D1)–(D5). Notice that each instance of one of the latter schemata is automatically a sentence of $\mathcal{L}_3^+$.

The notion of derivability for $\mathcal{L}_3^+$ is defined just as it is for $\mathcal{L}_3$. In particular, *modus ponens* is the only rule of inference. The notations

$$\Psi \vdash_3^+ X \qquad \text{and} \qquad \vdash_3^+ X$$

express the derivability of a sentence X from a set of sentences Ψ, and the derivability of X from the empty set, in $\mathcal{L}_3^+$. Sentences derivable from the empty set are said to be *logically derivable*. The models of $\mathcal{L}_3^+$ are the same as the models of the other logics.

The logic $\mathcal{L}_3$ is obviously a sublogic of $\mathcal{L}_3^+$: every formula of $\mathcal{L}_3$ is a formula of $\mathcal{L}_3^+$, and every derivation of a sentence X from a set of premises Ψ in $\mathcal{L}_3$ is also a derivation of X from Ψ in $\mathcal{L}_3^+$. Similarly, $\mathcal{L}^\times$ is a sublogic of $\mathcal{L}_3^+$. Clearly, every equation of $\mathcal{L}^\times$ is an equation of $\mathcal{L}_3^+$. It is also true that whenever an equation X is derivable from a set of equations Ψ in $\mathcal{L}^\times$, then X is derivable from Ψ in $\mathcal{L}_3^+$. However, one can no longer invoke the Completeness Theorem to prove this assertion.

Instead, one first establishes a series of lemmas – one for each axiom schema (B1)–(B10) of $\mathcal{L}^\times$ – to the effect that each instance of the schema is logically derivable in $\mathcal{L}_3^+$ (using definitional schemata (D1)–(D5)). For example, consider an instance of schema (B6), say,

$$A^{\smile\smile} \mathrel{\dot{=}} A. \tag{22}$$

By schema (D5), equation (22) is provably equivalent to the sentence

$$\forall v_0 \forall v_1 (v_0 A^{\smile\smile} v_1 \leftrightarrow v_0 A v_1). \tag{23}$$

Sentence (23) is derivable in $\mathcal{L}_3^+$ by two applications of (D2). (In the first application, $v_0 A^{\smile\smile} v_1$ is proved equivalent to $v_1 A^{\smile} v_0$, and in the second, $v_1 A^{\smile} v_0$ is proved equivalent to $v_0 A v_1$.) Consequently, (22) is derivable in $\mathcal{L}_3^+$. Axiom schema (A10) is needed in order to derive each instance of schema (B4) in $\mathcal{L}_3^+$.

In a similar fashion, one proves a lemma to the effect that the rule of replacement of equals by equals is a valid rule of inference in $\mathcal{L}_3^+$. The proof proceeds by induction on predicates. In addition to schemata (D1)–(D5), it uses the Theorem on Alphabetic Variants and the Replacement Theorem, both of which are provable in $\mathcal{L}_3^+$. As an example, assume the two equations

$$A \odot B \mathrel{\dot{=}} C \qquad \text{and} \qquad B \mathrel{\dot{=}} D \tag{24}$$

as hypotheses, with the goal of deriving the equation

$$A \odot D \mathrel{\mathring{=}} C. \tag{25}$$

First, invoke (D5) and (D3) (and detachment) to derive the sentences

$$\forall v_0 \forall v_1 [\exists v_2 (v_0 A v_2 \wedge v_2 B v_1) \leftrightarrow v_0 C v_1] \qquad \text{and} \qquad \forall v_0 \forall v_1 [v_0 B v_1 \leftrightarrow v_0 D v_1] \tag{26}$$

from the equations in (24). The sentence

$$\forall v_2 \forall v_1 [v_2 B v_1 \leftrightarrow v_2 D v_1] \tag{27}$$

is an alphabetic variant of the second sentence in (26), so it too is derivable in $\mathcal{L}_3^+$ from (24), by the Theorem on Alphabetic Variants. Use the first sentence in (26), sentence (27), and the Replacement Theorem to derive the sentence

$$\forall v_0 \forall v_1 [\exists v_2 (v_0 A v_2 \wedge v_2 D v_1) \leftrightarrow v_0 C v_1].$$

Finally, invoke (D5) and (D3) one more time to obtain (25).

To prove that derivability in $\mathcal{L}^\times$ implies derivability in $\mathcal{L}_3^+$, one proceeds by induction on the definition of a derivation in $\mathcal{L}^\times$. Every equation Z in a derivation in $\mathcal{L}^\times$ from a set of hypotheses Ψ is either an instance of a logical axiom schema, or an equation in Ψ, or a consequence of two earlier equations in the derivation using the rule of replacement. If Z is an instance of one of axiom schemata (B1)–(B10), then it is logically derivable in $\mathcal{L}_3^+$, by one of the series of lemmas of the penultimate paragraph. If Z is an equation in Ψ, then it is trivially derivable in $\mathcal{L}_3^+$ from Ψ. Finally, if Z follows from two earlier equations in the derivation by an application of the rule of replacement, then Z is derivable from these two equations in $\mathcal{L}_3^+$, by the lemma of the previous paragraph. Thus, each equation in the derivation is derivable from Ψ in $\mathcal{L}_3^+$.

19 The Equivalence of $\mathcal{L}_3$ and $\mathcal{L}_3^+$

The logics $\mathcal{L}_3$ and $\mathcal{L}_3^+$ are actually equivalent in means of expression and proof. This equivalence can be established using the same translation function G that was used to prove the equivalence of the logics $\mathcal{L}$ and $\mathcal{L}^+$. The first theorem says that $\mathcal{L}_3$ and $\mathcal{L}_3^+$ have the same expressive power: every formula of the latter logic is equivalent to a formula of the former. (The converse is trivial, since $\mathcal{L}_3$ is a sublogic of $\mathcal{L}_3^+$.) Its proof is identical to the proof of Theorem 2.

Theorem 11 *For every formula X of $\mathcal{L}_3^+$, its translation $G(X)$ is a formula of $\mathcal{L}_3$, and*

$$\vdash_3^+ [\![X \leftrightarrow G(X)]\!].$$

The equivalence of the two logics in means of proof is a consequence of the following mapping theorem and the fact that $\mathcal{L}_3$ is a sublogic of $\mathcal{L}_3^+$. The proof of the theorem is identical to the proof of Theorem 3.

Theorem 12 (Mapping Theorem for $\mathcal{L}_3$ and $\mathcal{L}_3^+$) *For every sentence X and every set of sentences Ψ of $\mathcal{L}_3^+$,*

$$\Psi \vdash_3^+ X \qquad \textit{if and only if} \qquad \{G(Y) : Y \textit{ in } \Psi\} \vdash_3 G(X).$$

The equivalence of $\mathcal{L}_3$ and $\mathcal{L}_3^+$ implies that the logics have isomorphic lattices of theories. The proof is identical to the proof of Corollary 4. (Recall that Σ_3 is the set of sentences of $\mathcal{L}_3$.)

Corollary 13 *The correspondence*

$$\Theta \longmapsto \{X : \Theta \vdash_3^+ X\}$$

is an isomorphism from the lattice of theories of $\mathcal{L}_3$ *to the lattice of theories of* $\mathcal{L}_3^+$. *Its inverse is the correspondence*

$$\Delta \longmapsto \Delta \cap \Sigma_3 .$$

The isomorphisms preserve the properties of consistency, completeness, finite axiomatizability, undecidability, essential undecidability, and hereditary undecidability.

20 The Equivalence of $\mathcal{L}^\times$ with $\mathcal{L}_3$ and $\mathcal{L}_3^+$

To show that $\mathcal{L}^\times$ is equivalent with $\mathcal{L}_3$ in means of expression and proof, it suffices to show that it is equivalent with $\mathcal{L}_3^+$, by the results of the preceding section. We have already seen that $\mathcal{L}^\times$ is a sublogic of $\mathcal{L}_3^+$: every equation of the former logic is an equation of the latter, and every equation derivable from a set of equations in the former logic is also derivable from the same set of equations in the latter (though the derivations themselves are different). It must now be shown that, conversely, every sentence of $\mathcal{L}_3^+$ can be translated into an equivalent equation of $\mathcal{L}^\times$ and every derivation (from a set of hypotheses) in $\mathcal{L}_3^+$ can be translated into an equivalent derivation in $\mathcal{L}^\times$. To this end, we need to define a translation function H mapping sentences of $\mathcal{L}_3^+$ to equations of $\mathcal{L}^\times$, one that preserves the meaning of the sentences and for which a mapping theorem can be established.

The definition of H and the proof of the mapping theorem have a quite complicated appearance, though the underlying intuition is not difficult to discern. The reason is that the definition involves a large number of cases. Therefore, it is helpful to clarify at the outset the main properties of H. It maps the set of formulas of $\mathcal{L}_3^+$ to the set of quantifier-free formulas of $\mathcal{L}_3^+$. Every formula X and its translation $H(X)$ have the same free variables, and the two are provably equivalent in the sense that

$$\vdash_3^+ [\![X \leftrightarrow H(X)]\!].$$

If X has three free variables, its translation $H(X)$ will be a conjunction of disjunctions of atomic formulas – and in fact a conjunction of formulas of the form

$$(v_0 A_k v_1) \vee (v_0 B_k v_2) \vee (v_1 C_k v_2), \tag{28}$$

where A_k, B_k, and C_k are predicates. If X has fewer than three free variables, its translation will be an atomic formula. In particular, when X is a sentence, its translation will be an equation of the form $C \mathrel{\mathring{=}} 1$.

The definition of H is by recursion on the definition of formulas in $\mathcal{L}_3^+$. We shall use a slightly simplified version of the definition given in Section 3.9 of [36]. Write

$$X \quad \overset{H}{\longmapsto} \quad Y$$

to express that H maps the formula X to the formula Y. On atomic formulas, H is defined by the rules

$$v_i A v_j \quad \xmapsto{H} \quad \begin{cases} v_i A v_j & \text{if } i \leq j, \\ v_j A^{\smile} v_i & \text{if } i > j, \end{cases}$$

and

$$A \mathrel{\mathring{=}} B \quad \xmapsto{H} \quad \begin{cases} A \mathrel{\mathring{=}} B & \text{if } B = 1, \\ A \cdot B + A^{-} \cdot B^{-} \mathrel{\mathring{=}} 1 & \text{if } B \neq 1. \end{cases}$$

The purpose of the final clause is to ensure that H maps every equation to an equation of the form $C \mathrel{\mathring{=}} 1$.

Suppose, now, that $H(Y)$ and $H(Z)$ have been defined. The definitions of

$$H(\neg Y) \qquad , \qquad H(Y \rightarrow Z) \qquad , \qquad H(\forall v_i Y)$$

depend of course on the forms of $H(Y)$ and $H(Z)$. Begin with the definition of

$$H(\neg Y). \tag{29}$$

There are three cases to consider. Two are easy:

$$\begin{aligned} Y \quad &\xmapsto{H} \quad A \mathrel{\mathring{=}} 1 \quad \text{implies} \quad \neg Y \quad \xmapsto{H} \quad 1 \odot A^{-} \odot 1 \mathrel{\mathring{=}} 1, \\ Y \quad &\xmapsto{H} \quad v_i A v_j \quad \text{implies} \quad \neg Y \quad \xmapsto{H} \quad v_i A^{-} v_j. \end{aligned}$$

The case when formula $H(Y)$ has three free variables is more involved. This formula is, by the induction hypothesis, a conjunction of formulas of the form (28). Suppose, for simplicity, that there are just two conjuncts, so that $H(Y)$ has the form

$$(v_0 A_0 v_1 \vee v_0 B_0 v_2 \vee v_1 C_0 v_2) \wedge (v_0 A_1 v_1 \vee v_0 B_1 v_2 \vee v_1 C_1 v_2). \tag{30}$$

Form the negation of (30) – the formula $\neg H(Y)$ – and transform it step by step into a series of provably equivalent formulas as follows. First, distribute the negation across the conjunction and the disjunctions to obtain a disjunction of conjunctions of negations of atomic formulas. The negations may be absorbed into the predicates as complements, just as was done in a preceding clause of the definition, to arrive at

$$(v_0 A_0{}^{-} v_1 \wedge v_0 B_0{}^{-} v_2 \wedge v_1 C_0{}^{-} v_2) \vee (v_0 A_1{}^{-} v_1 \wedge v_0 B_1{}^{-} v_2 \wedge v_1 C_1{}^{-} v_2). \tag{31}$$

Next, distribute the disjunction over the conjunctions to write (31) as the conjunction of nine formulas, each of which is a disjunction of two atomic formulas:

$$\begin{aligned} &(v_0 A_0{}^{-} v_1 \vee v_0 A_1{}^{-} v_1) \wedge (v_0 A_0{}^{-} v_1 \vee v_0 B_1{}^{-} v_2) \wedge (v_0 A_0{}^{-} v_1 \vee v_1 C_1{}^{-} v_2) \\ &\qquad \wedge \ldots \wedge (v_1 C_0{}^{-} v_2 \vee v_1 C_1{}^{-} v_2). \end{aligned} \tag{32}$$

Finally, combine disjuncts with the same variables by forming the sum of the predicates, and adjoin a vacuously false atomic formula to fill in any missing disjuncts, so that the result is a conjunction of nine formulas of the form (28)

$$\begin{aligned}(v_0(A_0{}^- + A_1{}^-)v_1 \vee v_0 0 v_2 \vee v_1 0 v_2) \wedge (v_0 A_0{}^- v_1 \vee v_0 B_1{}^- v_2 \vee v_1 0 v_2)\\ \wedge (v_0 A_0{}^- v_1 \vee v_0 0 v_2 \vee v_1 C_1{}^- v_2)\\ \wedge \ldots \wedge (v_0 0 v_1 \vee v_0 0 v_2 \vee v_1(C_0{}^- + C_1{}^-)v_2). \quad (33)\end{aligned}$$

This last formula is taken as the definition of (29) in this case. Observe that the equivalence of the negation of (30) with (31), of (31) with (32), and of (32) with (33) are all provable in $\mathcal{L}_3^+$, using well-known laws of sentential logic (all of which hold in $\mathcal{L}_3^+$), the definitional schemata (D1) and (D2), and the Replacement Theorem. Thus, the formula $\neg H(Y)$ is provably equivalent to (29) in $\mathcal{L}_3^+$.

Defining

$$H(Y \to Z) \tag{34}$$

is tantamount to defining the value of H on the formula $\neg Y \vee Z$. The value (29) on the first disjunct has just been defined, so essentially it is only necessary to define H on disjunctions. Of course, disjunction is not a primitive sentential connective of $\mathcal{L}_3^+$, so this remark is only heuristic. Keeping the heuristic in mind, however, we first form (29), as above. The definition of (34) depends on the particular forms of (29) and $H(Z)$, so there are many cases to consider. We shall illustrate some of them without attempting to give an exhaustive definition. For the first four cases, assume that

$$Z \quad \stackrel{H}{\longmapsto} \quad v_0 A_1 v_1 \vee v_0 B_1 v_2 \vee v_1 C_1 v_2. \tag{35}$$

If

$$\neg Y \quad \stackrel{H}{\longmapsto} \quad v_0 A_0 v_1 \vee v_0 B_0 v_2 \vee v_1 C_0 v_2,$$

then

$$Y \to Z \quad \stackrel{H}{\longmapsto} \quad v_0(A_0 + A_1)v_1 \vee v_0(B_0 + B_1)v_2 \vee v_1(C_0 + C_1)v_2.$$

If

$$\neg Y \quad \stackrel{H}{\longmapsto} \quad v_0 B_0 v_2,$$

then

$$Y \to Z \quad \stackrel{H}{\longmapsto} \quad v_0 A_1 v_1 \vee v_0(B_0 + B_1)v_2 \vee v_1 C_1 v_2.$$

If

$$\neg Y \quad \stackrel{H}{\longmapsto} \quad v_1 B_0 v_1,$$

then

$$Y \to Z \quad \stackrel{H}{\longmapsto} \quad v_0[1 \odot (B_0 \cdot \mathring{1}) + A_1]v_1 \vee v_0 B_1 v_2 \vee v_1 C_1 v_2.$$

Finally, if

$$\neg Y \quad \overset{H}{\longmapsto} \quad A_0 \mathrel{\mathring{=}} 1,$$

then

$$Y \to Z \quad \overset{H}{\longmapsto} \quad v_0([0 \oplus A_0 \oplus 0] + A_1)v_1 \vee v_0 B_1 v_2 \vee v_1 C_1 v_2.$$

The above examples illustrate the definition of (34) under the assumption that (35) holds. Two further examples illustrate the definition under different assumptions. If

$$\neg Y \quad \overset{H}{\longmapsto} \quad v_1 A v_1 \qquad \text{and} \qquad Z \quad \overset{H}{\longmapsto} \quad v_0 B v_2,$$

then

$$Y \to Z \quad \overset{H}{\longmapsto} \quad v_0[1 \odot (A \cdot \mathring{1})]v_1 \vee v_0 B v_2 \vee v_1 0 v_2,$$

and if

$$\neg Y \quad \overset{H}{\longmapsto} \quad v_2 A v_2 \qquad \text{and} \qquad Z \quad \overset{H}{\longmapsto} \quad v_0 B v_2,$$

then

$$Y \to Z \quad \overset{H}{\longmapsto} \quad v_0[1 \odot (A \cdot \mathring{1}) + B]v_2.$$

It is not difficult to show that formula (34) is provably equivalent in $\mathcal{L}_3^+$ to the formula

$$H(Y) \to H(Z). \tag{36}$$

Consider, as an example, the case when (29) is the equation $A_0 \mathrel{\mathring{=}} 1$, and $H(Z)$ is determined by (35). In $\mathcal{L}_3^+$ the following formulas are provably equivalent, using definitional schemata (D1)–(D5) and their consequence (D6), together with the Theorem on Alphabetic Variants and the Replacement Theorem:

$$v_0[0 \oplus A_0 \oplus 0]v_1,$$

$$\forall v_2(\forall v_1[v_0 0 v_1 \vee v_1 A_0 v_2] \vee v_2 0 v_1),$$

$$\forall v_2 \forall v_1(v_1 A_0 v_2),$$

$$\forall v_0 \forall v_1(v_0 A_0 v_1),$$

$$\forall v_0 \forall v_1(v_0 A_0 v_1 \leftrightarrow v_0 1 v_1),$$

$$A_0 \mathrel{\mathring{=}} 1.$$

Consequently, (34), namely the formula

$$v_0([0 \oplus A_0 \oplus 0] + A_1)v_1 \vee v_0 B_1 v_2 \vee v_1 C_1 v_2,$$

is provably equivalent to the disjunction of the four atomic formulas

$$A_0 \mathrel{\mathring{=}} 1 \qquad , \qquad v_0 A_1 v_1 \qquad , \qquad v_0 B_1 v_2 \qquad , \qquad v_1 C_1 v_2,$$

by (D1). In view of the induction hypothesis (and the assumptions of the case we are in), $\neg H(Y)$ is provably equivalent to the first of these formulas, and $H(Z)$ is obviously equivalent to the disjunction of the remaining three. Consequently, (34) is provably equivalent to

$$\neg H(Y) \vee H(Z),$$

and hence also to (36).

To complete the definition of H, its values must be defined on formulas beginning with a universal quantifier. As an illustration, consider the definition of

$$H(\forall v_1 Y), \tag{37}$$

and treat four cases. If the variable v_1 does not occur free in Y, then (37) is defined to be $H(Y)$. Suppose v_1 does occur free in Y. The following two examples illustrate the cases when Y has one or two free variables:

$$Y \quad \stackrel{H}{\longmapsto} \quad v_1 A v_1 \quad \text{implies} \quad \forall v_1 Y \quad \stackrel{H}{\longmapsto} \quad A + \overset{\circ}{0} \mathrel{\overset{\circ}{=}} 1,$$

$$Y \quad \stackrel{H}{\longmapsto} \quad v_0 A v_1 \quad \text{implies} \quad \forall v_1 Y \quad \stackrel{H}{\longmapsto} \quad v_0 (A \oplus 0) v_0.$$

Finally, consider the case when Y has three free variables. By the induction hypotheses, the formula $H(Y)$ is a conjunction of formulas of the form (28). Since a universal quantifier distributes across conjunctions (this metatheorem is provable in $\mathcal{L}_3^+$), it suffices to treat the case when

$$Y \quad \stackrel{H}{\longmapsto} \quad v_0 A v_1 \vee v_0 B v_2 \vee v_1 C v_2.$$

Under this hypothesis, define

$$\forall v_1 Y \quad \stackrel{H}{\longmapsto} \quad v_0[(A \oplus C) + B] v_2.$$

Once again, it is not difficult to show that (37) is provably equivalent to

$$\forall v_1 H(Y) \tag{38}$$

in $\mathcal{L}_3^+$. For instance, in the case of the final clause, use schemata (D2) and (D6), together with the Theorem on Alphabetic Variants, the Replacement Theorem, and the (provable) distributivity of a universal quantifier over a disjunction in which one of the two disjuncts does not contain the quantified variable, to derive the equivalence of the following formulas in $\mathcal{L}_3^+$:

$$v_0[(A \oplus C) + B] v_2,$$

$$v_0(A \oplus C) v_2 \vee v_0 B v_2,$$

$$\forall v_1 (v_0 A v_1 \vee v_1 C v_2) \vee v_0 B v_2,$$

$$\forall v_1 (v_0 A v_1 \vee v_1 C v_2 \vee v_0 B v_2),$$

$$\forall v_1 (v_0 A v_1 \vee v_0 B v_2 \vee v_1 C v_2).$$

The first formula is just (37) (under the conditions of the final clause), while the last formula is (38). This completes the discussion of the definition of translation function H and simultaneously illustrates how one proves, by induction on formulas of $\mathcal{L}_3^+$, the following theorem.

Theorem 14 *For every formula X of $\mathcal{L}_3^+$, its translation $H(X)$ is a quantifier-free formula of $\mathcal{L}_3^+$ with the same free variables, and*

$$\vdash_3^+ [\![X \leftrightarrow H(X)]\!].$$

Every sentence of $\mathcal{L}_3^+$ is mapped to an equation of the form $C \mathrel{\mathring{=}} 1$.

It is an immediate consequence of the preceding theorem that H maps every sentence of $\mathcal{L}_3^+$ to a provably equivalent equation of $\mathcal{L}^\times$. In other words, $\mathcal{L}_3^+$ and $\mathcal{L}^\times$ are equivalent in means of expression.

A stronger form of the theorem holds for equations of $\mathcal{L}^\times$. The proof uses the validity of the standard Boolean laws in $\mathcal{L}^\times$, and is almost trivial.

Theorem 15 *Every equation X is provably equivalent in $\mathcal{L}^\times$ to its translation $H(X)$.*

To show that $\mathcal{L}^\times$ and $\mathcal{L}_3^+$ are equivalent in means of proof, one must establish the following mapping theorem.

Theorem 16 (Mapping Theorem for $\mathcal{L}^\times$ and $\mathcal{L}_3^+$) *Let X be any sentence and Ψ any set of sentences of $\mathcal{L}_3^+$. Then*

$$\Psi \vdash_3^+ X \qquad \textit{if and only if} \qquad \{H(Y) : Y \textit{ in } \Psi\} \vdash^\times H(X).$$

The proof in one direction is not difficult. Suppose that $H(X)$ is derivable from the set of hypotheses

$$\{H(Y) : Y \text{ in } \Psi\} \tag{39}$$

in $\mathcal{L}^\times$. Then $H(X)$ is derivable from (39) in $\mathcal{L}_3^+$, since $\mathcal{L}^\times$ is a sublogic of $\mathcal{L}_3^+$. Each sentence Y of Ψ is provably equivalent in $\mathcal{L}_3^+$ to $H(Y)$, and similarly for the sentence X, by Theorem 14. Consequently, X is derivable from Ψ in $\mathcal{L}_3^+$.

The proof of the reverse direction of the theorem is more complicated. First, for each logical axiom schema of $\mathcal{L}_3^+$ one proves a lemma to the effect that the translation $H(Z)$ of each instance Z of the schema is logically derivable in $\mathcal{L}^\times$. The proof of each of these lemmas is complicated by the number of formula metavariables that occur in the schema and by the number of cases to be considered for each formula metavariable. The interested reader should consult Section 3.9 of [36] for examples of such derivations.

One then proceeds by induction on the definition of derivations in $\mathcal{L}_3^+$ to show that the translation $H(Z)$ of each sentence Z in a derivation of X from Ψ is derivable from (39) in $\mathcal{L}^\times$. If Z is an instance of a logical axiom schema, then $H(Z)$ is logically derivable in $\mathcal{L}^\times$ by one of the lemmas. If Z is a sentence in Ψ, then $H(Z)$ is trivially derivable in $\mathcal{L}^\times$ from (39). Suppose, finally, that Z is inferred from two earlier sentences Y and $Y \to Z$ in the derivation using the rule of detachment. We shall show that $H(Z)$ is derivable from

$$H(Y) \qquad \text{and} \qquad H(Y \to Z) \tag{40}$$

in $\mathcal{L}^\times$. Since H maps sentences to equations of the form $C \mathrel{\mathring{=}} 1$, by Theorem 14, we may assume that

$$Y \quad \overset{H}{\longmapsto} \quad A \mathrel{\mathring{=}} 1 \qquad \text{and} \qquad Z \quad \overset{H}{\longmapsto} \quad B \mathrel{\mathring{=}} 1.$$

It follows from the definition of H that

$$\neg Y \quad \overset{H}{\longmapsto} \quad 1 \odot A^{-} \odot 1 \mathrel{\mathring{=}} 1,$$

and consequently that

$$Y \to Z \quad \overset{H}{\longmapsto} \quad [0 \oplus (1 \odot A^{-} \odot 1) \oplus 0] + B \mathrel{\mathring{=}} 1.$$

The induction hypothesis is that the equations in (40), that is, the equations

$$A \mathrel{\mathring{=}} 1 \qquad \text{and} \qquad [0 \oplus (1 \odot A^{-} \odot 1) \oplus 0] + B \mathrel{\mathring{=}} 1, \tag{41}$$

are derivable in $\mathcal{L}^{\times}$ from (39). From the first equation in (41), derive in $\mathcal{L}^{\times}$ the equivalent equations

$$1 \odot A^{-} \odot 1 \mathrel{\mathring{=}} 0 \qquad \text{and} \qquad 0 \oplus (1 \odot A^{-} \odot 1) \oplus 0 \mathrel{\mathring{=}} 0. \tag{42}$$

From the second equations in (41) and (42) derive the equation

$$B \mathrel{\mathring{=}} 1.$$

This last equation is just $H(Z)$. Thus, $H(Z)$ is derivable from (40) in $\mathcal{L}^{\times}$. This completes the outline of the proof of Mapping Theorem 16.

The equivalence of $\mathcal{L}^{\times}$ and $\mathcal{L}_3^{+}$ implies that the logics have isomorphic lattices of theories. The proof is very similar to the proof of Corollary 4 but uses Theorem 15. (See the remarks following Corollary 13.)

Corollary 17 *The correspondence*

$$\Theta \longmapsto \{X : \Theta \vdash_3^{+} X\}$$

is an isomorphism from the lattice of theories of $\mathcal{L}^{\times}$ to the lattice of theories of $\mathcal{L}_3^{+}$. Its inverse is the correspondence

$$\Delta \longmapsto \Delta \cap \Sigma^{\times}.$$

The isomorphisms preserve the properties of consistency, completeness, finite axiomatizability, undecidability, essential undecidability, and hereditary undecidability.

The equivalence of the logics $\mathcal{L}^{\times}$ and $\mathcal{L}_3$ in means of expression and proof follows at once from Theorems 11 and 14, and from Theorems 12 and 16. More precisely, every sentence X of $\mathcal{L}_3$ is provably equivalent (in $\mathcal{L}_3^{+}$) to the equation $H(X)$ of $\mathcal{L}^{\times}$, and every equation X of $\mathcal{L}^{\times}$ is provably equivalent to the sentence $G(X)$ of $\mathcal{L}_3$. Thus, $\mathcal{L}_3$ and $\mathcal{L}^{\times}$ have the same expressive power. Furthermore, every sentence X is derivable from a set of sentences Ψ in $\mathcal{L}_3$ just in case the equation $H(X)$ is derivable from the set of equations $\{H(Y) : Y \text{ in } \Psi\}$ in $\mathcal{L}^{\times}$, and, similarly, every equation X is derivable from a set of equations Ψ in $\mathcal{L}^{\times}$ just in case the sentence $G(X)$ is derivable from the set of sentences $\{G(Y) : Y \text{ in } \Phi\}$ in $\mathcal{L}_3$. In other words, $\mathcal{L}_3$ and $\mathcal{L}^{\times}$ have the same deductive power. In summary, the logical strength of $\mathcal{L}^{\times}$ is just that of a first-order logic with three variables, namely, the logic $\mathcal{L}_3$.

There is another, two-step approach to defining a translation function from $\mathcal{L}_3^{+}$ to $\mathcal{L}^{\times}$. In the first step, an auxiliary function H' mapping each formula of $\mathcal{L}_3^{+}$ to a quantifier-free formula of $\mathcal{L}_3^{+}$ – one that is *always* a conjunction of formulas of the form (28) – is defined. The definition is similar in spirit to that of H, but in the atomic and universal quantifier clauses, one adds vacuously false atomic formulas to fill in missing disjuncts. In the second step, for each sentence X an equivalent equation of $\mathcal{L}^\times$ is constructed from the translation $H'(X)$. The advantage of this approach is that it reduces substantially the number of cases that must be considered. For instance, the inductive clauses for negation and implication each involve only one case. The disadvantages of the approach are the following: (1) in general, the formula $H'(X)$ no longer has the same free variables as X; (2) the length of $H'(X)$ is usually substantially longer than the length of $H(X)$; (3) the equation equivalent to a given sentence X must be constructed in a second, separate step; (4) the intuition underlying the definition of H' is perhaps less immediate than the one underlying the definition of H. See [23] for details.

21 Consequences of the Equivalence of $\mathcal{L}^\times$ with $\mathcal{L}_3$

It has already been mentioned that the Completeness Theorem fails for the logic $\mathcal{L}^\times$. Indeed, equation (6) is true in all models but it is not derivable in $\mathcal{L}^\times$. It follows from the equivalence of $\mathcal{L}^\times$ and $\mathcal{L}_3$ that the translation of (6) under the mapping G is true in all models but is not derivable in $\mathcal{L}_3$ (or even in $\mathcal{L}_3^+$). Conclusion: the Completeness Theorem fails to hold for $\mathcal{L}_3$ (and $\mathcal{L}_3^+$). As in the case of $\mathcal{L}^\times$, this failure is not due to a deficient selection of logical axiom schemata. The equivalence of $\mathcal{L}_3$ with $\mathcal{L}^\times$ and Maddux's theorem concerning the incompleteness of finitely schematized extensions of $\mathcal{L}^\times$ imply that in every extension of $\mathcal{L}_3$ obtained by adding finitely many (true) axiom schemata, there will be true sentences that are not logically derivable. Recently, Robin Hirsch, Ian Hodkinson, and Maddux have shown (in [14] and [15]) that, for each integer $n \geq 5$, there are true sentences of $\mathcal{L}_3$ whose proof *requires n* variables (in a formalism similar to $\mathcal{L}_3$ but provided with n variables). In other words, there are three-variable sentences that are logically derivable if one uses n variables but not if one uses fewer than n variables.

In the logic $\mathcal{L}^\times$, the set of true equations is hereditarily undecidable, and the set of logically provable equations is undecidable, by Theorem 10. In view of Corollaries 17 and 13, the analogous results hold for the logics $\mathcal{L}_3^+$ and $\mathcal{L}_3$. Thus, in $\mathcal{L}_3$, the set of true sentences is hereditarily undecidable, and the set of logically provable sentences is undecidable. In particular, there are no algorithms for deciding whether first-order sentences with three variables are true or whether they are derivable in $\mathcal{L}_3$.

A third consequence of equivalence of $\mathcal{L}_3$ and $\mathcal{L}^\times$ is that virtually every system $\mathcal{S}$ of set theory formalized in first-order logic can equivalently be formalized in $\mathcal{L}_3$. Indeed, suppose Ψ is the set of non-logical (set-theoretical) axioms of $\mathcal{S}$, and assume the sentence $G(Q_{AB})$ is derivable from Ψ in $\mathcal{L}$. Correlate with $\mathcal{S}$ the system $\mathcal{S}_3$ formalized in $\mathcal{L}_3$ whose set of axioms is the set of three-variable sentences

$$\{\forall v_0 \forall v_1 G(v_0 M(Y) v_1) : Y \text{ in } \Psi\} \cup \{G(Q_{AB})\}.$$

Then $\mathcal{S}$ and $\mathcal{S}_3$ are equivalent in means of expression and proof. Conclusion: practically every system of set theory can be developed in a logic with just three variables. In other words, *three variables suffice for doing all of mathematics*. István Németi [29] has shown that set theory can even be formalized in the three-variable logic obtained from $\mathcal{L}_3$ by deleting all instances of associativity schema (A10).

22 $\mathcal{L}^\times$ as a Fragment of Sentential Logic

The equational logic $\mathcal{L}^\times$ may be simplified still further. Equations may be entirely eliminated from the language, so that predicates remain as the only valid expressions, serving in the roles of both predicates and sentences. Furthermore, an axiomatization may be given for this reduced language that is reminiscent of sentential logic, and permits a form of *modus ponens* to be used as the only rule of inference. In fact, the resulting logic proves to be a fragment of sentential logic. Consequently, a fragment of sentential logic provides an adequate basis for the development of all of mathematics.

Every equation $A \mathrel{\mathring{=}} B$ is equivalent in $\mathcal{L}^\times$ to the equation

$$A \cdot B + A^- \cdot B^- \mathrel{\mathring{=}} 1.$$

Thus, one could in principle create a new equational logic similar in structure to $\mathcal{L}^\times$ but in which every equation has the form $C \mathrel{\mathring{=}} 1$. This new logic is easily seen to be equivalent to $\mathcal{L}^\times$ in means of expression and proof. The new logic motivates another modification: dispense entirely with the equality symbol "$\mathring{=}$," and use each predicate with a double meaning: as the designation of a binary relation, and as a sentence asserting that the relation denoted by the predicate is the universal relation. Such a logic might be called the *reduced version* of $\mathcal{L}^\times$, denoted by $\mathcal{L}^\times_r$.

There is no difficulty in making this intuitive description of $\mathcal{L}^\times_r$ precise. The symbols of $\mathcal{L}^\times_r$ coincide with the symbols of $\mathcal{L}^\times$, except that the equality symbol "$\mathring{=}$" is deleted. Predicates are defined in $\mathcal{L}^\times_r$ precisely as they are in $\mathcal{L}^\times$. A formula is just a predicate. The logical axioms of $\mathcal{L}^\times_r$ are the predicates of the form

$$A \cdot B + A^- \cdot B^-,$$

where $A \mathrel{\mathring{=}} B$ is an axiom of $\mathcal{L}^\times$. For example, each predicate of the form

$$[(A \odot \mathring{1}) \cdot A] + [(A \odot \mathring{1})^- \cdot A^-]$$

is an axiom of $\mathcal{L}^\times_r$ – one could say it is the predicate form of an instance of schema (B5). A derivation in $\mathcal{L}^\times_r$ of a predicate X from a set of predicates Ψ is a sequence of predicates such that the final predicate in the sequence is X and such that each predicate in the sequence is either (a) a logical axiom, or (b) a predicate in Ψ, or else (c) a consequence of two earlier predicates

$$A \qquad \text{and} \qquad C \cdot D + C^- \cdot D^-$$

in the sequence, obtained by replacing some occurrence of the predicate C in A with the predicate D.

The models of $\mathcal{L}^\times_r$ are defined to be just the models of $\mathcal{L}^\times$: structures consisting of a nonempty universe and a binary epsilon relation. The notion of the denotation of a predicate in a model is defined exactly as it is for $\mathcal{L}^\times$. A predicate-sentence is said to be *true* in a model if it denotes the universal relation on the universe.

The definition of $\mathcal{L}^\times_r$ makes clear that it is equivalent to $\mathcal{L}^\times$ in means of expression and proof. A precise proof of this statement is simplified by constructing a common extension of the two logics. This task does not present any substantial difficulties, and we shall not stop to do it here. The reader is referred to Section 5.4 of [36].

From an aesthetic point of view, the definition of the notion of derivability in $\mathcal{L}^\times_r$ is rather inelegant. The logical axioms seem somewhat unnatural, and the rule of replacement takes a rather awkward form. There is a different way of defining derivability for $\mathcal{L}_r^{\times}$ that is more natural and in fact close in spirit to the definition of derivable for the first-order logic $\mathcal{L}$. It uses a form of *modus ponens* as the only rule of inference.

To help explain this definition, we first introduce two new operators on predicates, $\to$ and $\Rightarrow$. For any predicates A and B of $\mathcal{L}_r^{\times}$,

$$A \to B \quad \text{abbreviates} \quad 1 \odot A^{-} \odot 1 + B \tag{43}$$

and

$$A \Rightarrow B \quad \text{abbreviates} \quad A^{-} + B. \tag{44}$$

To grasp the difference in meaning between these two operators, notice that $A \to B$ and $A \Rightarrow B$ are, respectively, equivalent to the sentences

$$\forall v_0 \forall v_1 (v_0 A v_1) \to \forall v_0 \forall v_1 (v_0 B v_1) \qquad \text{and} \qquad \forall v_0 \forall v_1 (v_0 A v_1 \to v_0 B v_1)$$

of $\mathcal{L}^{+}$.

As the new set of logical axioms of $\mathcal{L}_r^{\times}$, take all instances of the following 13 schemata.

$$(A + A) \Rightarrow A. \tag{C1}$$

$$A \Rightarrow (A + B). \tag{C2}$$

$$(B \Rightarrow C) \Rightarrow [(A + B) \Rightarrow (C + A)]. \tag{C3}$$

$$[(A \odot B) \odot C] \Rightarrow [A \odot (B \odot C)]. \tag{C4}$$

$$[(A + B) \odot C] \Rightarrow [(A \odot C) + (B \odot C)]. \tag{C5}$$

$$(A^{\smile} \odot B)^{\smile} \Rightarrow (B^{\smile} \odot A). \tag{C6}$$

$$[A^{\smile} \odot (A \odot B)^{-}] \Rightarrow B^{-}. \tag{C7}$$

$$(A \odot \mathring{1}) \Rightarrow A. \tag{C8}$$

$$A \Rightarrow (A \odot \mathring{1}). \tag{C9}$$

$$(A \Rightarrow B) \to (A \to B). \tag{C10}$$

$$(A \Rightarrow B) \to [(A \odot C) \Rightarrow (B \odot C)]. \tag{C11}$$

$$(A^{\smile} \Rightarrow B) \to (A \Rightarrow B^{\smile}). \tag{C12}$$

$$(A \Rightarrow B^{\smile}) \to (A^{\smile} \Rightarrow B). \tag{C13}$$

The rule of inference used with this axiom system is a form of the rule of detachment for the operator $\to$: from predicates

$$A \qquad \text{and} \qquad A \to B$$

infer the predicate B.

The new notion of derivability for $\mathcal{L}_r^{\times}$ is defined as follows. A derivation of a predicate X from a set of predicates Ψ is a sequence of predicates such that last term of the sequence is X and such that each term in the sequence is either (a) an instance of one of the logical axiom schemata (C1)–(C13), or (b) a predicate in Ψ, or else (c) a consequence of two earlier predicates in the sequence, obtained by an application of the rule of detachment. It is not a trivial matter to show that the two notions of derivability for $\mathcal{L}_r^\times$ are equivalent, but there are no substantial difficulties involved.

As was pointed out above, the logic $\mathcal{L}_r^\times$ is equivalent to the logic $\mathcal{L}^\times$. The set of true equations and the set of logically derivable equations of $\mathcal{L}^\times$ are, respectively, hereditarily undecidable and undecidable. Therefore, the same is true of $\mathcal{L}_r^\times$: its set of true sentences and its set of logically derivable sentences are, respectively, hereditarily undecidable and undecidable.

23 Sentential Logic

To see that $\mathcal{L}_r^\times$ may be viewed as a fragment of sentential logic, we must formulate the latter within the language of $\mathcal{L}_r^\times$. We use $\mathcal{T}$ to refer to the particular form of sentential logic presented here. The language of $\mathcal{T}$ contains no sentential variables. It has two sentential constants (atomic sentences): the logical constant

$$\mathring{1} \tag{45}$$

functioning as a fixed true sentence (usually "T" is used instead of (45)), and a non-logical constant E functioning as a fixed sentence whose truth value is not determined. The language also contains four sentential connectives,

$$+ \quad , \quad \odot \quad , \quad {}^- \quad , \quad {}^\smile .$$

The first two are the binary connectives of disjunction and conjunction, respectively. They are synonyms for the usual connectives $\vee$ and $\wedge$. The last two are the unary connectives of negation and affirmation, respectively. The first is a synonym for $\neg$. The second may be thought of as expressing the phrase "it is the case that." It occurs rather infrequently in formalizations of sentential logic and is essentially superfluous.

Compound sentences of $\mathcal{T}$ are built from atomic sentences in the usual way: if A and B are sentences, then so are

$$A + B \quad , \quad A \odot B \quad , \quad A^- \quad , \quad A^\smile$$

(and every sentence is built up from atomic sentences in this fashion). Thus, the set of sentences of $\mathcal{T}$ coincides with the set of predicates of $\mathcal{L}_r^\times$ and $\mathcal{L}^\times$. The abbreviations (43) and (44) may be used in $\mathcal{T}$ as well. In addition, we introduce two more abbreviations:

$$A \cdot B \qquad \text{abbreviates} \qquad (A^- + B^-)^-$$

and

$$A \Leftrightarrow B \qquad \text{abbreviates} \qquad (A \Rightarrow B) \cdot (B \Rightarrow A),$$

that is, it abbreviates the sentence

$$[(A^- + B)^- + (B^- + A)^-]^- .$$

The notion of derivability for $\mathcal{T}$ is defined using the rule of detachment for $\Rightarrow$ (not $\rightarrow$) as the only rule of inference. The set of logical axiom schemata is chosen in a way that ensures the set of logically true sentences coincides with the set of derivable sentences. Following Church [5], p. 137, we use (C1)–(C3) as the schemata involving just disjunction and negation. (These schemata, together with the rule of detachment, are sufficient to ensure that every true sentence involving only disjunction and negation is derivable in $\mathcal{T}$.) The remaining schemata serve as definitions of the other constants and connectives of $\mathcal{T}$:

$$\mathring{1} \Leftrightarrow E + E^{-}, \tag{E1}$$

$$A \odot B \Leftrightarrow A \cdot B, \tag{E2}$$

$$A^{\smile} \Leftrightarrow A, \tag{E3}$$

for all sentences A and B of $\mathcal{T}$. The rule of detachment for $\mathcal{T}$ says: from two sentences

$$A \qquad \text{and} \qquad A \Rightarrow B,$$

infer B. A derivation in $\mathcal{T}$ of a sentence X from a set of sentences Ψ is a sequence of sentences such that the final sentence is X and such that each sentence in the sequence is either (a) an instance of one of logical axiom schemata (C1)–(C3) and (E1)–(E3), or (b) a sentence in Ψ, or else (c) a consequence of two earlier sentences in the sequence, obtained by applying the rule of detachment.

The models of $\mathcal{T}$ are quite different from the models of $\mathcal{L}_r^{\times}$. Sentences of sentential logic are usually assumed to designate one of two *truth values*, truth T or falsehood F. A model of $\mathcal{T}$ is therefore a pair consisting of the universe

$$\{T, F\}$$

and a single truth value (as the interpretation of E), either T or F. Thus, $\mathcal{T}$ has only two models. The sentential connectives of $\mathcal{T}$ are interpreted in a model in the standard way, and on the basis of this interpretation, a recursive definition of the notion of a sentence of $\mathcal{T}$ being true in a model may be given in the usual fashion.

24 $\mathcal{L}_r^{\times}$ as a Sublogic of $\mathcal{T}$

It is clear from the discussion in the preceding section that the languages $\mathcal{L}_r^{\times}$ and $\mathcal{T}$ have totally different semantics. Syntactically, however, they are very close. In fact, $\mathcal{L}_r^{\times}$ proves to be a (syntactical) sublogic of $\mathcal{T}$. It is clear that every predicate-sentence of $\mathcal{L}_r^{\times}$ is a sentence of $\mathcal{T}$. Also, every instance of a logical axiom schema of $\mathcal{L}_r^{\times}$ is derivable in $\mathcal{T}$. In order to see this, it is helpful to use the Completeness Theorem for the sentential logic with the primitive notions of disjunction and negation, and based upon schemata (C1)–(C3) and *modus ponens* for $\Rightarrow$ (see Church [5]).

Consider any instance of schemata (C4)–(C13). As a concrete example, take an instance of (C4). The sentence

$$[(A \cdot B) \cdot C] \Rightarrow [A \cdot (B \cdot C)] \tag{46}$$

is certainly derivable in $\mathcal{T}$ from (C1)–(C3) and the definition of $\cdot$, by the Completeness Theorem for $\mathcal{T}$. The sentence (C4) is derivable in $\mathcal{T}$ from (46) using definition (E2) and replacement (which also holds, by the Completeness Theorem).

The derivations of all instances of the other schemata proceed in a similar fashion. To derive instances of (C10)–(C13), it is helpful to use the sentence

$$(A \rightarrow B) \Leftrightarrow (A \Rightarrow B), \tag{47}$$

which is derivable in $\mathcal{T}$ using definitional schemata (E1) and (E2), together with (43) and (44).

It follows from (47) that the rule of detachment for $\rightarrow$ is valid in $\mathcal{T}$. A simple proof by induction on the definition of derivations in $\mathcal{L}_r^\times$ now shows that if a predicate-sentence X is derivable from a set of premises Ψ in $\mathcal{L}_r^\times$, then X is derivable from Ψ in $\mathcal{T}$. This completes the proof that $\mathcal{L}_r^\times$ is a syntactic sublogic of $\mathcal{T}$.

In view of the preceding remarks, the set Λ of logically derivable sentences of sentential logic may be viewed as a theory in $\mathcal{L}_r^\times$. Quite surprisingly, Λ can be obtained by adding one extremely simple sentence to the logical axioms of $\mathcal{L}_r^\times$, namely the sentence (45). In other words, sentential logic can be axiomatized in $\mathcal{L}_r^\times$ by the single sentence (45). As is well known, there is a decision procedure for the set Λ. On the other hand, the set Θ of logically derivable sentences of $\mathcal{L}_r^\times$ is undecidable. Thus, Θ may be viewed as an undecidable "subtheory" of the ordinary, two-valued sentential logic.

The above observations seem rather paradoxical. Sentential logic is usually regarded as the simplest part of mathematical logic, almost trivial in content. Nevertheless, it has a sublogic – namely $\mathcal{L}_r^\times$ – that provides an adequate framework for developing the richest mathematical discipline, namely set theory, and hence for developing all of mathematics.

25 Conclusion

Motivated, perhaps, by what he viewed as a less than successful effort on the part of Löwenheim to show that the Peirce-Schröder calculus of relations provides an adequate basis for the foundation for mathematics, Tarski developed an elegant and exceptionally simple form of that calculus, and in a theorem that is regarded as a *tour de force*, he proved rigorously that his form of the calculus does provide a formally adequate framework for the development of set theory and number theory, and hence for all of classical mathematics. Some of the consequences of his theorem are so surprising they seem almost paradoxical: three variables suffice for doing all of mathematics in a first-order language, and there are even variable-free subsystems of the sentential calculus that are sufficient for that purpose. (In describing this last result in a postcard to Willard van Orman Quine, dated March 27, 1942, Tarski concluded with the following play on a French saying: "Isn't [it] a nice thing 'pour épater les logiciens-bourgeois'?")

Although the systems described in this paper may not be well suited for everyday work by mathematicians, their formal simplicity and lack of variables make them eminently well-suited for automated reasoning. The recursive translation mappings provide a concrete mechanism for automatic translation of mathematical theories and assertions, and even of formal proofs, into these systems. The principal theorems regarding equivalence in means of expression and proof may be viewed as guarantees that these systems are adequate in both their expressive and deductive powers for automated reasoning about all parts of classical mathematics.

In the late 1940s, Tarski and his students Louise Chin and Frederick Thompson developed an algebraic version of the full first-order logic that they called cylindric algebra. Tarski's work on this subject in the 1950s and 1960s, much of it carried out in collaboration with Leon Henkin and Donald Monk, culminated in the two volumes [12] and [13]. A different but related algebraic approach to first-order logic was developed by Paul Halmos under the name polyadic algebra; see [10].

During the last 30 years, work in algebraic logic, particularly in the tradition of Tarski, has taken on an international character, with centers in Brazil, England, Germany, Holland, and Hungary. Many of the researchers are particularly interested in applications to computer science and philosophy (including arrow logic, dynamic logic, the bounded fragment, and the Lambek calculus). Some of the recent developments are surveyed in the paper [1], the series of articles in [30], and the monograph [16]. Applications of algebraic logic are discussed in the books [2], [3], and [9]. The textbooks [7], [4], and [11] are excellent references for logic, model theory, and Boolean algebra, respectively.

Acknowledgement The author expresses his gratitude to Professor Belinfante for extending an invitation to participate in the Special Session on Automated Reasoning in Mathematics and Logic, held March 8–10, 2002, at the Georgia Institute of Technology.

References

1. Andréka, H., Németi, I., Sain, I.: Algebraic logic. In: Gabbay, D.M., Guenther, F. (eds.) Handbook of Philosophical Logic, vol. 2, pp. 133–247. Second edition, Kluwer, Dordrecht (2001)
2. van Benthem, J.: Exploring Logical Dynamics. Studies in Logic, Language, and Information, xi + 329 pp. CSLI Publications, Stanford, CA (1996)
3. Brink, C., Kahl, W., Schmidt, G. (eds.): Relational Methods in Computer Science. Adv. Comput. Sci. xv + 272 pp. Springer, Berlin Heidelberg New York (1997)
4. Chang, C.C., Keisler, H.J.: Model Theory. Studies in Logic and the Foundations of Mathematics, vol. 73, xvi + 650 pp. Third edition, North-Holland, Amsterdam (1990)
5. Church, A.: Introduction to Mathematical Logic, vol. 1, x + 378 pp. Princeton University Press, Princeton (1956)
6. De Morgan, A.: On the syllogism, no. IV, and on the logic of relations. Trans. Camb. Philo. Soc. **10**, 331–358 (1864)
7. Enderton, H.B.: A Mathematical Introduction to Logic, xiii + 295 pp. Academic, New York (1972)
8. Enderton, H.B.: Elements of Set Theory, xiv + 279 pp. Academic, New York (1977)
9. Gabbay, D.M., Kurucz, Á., Wolter, F., Zakharyaschev, M.: Many-Dimensional Modal Logics: Theory and Applications. Studies in Logic and the Foundations of Mathematics, vol. 148, xviii + 747 pp. Elsevier, Amsterdam (2003)
10. Halmos, P.R.: Algebraic Logic, 271 pp. Chelsea, New York (1962)
11. Halmos, P.R.: Lectures on Boolean Algebras. Van Nostrand Mathematical Studies, no. 1, ii + 147 pp. D. Van Nostrand Company, Princeton, NJ, (1963) (Reprinted by Springer, Berlin Heidelberg New York (1974))
12. Henkin, L., Monk, D., Tarski, A.: Cylindric Algebras, Part I. Studies in Logic and the Foundations of Mathematics, vol. 64, vi + 508 pp. North-Holland, Amsterdam (1971)
13. Henkin, L., Monk, D., Tarski, A.: Cylindric Algebras, Part II. Studies in Logic and the Foundations of Mathematics, vol. 115, ix + 302 pp. North-Holland, Amsterdam (1985)
14. Hirsch, R., Hodkinson, I., Maddux, R.D.: Relation algebra reducts of cylindric algebras and an application to proof theory. J. Symb. Log. **67**, 197–213 (2002)
15. Hirsch, R., Hodkinson, I., Maddux, R.D.: Provability with finitely many variables. Bull. Symb. Log. **8**, 348–379 (2002)
16. Hirsch, R., Hodkinson, I.: Relation Algebras by Games. Studies in Logic and the Foundations of Mathematics, vol. 147, xvii + 691 pp. Elsevier, Amsterdam (2002)

17. Hungtington, E.V.: New sets of independent postulates for the algebra of logic, with special reference to Whitehead and Russell's Principia Mathematica. Trans. Amer. Math. Soc. **35**, 274–304 (1933)
18. Hungtington, E.V.: Boolean algebra. A correction. Trans. Amer. Math. Soc. **35**, 557–558 (1933)
19. Löwenheim, L.: Über Möglichkeiten im Relativkalkül. Math. Ann. **76**, 447–470 (1915)
20. Löwenheim, L.: Einkleidung der Mathematik in Schröderschen Relativkalkul. J. Symb. Log. **5**, 1–15 (1940)
21. Maddux, R.D.: Topics in relation algebras. Doctoral dissertation, University of California, Berkeley, iii + 241 pp. (1978)
22. Maddux, R.D.: Nonfinite axiomatizability results for cylindric and relation algebras. J. Symb. Log. **54**, 951–974 (1989)
23. Maddux, R.D.: Relation algebras of formulas. In: Orlowska, E. (ed.) Logic at Work. Studies in Fuzziness and Soft Computing, vol. 24, pp. 613–636. Springer, Heidelberg and New York (1999)
24. McKenzie, R.N.: The representation of relation algebras. Doctoral dissertation, University of Colorado, Boulder, vii + 128 pp. (1966)
25. McKenzie, R.N.: Representations of integral relation algebras. Mich. Math. J. **17**, 279–287 (1970)
26. Monk, J.D.: On representable relation algebras. Mich. Math. J. **11**, 207–210 (1964)
27. Monk, J.D.: Nonfinitizability of classes of representable cylindric algebras. J. Symb. Log. **34**, 331–343 (1969)
28. Moore, G.H.: A house divided against itself: The emergence of first-order logic as the basis for mathematics. In: Phillips, E. (ed.) Studies in the History of Mathematics. Studies in Mathematics, vol. 26, pp. 98–136. The Mathematical Association of America (1987)
29. Németi, I.: Free algebras and decidability in algebraic logic (in Hungarian). Habilitation, Hungarian Academy of Sciences, Budapest, xviii + 169 pp. (1986) (An English translation is available from the author.)
30. Németi, I., Sain, I. (special eds.): Log. J. IGPL **8**, 373–591 (2000)
31. Peirce, C.S.: Note B. The logic of relatives. In: Peirce, C.S. (ed.) Studies in Logic by Members of the Johns Hopkins University, pp. 187–203. Little, Brown, and Company, Boston (1883) (Reprinted by John Benjamins, Amsterdam, 1983.)
32. Russell, B.: The Principles of Mathematics. Cambridge University Press (1903) (Reprinted by Allen & Unwin, London, 1948.)
33. Schröder, E.: Vorlesungen über die Algebra der Logik (exakte Logik), vol. III, Algebra und Logik der Relative, part 1, 649 pp. Publisher: B. G. Teubner, Leipzig (1895) (Reprinted by Chelsea, New York, 1966.)
34. Tarski, A.: On the calculus of relations. J. Symb. Log. **6**, 73–89 (1941)
35. Tarski, A.: A simplified formalization of predicate logic with identity. Arch. Math. Log. Grundl. Forsch. **7**, 61–79 (1965)
36. Tarski, A., Givant, S.: A Formalization of Set Theory without Variables. Colloquium Publications, vol. 41, xxi + 318 pp. Am. Math. Soc. Providence RI (1987)
37. Tarski, A., Mostowski, A., Robinson, R.M.: Undecidable Theories. Studies in Logic and the Foundations of Mathematics, xi + 98 pp. North-Holland, Amsterdam (1953)
38. Vaught, R.L.: On a theorem of Cobham concerning undecidable theories. In: Nagel, E., Suppes, P., Tarski, A. (eds.) Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress, pp. 14–25. Stanford University Press, Stanford (1962)