THE BULLETIN OF SYMBOLIC LOGIC
Volume 2, Number 2, June 1996

THE DISCOVERY OF MY COMPLETENESS PROOFS

LEON HENKIN

Dedicated to my teacher, Alonzo Church, in his 91st year.

§1. Introduction. This paper deals with aspects of my doctoral dissertation[1] which contributed to the early development of model theory. What was of use to later workers was less the results of my thesis, than the method by which I proved the completeness of first-order logic—a result established by Kurt Gödel in *his* doctoral thesis 18 years before.[2]

The ideas that fed my discovery of this proof were mostly those I found in the teachings and writings of Alonzo Church. This may seem curious, as his work in logic, and his teaching, gave great emphasis to the constructive character of mathematical logic, while the model theory to which I contributed is filled with theorems about very large classes of mathematical structures, whose proofs often by-pass constructive methods.

Another curious thing about my discovery of a new proof of Gödel's completeness theorem, is that it arrived in the midst of my efforts to prove an entirely different result. Such "accidental" discoveries arise in many parts of scientific work. Perhaps there are regularities in the conditions under which such "accidents" occur which would interest some historians, so I shall try to describe in some detail the accident which befell me.

Received November 17, 1995,and in revised form, January 4, 1996.

This paper was presented on August 24, 1993, at the XIXth International Congress of History of Science, held in Zaragoza, Spain. It was included in Symposium 6 of the Congress, Histoire de la théorie des modèles, organized by Hourya Sinaceur of CNRS, Paris. I am grateful to Dr. Sinaceur for inviting me to participate in the Symposium and facilitating my attendance at the Congress.

[1]The dissertation has never been published. It is entitled *The completeness of formal systems*, and was submitted to Princeton University in June, 1947. The contents are described below, in Section 3 of this paper. Parts were rewritten for later publication; see [Henkin, 1949], [Henkin, 1950], [Henkin, 1953], in the References at the end of this paper.

[2]Kurt Gödel's dissertation was submitted to the University of Vienna in 1929. It was re-written later that year for publication, [Gödel, 1930]. The text of the original dissertation was published in Gödel's collected works, [Feferman, 1986].

© 1996, Association for Symbolic Logic
1079-8986/96/0202-0001/$04.20

A mathematical discovery is an idea, or a complex of ideas, which have been found and set forth under certain circumstances. The *process of discovery* consists in selecting certain input ideas and somehow combining and transforming them to produce the new output ideas. The process that produces a particular discovery may thus be represented by a diagram such as one sees in many parts of science; a "black box" with lines coming in from the left to represent the input ideas, and lines going out to the right representing the output. To describe that discovery one must explain what occurs inside the box, i.e., how the outputs were obtained from the inputs.

In the present case we are primarily interested in a single output idea, the idea of my proof of completeness of first-order logic. However, the dissertation in which I first set this down contains other results, including another completeness proof. When we look into the black box we shall see that the production of the primary output cannot be understood without reference to the others, so the diagram for our discovery will have several outputs. (See Section 3, below.)

What about the inputs? In the case of a mature mathematician, the input ideas for a particular discovery may range very widely over studies and prior work compiled over an extended period of time. In the case of a doctoral dissertation the inputs can be identified more narrowly: Usually there is no prior work by the author, and one expects that lectures of teachers may have produced more significant inputs relative to input ideas from independent reading. In my own case there was indeed no prior work, and the number of teachers and of independent readings was reduced by circumstances relating to the history of the institutions where I studied and the military history of my nation; thus the number of inputs is small enough to permit a fairly complete listing. (See Section 2, below.)

In Section 2, below, I sketch the extent of ideas about logic which I encountered during the period 1938–1942, as a student. In Section 3 I outline my doctoral dissertation as it was accepted by Professor Church in June, 1947. In Section 4 I describe my efforts to write a dissertation, beginning in March, 1946.

§**2. Background.** In the Fall of 1938, in my second year as a student at Columbia University, I enrolled in a first course in logic offered in the Philosophy Department by Ernest Nagel, a distinguished philosopher of science who had helped found the Association for Symbolic Logic two years earlier. This course was not really mathematical in character, but it stimulated my curiosity in the subject and led me to browse in Bertrand Russell's *Principles of Mathematics*, [Russell, 1903], which I found by chance in a room of the library devoted to "books of general interest."

It was in that book that I first read about the principle of choice. I was enormously impressed by Russell's example of a shoe store with infinitely many pairs of shoes and of socks: How easy it was to specify one shoe from each pair in the shop, that one might wish to try on, and how seemingly impossible to specify one sock from each pair! As we shall see, it is just such a difficulty on which I focused when, eight years later, I began to work on my doctoral dissertation.

Russell's *Principle* led me to peek into *Principia Mathematica*, which he co-authored with Alfred Whitehead [Whitehead/Russell, 1910]. The volume of formalism in this work was too daunting for me to tackle, but I read the several sections of text introductory to the formal developments, and was impressed with the general ideas of the theory of types—which also lay at the start of my later dissertation work—and with the mysterious axiom of reducibility.

In Fall, 1939, when my third year of university studies began, I enrolled in an advanced course in logic taught by Nagel, and here came across my first experience with a mathematical treatment of a formal deductive system. The course treated systems of propositional and first-order logic taken from the little textbook by Hilbert and Ackermann.[3]

Most of the course consisted of constructing formal proofs. Metamathematical results such as normal forms were treated, but none of these linked semantical notions to the syntactical structures on which the course was based. In particular, the concept of completeness was never considered.

However, although Nagel did not incorporate this concept in the course itself, he did propose to me as a separate project the reading of W. V. Quine's proof of the completeness of propositional logic, that had appeared 18 months before in Volume 3 of the *Journal of Symbolic Logic*.[4] This was a stupendous experience in my education, not because of the subject of the paper, but because it showed me vividly that new work in logic, and more generally in mathematics, was being published, and that—with great difficulty—I could read and follow it. Although I took many courses in the Mathematics Department during my years at Columbia, this paper of Quine's was the only reading in mathematics, outside the textbooks of the courses, which any of my teachers suggested to me.

As to the concept of completeness which was the focus of Quine's paper, it did not get through to me. I simply noted that the aim of the paper was to show that every tautology had a formal proof in the system of axioms presented, and I expended my utmost effort to check Quine's reasoning that this was so, without ever reflecting on *why* author and reader were making this effort. This strictly limited objective also kept me from wondering how the author thought of putting the steps of the proof together; the result was

[3][Hilbert/Ackermann, 1928].

[4][Quine, 1938].

that I failed to get "the idea of the proof," the essential ingredient needed for discovery.

Just before I began this second course in logic taken with Nagel, the world entered a convulsive phase of its history when Poland was overrun by the German army, and World War II began. Alfred Tarski, a leading Polish logician, had left home a few days earlier to lecture at Harvard, at the invitation of Quine. Unable to return to his country and family, Tarski accepted invitations to lecture at other universities, eventually settling in Berkeley in 1942. Nothing of his work or his story was known to me when Nagel announced to his logic class that a famous Polish logician would come to Columbia to give a special lecture, and all of us students were urged to attend. I went eagerly, listened attentively. Tarski spoke of Kurt Gödel's work on undecidable propositions in the theory of types, published eight years earlier, and on decision procedures that had been found for some formal systems and shown not to exist for others. In the question period following the talk I asked whether there could be a decision procedure to tell whether a sentence of the system studied by Gödel was unprovable. It was very exciting for me to be in direct contact with "a famous logician."[5]

In class, subsequent to Tarski's talk, Nagel told us students that it had taken him six months to read Gödel's paper on the incompleteness of certain formal systems. At the time I inferred that this material could not be included in an ordinary course in logic, but would have to be made the subject of a special course all by itself. [6]

The two logic courses by Nagel were given in the Philosophy Department, and I took other philosophy courses as well, but my principal subject of study at Columbia during 1937–41 was mathematics. No courses in logic were given in the Mathematics Department, but in 1939–40, simultaneously with my second logic course, I studied projective geometry with a Professor Pfeiffer. Our textbook was the two-volume work by Oswald Veblen and J. W. Young, which begins with a metamathematical treatment of the axiom system used as the basis for deriving the theorems of the subject.[7] Pfeiffer was

[5]I met Tarski again at Princeton in 1946, when he spoke at a conference on logic held in connection with the celebration of the 200th anniversary of Princeton University. The following year I sent him a copy of my just-completed dissertation. After two post-doctoral years at Princeton, I took my first position as assistant professor at the University of Southern California in 1949. In 1952 I was invited to join Tarski at Berkeley, but I declined because of a "loyalty oath" required of all faculty members by the University of California at that time. Subsequently the oath was abolished, and I moved to Berkeley in 1953.

[6]Some years later, Nagel joined with a non-academic co-author to write a popular account of Gödel's work, [Nagel/Newman, 1958]. This work has been translated into several languages, of which the latest is a Hebrew translation in 1993 by Nitsa Hadar-Movshovitz and Yael Harpaz-Rubin. (Dr. Hadar-Movshovitz was the first of my Ph.D. students in the field of mathematics education.)

[7]See [Veblen/Young, 1910].

quite interested in these foundational details, including the independence of the axioms, the principle of duality, and the relation of the axioms to models defined within the theory of real numbers. I studied this material eagerly, and feel that at a deep level it provided a basis for the unexpected turn toward completeness that my dissertation work took seven years later, after having started in another direction.

As I stated above, the Mathematics Department at Columbia had no logician among its faculty, and offered no courses in logic during my years there. However, there was one faculty member who drew me into an important reading experience during my final year at Columbia, 1940–41. That was F. J. Murray, who worked on operator algebras, having collaborated with John von Neumann on publications in that field during and after a sojourn at the Institute for Advanced study.

The Princeton University Press had just brought out Gödel's monograph on the consistency of the axiom of choice and the generalized continuum hypothesis.[8] Although I had never been a student of Murray's, he knew that I was about the only one among mathematics students or faculty with an interest in logic; so he sought me out and proposed that the two of us work through the Gödel monograph together. I readily agreed, and ordered a copy of the monograph. As far as I can recall, Murray and I had one or two meetings to discuss the scope and the beginning of the work, and then he found himself too busy with other projects and I was left to work through Gödel's monograph on my own.

This event was probably my most important learning experience as an undergraduate. I gained much more of the content of Gödel's monograph than I had in reading Quine's paper the year before. I admired the metamathematical treatment whereby the comprehension schema of set-formation is obtained from finitely many axioms, and the sophisticated handling of inner-model constructions by means of the notion of the "absoluteness" of various set-theoretical notions. I was intrigued by the creation of a universal choice function in the realm of constructible sets, while none had been at hand in the realm of sets described by the original axioms, drawing my attention to a class of functions which were to be the starting point of my dissertation investigations five years later. Fortunately, an observation about choice functions that Gödel made was not included in the first printing of his monograph, otherwise it might have led me to discard outright the dissertation problem on which I embarked. See Observation A at the end of Section 4 below.

During my high-school studies I had thought of becoming a mathematics teacher, but in fall, 1940, I decided to apply for admission to a Ph.D. program in mathematics for the following year—without a clear idea of what sort of careers might follow. In the spring of 1941 I was accepted at three universities

[8]See [Gödel, 1940].

and chose Princeton, largely because I understood that there was a well known logician, Alonzo Church, in the Mathematics Department there. I had seen his listing as Editor of the *Journal of Symbolic Logic* when I had looked up Quine's paper in Volume 3, but had no idea what sort of work he did. In fact, I did not realize that creating and publishing mathematics was a regular part of a professor's work.

The Ph.D. program which I began at Princeton in fall, 1941, called for me to spend two years taking a variety of mathematics courses, then to pass a "qualifying" oral examination to show that I had a good grasp in three areas of mathematics, and then to write a dissertation containing original research results. In my first semester I took courses in logic, analysis, and general topology.

The logic course, given by Professor Church, extended over both semesters of the academic year. In comparison with present-day courses the material covered might be considered scanty. In the first semester various systems of propositional and first-order logic were introduced, normal forms were described and established and were used in proving completeness, and the (downward) Löwenheim-Skolem theorem was discussed. In the presentation of Gödel's completeness proof, emphasis was given to its reductive character: the provability of a logically valid formula is reduced first to the provability of its Skolem normal form, and then to the provability of some tautology in a specified set of propositional formulas.

In the second semester an applied second-order system for Peano arithmetic was studied in great detail, and the Gödel incompleteness results were derived for it; and from these followed the incompleteness of second-order logic. In connection with the incompleteness proofs, primitive-recursive functions received detailed examination in the course, but there was not time to study general recursive functions. However, their definition was mentioned, and their role in establishing the non-existence of certain decision procedures was described.

One detail of the second-order Peano theory deserves comment. The language of this theory did not contain any operation symbols, either constants or variables. Thus, binary operations on the domain of natural numbers such as addition and multiplication did not have names in the system, but the corresponding ternary relations—expressing that the sum or product of x and y is z—were represented by 3-place predicate constants. Of course the *possibility*, in general, of replacing symbols for functions by using associated symbols for relations, is a basic metamathematical result that could well be mentioned in a beginning logic course, but the idea of using such a replacement throughout the development of a formal theory of numbers now seems strange to me. At the time, however, I accepted it without question as a part of the formalization of mathematics within logical systems, since all first-order logical systems considered in the course were devoid of operation symbols.

I have sketched above the *content* of my logic course at Princeton, but the manner in which the material was presented by Church played an important part in generating the conception of logic that the students received from the course.

At every point of the course, Church would remind us that we were following "the logistic method" to study "logistic systems." This involved limiting our use of English, our meta-language, to set up and work with certain uninterpreted formal languages whose rules had to be specified with great exactitude in a completely effective way. This perspective is well set forth in the 68-page introductory chapter to Church's published textbook.[9]

A 12-page account of the logistic method forms Section 7 of the Introduction of the book, coming after 46 pages devoted to careful description of the linguistic components of languages such as those to be studied; Sections 8 (Syntax) and 9 (Semantics) conclude the Introduction. However, these two dimensions of language play very unequal roles in the deductive systems whose study is the proper role of logic, according to Church. This can be gleaned from the following passage, taken from Section 9.

"From time to time in the following chapters we shall interrupt the rigorous treatment of a logistic system in order to make an informal semantical aside . . . *Except in this Introduction, semantical passages will be distinguished from others by being printed in smaller type, the small type serving as a warning that the material is not part of the formal logistic development and must not be used as such.*"[10]

The one-year course in mathematical logic described above, taken at Princeton during 1941–42, contains all of my study of logic as a graduate student. In the middle of that year the U.S.A. was swept into World War II, requiring me to alter drastically my plans for graduate study. Instead of taking two years to prepare for my qualifying examination, I had to absorb parts of mathematics by reading, rather than by course work, in Spring 1942; I then passed the qualifying exam, received an M.A. degree, and left Princeton University for what were to be four years of work on military projects.[11]

[9][Church, 1956]. Early forms of Chapters I–IV were available in manuscript form in 1947, and contained the material covered in my first-semester course. Volume II of this work has never been published, but a "tentative table of contents of Volume II" is printed in Volume I, immediately following *its* table of contents, and shows that the material covered in my second-semester course was intended to appear in Chapters VI–VIII of Volume II.

[10]Emphasis of last sentence appears in the book.

[11]During the period May, 1942 - March, 1946 I worked as a mathematician, first on radar problems and then, beginning January 1943, on the design of a plant to separate uranium

§3. My Ph.D. dissertation. The dissertation submitted to Princeton University in June, 1947, contains my proof of completeness for first-order logic, as well as applications that we now consider to be part of the theory of models.[12]

In Section 4, below, I shall describe the year-long process of discovery of those results. To better understand that process, we set forth in the present section the main results of the dissertation, and the way in which they were presented there. We shall see later that the mode of organizing the dissertation serves to *hide* the process of discovery.

The dissertation contains four parts. **Part I** contains the new proof of the completeness of first-order logic, the discovery of which is to be described in the next section of this paper. Theorem I formulates what I call the *strong completeness* property for a system $\mathcal{L}$ of first-order logic that Church calls a *pure functional calculus of first order*. $\mathcal{L}$ has denumerably infinite lists of propositional symbols, of individual symbols, and of predicate symbols of each finite rank. A recursive description of the sentential formulas (called *well-formed formulas*, and abbreviated *wffs*) is given, and among these the formal axioms are identified by means of five schemas; the formal rules of inference are *detachment* and *generalization. Theorem I* states that *any set S of $\mathcal{L}$-sentences (wffs without free variables) that is formally consistent in the deductive system of $\mathcal{L}$, is satisfied by some denumerably infinite $\mathcal{L}$-structure $\mathcal{M}$.*

The key element of my proof of Theorem I is the enlargement of $\mathcal{L}$ to a new language $\mathcal{L}_1$, by adjoining an infinite sequence of new individual constants. The set C of all individual constants of $\mathcal{L}_1$ becomes the domain of individuals of the structure $\mathcal{M}$ that will satisfy all wffs of S. The n-place relation over C that is assigned to any n-ary predicate symbol of $\mathcal{L}$, in the structure $\mathcal{M}$, is defined by means of a certain consistent set S_1 of $\mathcal{L}_1$-sentences, obtained by enlarging S in two steps: First, for each formal theorem of existential form, a substitution instance (using a constant of C) is added to S, and then the resulting set is enlarged to a maximal consistent set of $\mathcal{L}_1$-sentences, to obtain $\mathcal{S}_1$.

Once Theorem I is established, three corollaries are easily obtained. *Corollary I* reads, *The pure first-order functional calculus is complete.* Completeness means that every wff that is logically valid (i.e., satisfied in every $\mathcal{L}$-structure), is formally provable using the formal axioms and rules of inference of $\mathcal{L}$. This corollary gives the content of the completeness theorem proved by Gödel in his doctoral dissertation in 1929.[13]

isotopes. Most of my work involved numerical analysis to obtain solutions of certain partial difference-differential equations. During this period I neither read, nor thought about, logic.

[12]The term "theory of models" did not gain wide usage until 1954, with the publication of [Tarski, 1954].

[13]The dissertation, [Gödel, 1929], answered the question of completeness posed in

Gödel *used* completeness to prove the statement given in our Theorem I, which I have called *strong* completeness. This nomenclature is justified because it is trivial to restate Theorem I in the following form: If S is any set of $\mathcal{L}$-sentences and r is any logical consequence of S (i.e., r is satisfied in every $\mathcal{L}$-structure that satisfies all sentences of S), then r is formally derivable from S (using the formal axioms and rules of inference of $\mathcal{L}$). When Theorem I is formulated in this way, Corollary I becomes a *special case* of Theorem I (the case where S is empty), so if the corollary expresses completeness, we can say that the theorem expresses strong completeness.

The remaining two corollaries of Theorem I are as follows. *Corollary II: If a set of wffs of $\mathcal{L}$ is satisfied in some $\mathcal{L}$-structure, then it is satisfied in some denumerably infinite $\mathcal{L}$-structure.* This, of course, is the content of the Skolem-Löwenheim theorem. *Corollary III: A set $\mathcal{S}$ of wffs of $\mathcal{L}$ is simultaneously satisfied in some $\mathcal{L}$-structure $\mathcal{M}$ if, and only if, each finite subset $\mathcal{S}_1$ of $\mathcal{S}$ is satisfied in some $\mathcal{L}$-structure $\mathcal{M}_1$.* This result is now called the *compactness property* of first-order logic, and has become one of the principal tools of model theory. The compactness property was not part of Gödel's dissertation [Gödel, 1929] but was added in the version written for publication [Gödel, 1930].

Part I of my dissertation contains a description of various first-order logical systems, differing in several ways from the pure first-order functional calculus $\mathcal{L}$, for which results analogous to Theorem I and its corollaries hold. First-mentioned are the *applied* first-order functional calculi, in which any finite or denumerably infinite set of constants is admitted in the formation of wffs. The constants may have the type of propositional, individual, or predicate symbols, and the variables (except for those of individual type) may be omitted.

In view of the fact that the formal logical axioms are given by means of schemas having infinitely many instances, the deductive system for $\mathcal{L}$ employed in the dissertation dispenses with substitution rules for the several types of variables; hence there is no difference in the treatment of variables and constants (except for individual variables which, appearing in quantifiers, have a special status affecting their occurrence in formal axioms).

It is stated that the applied calculi without predicate variables are used to formalize various mathematical theories in algebra and geometry, where the logical axioms given for propositional connectives are supplemented by axioms intended to express the character of the mathematical concepts of the theory. Many of these systems require an equality symbol for the relation

[Hilbert/Ackermann, 1928]; it was presented to the University of Vienna in 1929. Gödel's Ph.D. degree was granted in Feb., 1930. Gödel rewrote the dissertation material for publication, submitting it in October, 1929; it appeared in 1930. [Gödel, 1930]. The dissertation itself was finally published in 1986, in Volume I of the collected works of Gödel, edited by Feferman et al, [Feferman, 1986].

of identity; this can be treated as a logical constant, by introducing a binary predicate constant, Q, and adjoining a new set E of formal axioms to the deductive apparatus of $\mathcal{L}$. Gödel, in his dissertation, showed how a strong completeness theorem for first-order logic can be extended to cover the case of first-order logic with identity, and this method is borrowed to cover such applied first-order functional calculi in my dissertation.[14]

The final way mentioned, for generalizing Theorem I and its corollaries to a wider class of first-order deductive systems, is to consider applied first-order functional calculi $\mathcal{L}$ in which a *non-denumerable* set of individual and predicate symbols is used in the formation of wffs. The formal axiom schemas and rules of inference remain exactly as before. In such a system, one can consider a consistent set S of sentences that is non-denumerably infinite. A structure $\mathcal{M}$ satisfying each sentence of S can be obtained with minor and obvious changes to the method used in the proof of Theorem I.[15]

Part I of my dissertation ends with the formulation, for later reference, of a Theorem II. This records the observations that had been made, to the effect that the strong completeness property holds for a wide class of first-order systems, rather than only for the pure first-order functional calculus mentioned in Theorem I. Specifically, *Theorem II* states that *if $\mathcal{L}$ is any consistent set of sentences of an applied, extended, first-order functional calculus $\mathcal{L}$ (which may include an equality-symbol to refer to the identity relation in $\mathcal{L}$-structures), then there is some $\mathcal{L}$-structure $\mathcal{M}$, having a domain whose cardinality does not exceed the cardinal number of the set of all $\mathcal{L}$-symbols, which satisfies each sentence of S.*

Immediately following Theorem II there appears a *Corollary: A set $\mathcal{S}$ of sentences of a system $\mathcal{L}$ of the kind described in Theorem II is satisfied in some structure $\mathcal{M}$, if and only if each finite subset S_1 of S is satisfied in some structure $\mathcal{M}_1$.* This corollary expresses the compactness property for sets of

[14]The axioms E assure that in any structure $\mathcal{M}$ satisfying a given consistent set $\mathcal{S}$ of $\mathcal{L}$-sentences, the predicate symbol Q will denote a congruence relation Q' of the structure. Using this, one can form from $\mathcal{M}$ and Q' a "quotient structure", $\mathcal{M}^{\star}$, whose elements are the equivalence classes induced by Q' on the domain of individuals of $\mathcal{M}$; this parallels the algebraic construction of quotient groups and quotient rings. It is then easy to check that $\mathcal{M}^{\star}$ is a structure that also satisfies $\mathcal{S}$, in which the symbol Q denotes the identity relation. However, when $\mathcal{M}$ is denumerably infinite, $\mathcal{M}^{\star}$ may be finite; hence the formulation of Theorem I must be modified accordingly, when applied to first-order logic with identity.

[15]First, L is enlarged to L_1 by adjoining a set of new individual constants having the same cardinality k as the set of all symbols of L. This assures that the set C of all individual constants of of L_1, which serves as the domain of individuals for the structure $\mathcal{M}$ which is being formed, will have this same cardinality k. Then, in enlarging S to the set S_1 of L_1-sentences that is used to determine which n-place relation over C is assigned by $\mathcal{M}$ to each n-ary predicate symbol, the axiom of choice is used to arrange all L_1-sentences in a well-ordered sequence of length k, instead of using a simple enumeration of sentences as was done to form S_1 in the proof of Theorem I.

first-order sentences having any cardinality. It is the principal tool used for a series of applications that make up Part II of the dissertation.

Part II is entitled *Applications to algebra*. It begins by setting up an applied first-order calculus for ring theory. Following the example I encountered in my second-semester course with Church, in Spring 1942, I used no operation symbols, only a binary relation-symbol for equality and two ternary relation symbols to formalize addition and multiplication.[16]

Using these, I listed ten ring axioms that were to be added to the logical axioms in this applied calculus. At the same time I indicated how, formalizing results about a particular ring, $\mathcal{R}$, one could adjoin to the preceding system a set of individual constants correlated with the elements of $\mathcal{R}$, and add to the ring axioms what I called *the basic sentences* of $\mathcal{R}$—sentences employing the individual constants to indicate, for each ordered triple of elements, whether or not the sum (or the product) of the first two is equal to the third.[17]

The first application given of the compactness property for first-order logic consists of a new proof of the Boolean representation theorem, first shown by Marshall Stone in [Stone, 1936]. So *Theorem III* of the dissertation reads: *Every Boolean ring is isomorphic with some subring of the ring of all subsets of some fixed domain of individuals* (*in which the ring operations are intersection and union* (*modulo* 2)). The proof given for Theorem III uses the Corollary of Theorem II (compactness) to show that a given Boolean ring $\mathcal{R}$ can be extended to a Boolean ring that is atomistic; the representation of the latter by the ring of all sets of its atoms is easily accomplished. The given Boolean ring $\mathcal{R}$ is brought into the extended, first-order applied calculus by means of basic sentences involving individual constants correlated with elements of $\mathcal{R}$. That each finite subset of these is consistent with axioms for atomistic Boolean rings, is proved by showing that any finite subset generates a finite subring of $\mathcal{R}$, and that every finite Boolean ring is atomistic.

Theorem III, which is an algebraic theorem given a metamathematical proof, is used as an example to motivate a general theorem of model theory. First, definitions are given for the expressions *a type of algebraic structures, a particular algebraic structure, basic sentences of an algebraic structure*, and *elementary property* of a type of algebraic structures.[18] Then, *Theorem IV*

[16]When, in 1951, I re-wrote this part of my dissertation for publication (it was the third part to appear), I discontinued this anachronistic feature, and employed a first-order calculus containing operation symbols. But when, in 1948, I rewrote Part I of the dissertation for publication, the operation symbols were still lacking. See [Henkin, 1953] and [Henkin, 1949], respectively.

[17]This set of "basic sentences" of $\mathcal{R}$ has since come to be called *the diagram of* $\mathcal{R}$, following terminology introduced by Abraham Robinson, who independently found most of the results of Part II of my dissertation by similar methods.

[18]In the current terminology of model theory, *a type of algebraic structures* is an elementary class, i.e., the class of all models of some first-order sentence, *a particular algebraic structure*

is formulated as follows: *A necessary and sufficient condition that a given algebraic structure $\mathcal{A}$ be a sub-structure of an algebraic structure $\mathcal{B}$ of the same type which has a given elementary property P, is that it be possible to imbed each finite subset $\mathcal{A}_1$ of $\mathcal{A}$, in a structure $\mathcal{B}_1$ having property P, in such a way that the basic sentences for the elements of $\mathcal{A}_1$ are preserved.*

A discussion of elementary properties follows in the dissertation, making explicit that a property of structures defined by higher-order sentences may be elementary if there is an equivalent first-order sentence. This is exemplified by showing that, for the algebraic class of fields, and for any prime number q, the property to be of characteristic q is indeed elementary. Using this, the dissertation *Theorem V* is formulated: *if, for an infinite sequence of prime numbers q_i there exist fields of characteristic q_i having a certain elementary property P, then there exists a field of characteristic zero having property P.*

The following *Corollary I* is obtained by using a complete diagram for a given field of characteristic zero. *If, for a given infinite sequence of prime numbers q_i, every field of characteristic q_i has an elementary property P, then every field of characteristic zero can be embedded (as a subfield) in some field (of the same cardinality) having property P.*

Corollary II for Theorem V states that *the property of a field to be of prime characteristic is not elementary*, even though the property to be of characteristic q for any particular prime number q, *is* elementary. Then, a *quasi-elementary property* being identified as one defined by some *set* of first-order sentences, it is remarked that *the property of fields to be of characteristic zero is quasi-elementary but not elementary, while the property to be of prime characteristic is not even quasi-elementary.*

This concludes Part II of the dissertation.

Part III of the dissertation is entitled *The calculi of higher order*. This was intended to be the climax of the work because it throws new light on Gödel's incompleteness results.

In a theory of types it is possible to *define* the natural numbers and the arithmetical operations on them, so that one can develop the Peano theory of numbers in a pure logical calculus of high enough order. Even in second-order logic, because the Peano postulates are categorical it is possible to correlate with each sentence of arithmetic a sentence of the pure logical calculus of second order, such that the former is true for the system of natural numbers if, and only if, the latter is logically valid. Thus the sentences of arithmetic constructed by Gödel, which he showed to be true of the natural numbers but unprovable in an appropriate formal deductive system, provide

is an element of such a type of structures, the *basic sentences* of an elementary structure are the sentences of its diagram (as mentioned in Footnote 18), an *elementary property* of a type of algebraic structures, is a property defined by a first-order sentence of the language of that type of algebraic structures.

us with valid sentences of pure logical calculi which cannot be formally proved in the deductive systems with which these calculi are usually equipped.

The unprovable sentences constructed for given formal systems by Gödel are very special, and one may wonder whether there is some general criterion, involving the truth or falsity of sentences under suitable interpretations of the language, which can distinguish the provable sentences from the unprovable ones. This is exactly what is accomplished in Part III of my dissertation.

An interpretation of the formal languages created in mathematical logic is always made with respect to some *structure*. A structure $\mathcal{M}$ always possesses a non-empty domain D of elements, which serves (in the interpretation of a language $\mathcal{L}$) as the range of individual variables of $\mathcal{L}$. Additional components of $\mathcal{M}$ may be designated elements, operations, or relations over D, which serve as the denotations of any individual constants, operation constants, or predicate constants of $\mathcal{L}$. If $\mathcal{L}$ is of higher order, it may have some constant symbols of higher type. For example, if $\mathcal{L}$ is intended for use in the theory of topological spaces, it may have a unary predicate symbol R to serve as a name for the set of all open sets of points. In that case, any structure used to interpret $\mathcal{L}$ would have to possess a designated set R' of subsets of D, to serve as the denotation of the symbol R. If S is a given set of $\mathcal{L}$-sentences, any structure $\mathcal{M}$ for $\mathcal{L}$ which satisfies all sentences of S is called a *model of* S. Sometimes, even if no particular set of $\mathcal{L}$-sentences is specified, a structure $\mathcal{M}$ with components appropriate for interpreting $\mathcal{L}$ is called an $\mathcal{L}$-*model*.

Part III of my dissertation begins with a discussion of the pure functional calculus of second order, $\mathcal{L}_2$. This is obtained from the pure calculus of first order (mentioned above as the subject of Theorem I of Part I of the dissertation), by allowing propositional and predicate variables to appear in quantifiers, as well as individual variables. Since this system has no constant symbols, a structure $\mathcal{M}$ used for interpreting $\mathcal{L}_2$ need not have any component other than a domain D of individuals. For each positive integer n, the n-ary predicate variables of $\mathcal{L}_2$ (which always appear within existential or universal quantifiers in $\mathcal{L}_2$-*sentences*, though they have free occurrences in wffs that are not sentences, such as atomic wffs), range over the set of all n-place relations over D when $\mathcal{L}_2$ is interpreted with respect to $\mathcal{M}$.

Clearly, any two $\mathcal{L}_2$-structures whose domains of individuals have the same cardinality, will satisfy the same $\mathcal{L}_2$-sentences. A sentence of $\mathcal{L}_2$ that is satisfied by *every* $\mathcal{L}_2$-structure is *logically valid*. All formal theorems of $\mathcal{L}_2$ (sentences derivable from the formal axioms of $\mathcal{L}_2$ by using the formal rules of inference), are logically valid. But Gödel showed how to construct logically valid $\mathcal{L}_2$-sentences that are *not* formal theorems.

In Part III of my dissertation structures called *general models* are introduced, which can be used to interpret $\mathcal{L}_2$. The old structures are among them, and are called *standard models*. An $\mathcal{L}_2$ sentence that is satisfied in every general model is called *logically valid in the general sense*. It is easy to see that every formal axiom of $\mathcal{L}_2$ is satisfied by every general model, and that the formal rules of inference preserve this property, so that *every formal theorem of $\mathcal{L}_2$ is logically valid in the general sense*.[19] What is stated in Part III of my dissertation is that the converse holds, so that we get a *generalized completeness theorem: Every $\mathcal{L}_2$-sentence that is logically valid in the general sense, is formally provable in $\mathcal{L}_2$*.

The intuitive idea for the definition of general $\mathcal{L}_2$-models is simple. Such a structure $\mathcal{M}$, instead of consisting of only a single domain D, is to consist of infinitely many components, $(D_0, D_2, \ldots, D_n, \ldots)$. D_0 is to be an arbitrary non-empty set. D_1 is to be a domain of some subsets of D_0 and, for each $n > 1$, D_n is to be a domain of some n-place relations over D_0. When such a generalized model is used to interpret the $\mathcal{L}_2$-sentences, D_0 serves as the range of individual variables and, for each $n > 0$, D_n serves as the range for n-ary predicate variables. However, the sets and relations chosen as members of the domains cannot be selected arbitrarily.

The intuitive idea given above for general $\mathcal{L}_2$-structures must be complicated, by providing conditions to ensure that every formally provable $\mathcal{L}_2$-sentence is true for all general models. This is because, among these formally provable $\mathcal{L}_2$-sentences, are instances of the *comprehension principle for set theory*. For example, being given any wff b of $\mathcal{L}_2$ containing free occurrences of individual variables x_1 and x_2 (and having no other variables occurring freely in it), there is a formal theorem asserting the existence of a 2-place relation R consisting of just those ordered pairs of individuals that satisfy b. Hence, we need to make sure that in our generalized models, the domain D_2 contains such a relation R.

This requirement results in a definition of general models $\mathcal{M}$ whose domains D_n, for $n > 0$, are closed under Boolean operations, and such that D_n contains all projections of elements of D_{n+1}, for each $n > 0$. These conditions arise from the sentential connectives, and the quantifiers containing individual variables of $\mathcal{L}$. However, further conditions are needed to cover requirements connected with quantifiers containing predicate variables. This makes the precise definition of generalized models complicated, and it becomes unclear, without a proof, that there *are* any general models for $\mathcal{L}_2$ other than the standard model (in which, for each $n > 0$, D_n is the set of *all* n-place relations over D_0).

The generalized completeness theorem for $\mathcal{L}_2$ cited above shows, in the light of Gödel's incompleteness results for $\mathcal{L}_2$, that indeed there must be non-standard general models for $\mathcal{L}_2$. The proof of that completeness theorem gives a general method for constructing such non-standard models.

[19]This result is the *general soundness theorem for $\mathcal{L}_2$*.

The discussion (in Part III of my dissertation) of the pure logical calculus of second order, $\mathcal{L}_2$, is not more detailed than what has been said above; in particular, the definition of general models of $\mathcal{L}_2$ is not made precise, and no proof is given for the generalized completeness theorem. Instead, a detailed account of these matters is given for a logical system $\mathcal{T}$ of infinite order, in the form of a theory of types formulated by Church.[20]

We shall not reproduce details of the system $\mathcal{T}$ at this point, as we shall need to do that in Section 4 of this paper, below. Here, we merely set down the content of Theorem VI of the dissertation, which states a strong completeness property for $\mathcal{T}$ with respect to general models. Three corollaries are then listed, analogous to the corollaries of Theorem I that were given for the system $\mathcal{L}$ of first-order logic in Part I.

THEOREM VI. *If S is any formally consistent set of $\mathcal{T}$-sentences, then there is a denumerable general model $\mathcal{M}$ of $\mathcal{T}$ for which each sentence of S is true.* (*Each domain D_n of $\mathcal{M}$ is denumerable.*)

COROLLARY I. *The deductive system of $\mathcal{T}$ is complete* (*with respect to interpretations by general models*).

COROLLARY II. *If S is a set of $\mathcal{T}$-sentences satisfied by some general model, then there is a denumerable general model satisfying S.*

COROLLARY III. *If S is a set of $\mathcal{T}$-sentences such that each finite subset S_1 is satisfied by some general model $\mathcal{M}_1$, then there is some general model $\mathcal{M}$ that satisfies all sentences of S.*

There is also a Theorem VII, extending Theorem VI to cover applied logical systems of type-theory, obtained by adjoining a set of constants (of various types) of any cardinality. This is analogous to the passage from Theorem I to Theorem II in Part I of the dissertation. A compactness corollary for Theorem VII, paralleling Corollary III of Theorem VI, completes Part III of the dissertation.

Part IV of the dissertation is entitled *Applied systems of logic*. It begins by confessing that the compactness principle for higher-order formal languages is unlikely to find applications in various parts of mathematics, which were exemplified for first-order languages in Part II of the dissertation. This is because formal definitions for higher-order concepts used in mathematics, such as the concept of *topological space*, change their meaning when interpreted with respect to a general model.[21]

[20]See [Church, 1940].

[21]One minor application of compactness is given, to obtain a new proof of a result obtained in 1947 by C. J. Everett and G. Whaples. This gives a necessary and sufficient condition on a set M of finite sets, for the existence of a choice function f on M such that $f(A) \neq f(B)$ whenever A and B are distinct elements of M.

Instead of new results about structures that have been considered in mathematics, Part IV contains descriptions of applied formal languages for set theory, and for number theory, and calls attention to the possibility of finding non-standard models for axiom systems used in these areas. Examples considered are Bernays/Gödel set-theory formalized as a second-order system (or extensions to higher-order systems), and Peano number-theory (formalized as a second-order system).

It is shown how non-standard models of number theory arise when compactness is used to get a general model satisfying a set of sentences of the form $u \neq 0, u \neq 1, \ldots$, where u is an individual constant adjoined to the second-order system in which Peano axioms are formalized. The theory of such a model does not satisfy the condition of ω-consistency, used by Gödel in formulating his incompleteness results. It is shown that the order-type of such a *denumerable* non-standard model must be $\omega + \eta(\omega^* + \omega)$.

The dissertation ends on a philosophical note concerning non-standard models. It suggests that Gödel's incompleteness results can be considered as stating a fundamental inability to communicate the kind of mathematical systems we are examining, rather than an inability to establish facts about such a system.

§4. Working on my dissertation—the discovery. In March, 1946, I returned to Princeton to complete my work for the Ph.D. degree after an interruption of almost four years. The sole remaining requirement for the degree was that I write an acceptable dissertation. I was awarded a pre-doctoral fellowship by the National Research Council, part of the government's effort to refill colleges and universities with students and faculty from among those returning from war-time work.

Immediately upon my return I began to attend a course in logic that Professor Church had begun the preceding month. The subject was Frege's theory of sense and denotation, of which I had never heard before. Through careful study of examples, Church made a convincing case for Frege's thesis that to understand how language functions in conveying meaningful communications, it is not sufficient to study the relation between the symbolic linguistic structure and the universe of objects to which it refers—it is necessary to posit a third realm of abstract entities called *senses*, or *concepts*. Under this theory a symbolic expression functioning as a name *denotes* an object of the universe of discourse, and *expresses* some sense of that object; a sentence is construed as a name of its truth value, and the sense it expresses is called a *proposition*.

It was Church's aim to develop a mathematical theory of senses and their relation to the objects to which they refer. To obtain utmost precision, this was to be a formalized axiomatic theory. As a vehicle, he chose a formulation of the simple theory of types that he had published in 1940, and elaborated it by adjoining a new hierarchy of types of senses.[22]

The Church–Frege theory of sense and denotation has been of continuing interest to me, but Church's 1940 theory of types, which we shall here call $\mathcal{T}$, enthralled me from the moment it appeared in the 1946 course. Within a few weeks I had formulated a conjecture about it, and set out to find a proof which I hoped to incorporate in a dissertation. To follow the course of these ideas we must examine the theory $\mathcal{T}$ in some detail.

In $\mathcal{T}$, there are *two* domains or types, at the base of the hierarchy of types. D_1 is the *type of individuals*, and D_0 is *the type of truth-values*.[23] Further types are built up from D_0 and D_1 by providing, for each types D_a and D_b, a type $D_{(ab)}$ of functions from D_b to D_a. [A *subset* of D_a is identified with the function of $D_{(0a)}$ which assigns to all of its elements, and to no others, the value T (truth); thus, $D_{(0a)}$ serves as the type of all subsets of D_a. A *function of* $n + 1$ *variables* is identified with a function of its first variable, whose values are functions of its last n variables. Thus, a binary relation between elements of D_a and of D_b is identified with a function of type $D_{((0b)a)}$, and the latter serves as the type of all such relations.[24]]

Now we are ready to describe the formal language of $\mathcal{T}$ which, when interpreted, may be used to make statements about the hierarchy of types described above.

As to *symbols* of $\mathcal{T}$, there are *variables*, *constants*, and three *improper symbols* λ,), and (. For each type symbol a, an infinite list of variables of type a is given, e.g., $c_a, d_a, \ldots, q_a, c'_a, \ldots$. There are constants N_{00}, A_{000}, and, for each type symbol a, constants $\pi_{0(0a)}$ and $\iota_{a(0a)}$.

Certain strings of symbols are called *well-formed formulas* (*wffs*), and each is given a *type*. First of all, any variable or constant symbol by itself is a wff, and has the type of its subscript. Then, there are two ways to build a longer wff from two given ones. (i) If $\mathbf{x}_b$ is any variable of type b and $\mathbf{M}_a$ is any wff of type a, then $(\lambda \mathbf{x}_b \mathbf{M}_a)$ is a wff of type (ab). (ii) If $\mathbf{F}_{ab}$ and $\mathbf{B}_b$ are

[22]See [Church, 1940] for the original theory. A version of the theory incorporating types of senses was published later, [Church, 1951].

[23]For ease in publication we make various typographical changes from the theory $\mathcal{T}$ as it appears in [Church, 1940]. For example, Church calls D_0 *the type of propositions*, and states, "We purposely refrain from making more definite the nature of the types D_0 and $D_1, \ldots$". However, in the terminology of the theory of sense and denotation propounded in his 1946 course, the formal sentences of $\mathcal{T}$ are wffs of type 0; since sentences denote truth values, *these* must be the elements of D_0; propositions, being senses *expressed* by sentences, lie in a separate type which is not a component of the theory $\mathcal{T}$.

[24]Hereafter we drop outer parentheses in writing type symbols, and we use a sequence of more than two type symbols to abbreviate the type symbol obtained by associating to the left. Thus, D_{0ba} is used as the type of binary relations between elements of D_a and of D_b.

wffs of types ab and b respectively, then $(\mathbf{F}_{ab}\mathbf{B}_b)$ is a wff of type a.[25] Every occurrence of a variable $\mathbf{x}_b$ in a wff $\mathbf{M}_a$ is free, unless it is within a part of $\mathbf{M}_a$ of the form $(\lambda\mathbf{x}_b\mathbf{N}_c)$, in which case it is bound. A *closed wff* (*cwff*) is a wff with no free variables.

When the language of $\mathcal{T}$ is used to make statements about the hierarchy of types, each wff $\mathbf{B}_b$ refers to an element of the type D_b. A wff of form $\mathbf{F}_{ab}\mathbf{B}_b$ refers to the element of D_a obtained by applying the function designated by $\mathbf{F}_{ab}$ to the argument designated by $\mathbf{B}_b$. A wff of the form $\lambda\mathbf{x}_b\mathbf{M}_a$ refers to the function in D_{ab} which, when applied to an argument z_b, yields as value the element designated by the wff $\mathbf{M}_a$ when each free occurrence of $\mathbf{x}_b$ in $\mathbf{M}_a$ is assigned the value z_b. Thus, an occurrence of $\lambda\mathbf{x}_b$ in a wff serves as a functional abstractor for the wff $\mathbf{M}_a$ to which it is prefixed.

To complete the description of the intended meanings of wffs, as referring to elements of the types of the theory $\mathcal{T}$, we must indicate the elements to which variables and constants refer. Of course, a variable $\mathbf{x}_b$ ranges over all of the domain D_b; it only refers to a particular element of the latter when one is assigned to it in some linguistic context. Similarly, a wff $\mathbf{M}_a$ which contains free variables of one or more types, will not refer to a particular element of D_a unless values of appropriate type are assigned to those variables.

The constant N_{00} denotes the negation function of D_{00} which, acting on either truth value of D_0 gives the other one as its value. The constant A_{000} denotes the disjunction function of D_{000}, so that the wff $(A_{000}p_0)q_0$ denotes T iff either of the variables p_0, q_0 is assigned the value T, and denotes F if both variables are assigned F. Because of these meanings we make contact with traditional symbols for propositional connectives by writing $\sim \mathbf{B}_0$ for $N_{00}\mathbf{B}_0$, $\mathbf{B}_0 \vee \mathbf{C}_0$ for $(\mathcal{A}_{000}\mathbf{B}_0)\mathbf{C}_0$, and $\mathbf{B}_0 \supset \mathbf{C}_0$ for $(\sim \mathbf{B}_0) \vee \mathbf{C}_0$, for any wffs $\mathbf{B}_0$ and $\mathbf{C}_0$.

Next, for each type a, the constant $\pi_{0(0a)}$ denotes the function of $D_{0(0a)}$ which, when applied to any element of D_{0a} (identified with a subset of D_a), produces the value T if the subset is the whole of D_a, and produces the value F otherwise. From this it follows that a wff of the form $\pi_{0(0a)}(\lambda\mathbf{x}_a\mathbf{B}_0)$ will refer to the truth value T if, and only if, $\mathbf{B}_0$ refers to T for *every* assignment of an element of D_a as a value for the free occurrences of the variable $\mathbf{x}_a$ in $\mathbf{B}_0$. Because of this intended meaning of the constant $\pi_{0(0a)}$, we may introduce universal quantifiers by writing $(\forall\mathbf{x}_a)\mathbf{B}_0$ for $\pi_{0(0a)}(\lambda\mathbf{x}_a\mathbf{B}_0)$, for any wff $\mathbf{B}_0$ and variable $\mathbf{x}_a$.

Using such a quantifier we can obtain, for each type a, a wff which, under the intended interpretation of $\mathcal{T}$ will denote the identity relation for elements of the type D_a; we shall introduce the symbol Q_{0aa} as an abbreviation for it,

[25] Hereafter we use subscripts to indicate the type of any wff under consideration. We abbreviate notation for wffs by omitting parentheses with the same convention as for type symbols (see preceding footnote).

and then write $\mathbf{B}_a = \mathbf{C}_a$ for $(\mathbf{Q}_{0aa}\mathbf{B}_a)\mathbf{C}_a$, where $\mathbf{B}_a$ and $\mathbf{C}_a$ are any wffs of type a. The formula for $\mathbf{Q}_{0aa}$ is $\lambda y_a \lambda z_a (\forall g_{0a})(g_{0a} y_a \supset g_{0a} z_a)$, so that the wff $(\mathbf{Q}_{0aa}\mathbf{B}_a)\mathbf{C}_a$ denotes the same element as $(\forall g_{0a})(g_{0a}\mathbf{B}_a \supset g_{0a}\mathbf{C}_a)$. Recalling that elements of type $(0a)$ have been identified with subsets of D_a, this wff expresses the fact that every subset of D_a containing the element B_a will also contain the element C_a, and this clearly has the value T if, and only if, B_a is the same element as C_a.

Finally, we come to the constants $\iota_{a(0a)}$ that have been provided in the formal language of $\mathcal{T}$; they are intended to play the role of selection operators when the language is interpreted. That is, for any type a, $\iota_{a(0a)}$ denotes a function which, acting on any argument z_{0a} (regarded as a subset of D_a), assigns to it an element of z_{0a} (if z_{0a} is not empty). It follows that if $\mathbf{B}_0$ is a wff containing free occurrences of some variable $\mathbf{x}_a$, and if *there is one and only one* element of D_a which, when assigned as value to these free occurrences of $\mathbf{x}_a$, produces the value T for the wff $\mathbf{B}_0$, then $\iota_{a(0a)}(\lambda \mathbf{x}_a \mathbf{B}_0)$ denotes that unique element of D_a. Thus, the notation $(\imath \mathbf{x}_a)\mathbf{B}_0$ is introduced as an abbreviation for $\iota_{a(0a)}(\lambda \mathbf{x}_a \mathbf{B}_0)$, and the part $(\imath \mathbf{x}_a)$ of this notation serves as a *description operator*.[26]

Having described the syntax of the formal language for $\mathcal{T}$, and having indicated the intended interpretation of that language, it remains to describe the formal deductive apparatus of axioms and rules of inference needed to qualify $\mathcal{T}$ as a logistic system.

There are six formal rules of inference. The first three describe the process of λ-conversion, allowing for a change of bound variable in a part of a wff, and for the replacement of a part $(\lambda \mathbf{x}_b \mathbf{M}_a)\mathbf{N}_b$ of a wff by the result of replacing all occurrences of $\mathbf{x}_b$ in $\mathbf{M}_a$ by occurrences of $\mathbf{N}_b$ (under suitable restrictions on free and bound variables), or *vice versa*. Then come familiar logical rules of substitution, detachment, and generalization.

As to the list of formal axioms, this begins with a standard set of four axioms for propositional calculus, followed by the following two axiom *schemas* for handling quantifiers with variables of any type a.

Schema 5^a. $(\pi_{0(0a)} f_{0a}) \supset (f_{0a} x_a)$.

Schema 6^a. $(\forall x_a)(p_0 \vee f_{0a} x_a) \supset (p_0 \vee \pi_{0(0a)} f_{0a})$.

These are all of the axioms needed for the logical functional calculus, although we could add the following *axiom schema of descriptions* to enable us to incorporate the logical use of the English word "the".

Schema 9^a. $(f_{0a} x_a) \supset [(\forall y_a)(f_{0a} y_a \supset x_a = y_a) \supset f_{0a}(\iota_{a(0a)} f_{0a})]$.

Of course you will wonder about Axioms (Schemas?) numbered 7 and 8. If they are not part of the logical functional calculus, what are they doing in

[26]When used in a context $(\imath \mathbf{x}_a)$ in Church's paper, the symbol $\imath$ is an inverted iota. In any case, as part of a description operator it functions like the English word "the".

this formal deductive system? And what are they? The answer is that Church wished to show how a logistic system can be applied to provide a foundation for mathematics, or at least for Peano arithmetic and real analysis. The Axioms 7 and 8 together have the effect of an axiom of infinity, Axiom 7 being essentially $\sim (\forall x_1)(\forall y_1)(x_1 = y_1)$, and Axiom 8 ensuring that D_1 cannot be finite. Schemas 10^{ab} and 11^{a} provide axioms of extensionality and choice, respectively.

Schema 10^{ab}. $(\forall x_b)[f_{ab}x_b = g_{ab}x_b] \supset (f_{ab} = g_{ab})$.

Schema 11^{a}. $f_{0a}x_a \supset f_{0a}(\iota_{a(0a)}f_{0a})$.

Of course in the presence of axioms 11^{a}, Schema 9^{a} could be dropped, as it is directly derivable from Axioms 1–4 and 11^{a}; but Church set down both 9^{a} and 11^{a}, for he thought it desirable to investigate the consequences of Axioms 1–9^{a} without 10^{ab} and 11^{a}.

Several features of the theory $\mathcal{T}$ sketched above were interesting to me, but I was especially attracted by the neatness and shortness of the formula expressing the axiom of choice. It seemed to me that the symbol $\iota_{a(0a)}$ was put into the formal language of $\mathcal{T}$ originally to serve the function of the definite article "the", as expressed in Axiom 9^{a}, and that its availability to provide such a succinct formulation of the axiom of choice was a fortuitous circumstance that must have come to Church as an inspired afterthought.

In describing the intended meanings of the formulas of the formal language, above, I have been careful, but not mathematically precise. In this I was following Church. I had never seen, in his courses, a Tarskian definition of truth for any formal language, nor is one given in the paper [Church, 1940]. Yet Church seemed to have a crystal clear vision of the structure of meanings for the language of $\mathcal{T}$, and indeed this was precisely the object of study in the course on sense and denotation in which I encountered this language.

I enjoyed the fact that the functional abstractor symbol, λ, enables us to name many elements in the hierarchy of types. This is in contrast to the simplified version of type-theory fashioned by Gödel from *Principia Mathematica*, which is called PM in [Gödel, 1931]. PM is essentially a pure monadic logical calculus of order ω, converted to an applied system by choosing the domain of individuals to be the natural numbers, adding constants to denote zero and the successor function, and using these to formulate Peano axioms that are added to the logical axioms in the deductive system. Of course formulas with free variables in PM are associated with sets and relations, and Gödel takes us through a long list of these; but there are no *names* for these.

I decided to try to see just which objects of the hierarchy of types did have names in $\mathcal{T}$. The natural place to start was in the two domains, D_0 and D_1, at the base of the hierarchy. Of course the two elements of D_0, T and F, had names, e.g., $(\pi_{0(00)} = \pi_{0(00)})$ and $N_{(00)0}(\pi_{0(00)} = \pi_{0(00)})$. But there were *no* names of individuals in D_1; and indeed, as long as there was no specification of a *particular* domain of individuals, it made no sense to ask for names of particular individuals. So I decided that for my project I would take D_1 to be the set of natural numbers, and I would add to the language of $\mathcal{T}$ a constant, 0_1, to serve as a name for the number 0, and a constant S_{11} to serve as a name for the successor function. In this I was following the example of Gödel when he set down the language for PM.[27]

Of course, with constants 0_1 and S_{11} added to the language of $\mathcal{T}$, and with D_1 now chosen to be the set of natural numbers, *every* element of D_1 has a name (containing 0_1 and repeated occurrences of S_{11}). Proceeding up the hierarchy of types, I quickly came to the type D_{01} whose elements, already identified with subsets of D_1, I knew could be used as real numbers under suitable definitions. On the basis of cardinality, *not* every element of D_{01} can be named by a wff, since there are only a denumerable number of the latter.

Before trying to make some sort of general survey of the nameable functions in the various types D_{ab}, I went over the recursive process for identifying the denotation of any wff. Everything seemed perfectly clear to me except for one thing: The function assigned as denotation to each of the constants N_{00}, A_{000}, $\pi_{0(0a)}$, and to the additional constant S_{11}, were completely definite, but in the case of the constants $\iota_{a(0a)}$ there was ambiguity. In discussing meanings, it was said only that the denotation of $\iota_{\alpha(0\alpha)}$ would be *some* choice function for the non-empty elements of $D_{0\alpha}$ (considered as subsets of D_α), but no particular one had been mentioned. As a result, the element of D_a named by a wff M_a containing some symbol $\iota_{b(0b)}$ would not be definitely determined. I wondered whether I could not remedy this "defect."

Of course there was no trouble in choosing $\iota_{1(01)}$ to be the function such that, for any element f_{01}, $\iota_{1(01)} f_{01}$ = the least element of f_{01} if there is *some* natural number x_1 for which $f_{01} x_1 = T$, and setting $\iota_{1(01)} f_{01} = 0_1$ if there is *no* such x_1. However, when I tackled the problem of describing some *particular* choice function to serve as the denotation of $\iota_{(01)(0(01))}$, I

[27] Gödel had available the material for a theory of natural numbers within the hierarchy of types in PM *without* dedicating the type of individuals to serve as natural numbers. Namely, he could have used the Frege–Russell definition of numbers as being particular sets of sets of individuals, i.e., elements of the third-level type. He explains in a note that the introduction of the Peano axioms for the type of individuals was only to simplify his exposition of the proof of incompleteness. Church, too, was interested in describing a development of the theory of numbers in $\mathcal{T}$, and the paper [Church, 1940] gives some details of this. He works with a definition of natural numbers which identifies them with particular functions of type $(11)(11)$ (abbreviated as $1'$); for example, the number 3 is identified with the function $3_{1'}$ such that, for any function f_{11}, $(3_{1'} f_{11})$ is the function $\lambda x_1(f_{11}(f_{11}(f_{11} x_1)))$. This definition of natural numbers is adapted from the system of λ calculus which Church used as a foundation for his work on undecidable theories. My decision to bring the Peano postulates for D_1 into $\mathcal{T}$ was specifically to create named objects in every type.

was stumped. Although I believed without question that there are choice functions which select an element from each non-empty set of real numbers, I saw no way to separate a single one of these choice functions from all the others, to serve as the denotation of the symbol $\iota_{(01)(0(01))}$.

Something about my failure to specify any one particular choice function for all non-empty sets of real numbers led me to think that perhaps the nature of the problem made it intrinsically unsolvable, and I began to wonder how I might possibly show that. Could I make precise what it might mean to say that it is "intrinsically impossible" to specify any particular choice function for non-empty sets of real numbers? No, I really couldn't. But I could make a precise weaker statement that would be pretty interesting, if I could prove it.

To state this, we need a definition. Keep in mind that a wff $\mathbf{M}_a$ without free variables denotes an element of D_a whose determination depends on which choice-functions for non-empty subsets of D_b are assigned as the denotations of the constants $\iota_{b(0b)}$ occurring in $\mathbf{M}_a$. Let us say that $\mathbf{M}_a$ denotes an element of D_a *absolutely* if it denotes that same element *no matter which* choice-functions are taken as the denotations of the constants $\iota_{b(0b)}$ occurring in it. Then I made the following *Conjecture: No choice-function for non-empty sets of real numbers is denoted absolutely by a wff* $\mathbf{M}_{(01)(0(01))}$ *without free variables*.

It must have been about mid-April, 1946, when I formulated this conjecture and set out to find a proof, intending this work to be the centerpiece of my dissertation. Mostly I worked in my head, with very few forays into the literature. One paper that I did read carefully, hoping to pick up ideas that would be relevant for my task, was Mostowski's paper on the independence of the axiom of choice for systems of set theory which admit "urelemente"; but in the end I did not find a way to use it.

While the conjecture I sought to prove is formulated in terms of the *absolute* denotations of wffs, my plan of action was to start with an arbitrary, but fixed, assignment of denotations to the constants $\iota_{a(0a)}$, and find out all I could about those elements of the type domains D_a that were named by some wff of the language for $\mathcal{T}$. I called these *the nameable elements* of the type hierarchy.

Although I worked for a year at finding out something about these nameable functions, I had very little success. I must have had at least two results, though, because I remember being asked to talk about my work at a department colloquium in September, 1946, and choosing the title "Nameable functions."

I do remember one small result that I found pretty early, and since it is important for our story, I want to describe it here.

For each type a, the nameable elements of type a form a certain subset, D^n_a, of the domain D_a. We observe that any element of D^n_{ab} is a function that maps D_b^n to D_a^n, so the set of all domains D_a^n itself forms a hierarchy of types. Looking at it, I thought I should make it a little neater by "trimming the fat" from each function in any domain D_{ab}^n. By this I meant that each element of D_{ab}^n has D_b, rather than D_b^n, as its domain, so I thought I should replace each element f of D_{ab}^n by f^*, the *restriction* of f to D_b^n, and then work with the resulting sets, say D_{ab}^{n*}, to get a neater representation of the hierarchy of types of nameable functions.[28]

There was, however, a problem with this idea: What if the hierarchy contracted under the proposed reduction of the domains of functions? In other words, could there be *distinct* functions f and g in some D_{ab}^n, such that $f^* = g^*$? After some worry, I realized that this could not happen. The reason is that if f and g are elements of D_{ab}^n, there are wffs $\mathbf{F}_{ab}$ and $\mathbf{G}_{ab}$ which denote them respectively. Then, letting $\mathbf{X}_{0b}$ be the wff $\lambda x_b \sim (\mathbf{F}_{ab} x_b = \mathbf{G}_{ab} x_b)$, we see that if $f \neq g$ then $\mathbf{X}_{0b}$ denotes a non-empty subset of D_b, so that $\imath_{b(0b)} \mathbf{X}_{0b}$ denotes an element y of D_b^n for which $fy \neq gy$, showing that $f^* \neq g^*$.

With the realization that the passage from functions in D_a^n to those in D_a^{n*}, by reduction of domains, does not result in contraction, I had a self-contained hierarchy of type domains D_a^{n*} which, in an obvious sense, is isomorphic to the hierarchy of nameable functions within the original domains D_a.[29] Henceforth I concentrated my effort to show the non-existence of absolutely nameable choice functions for non-empty sets of real numbers, by searching within the hierarchy of domains D_a^{n*}.

March, 1947, arrived, the anniversary of my return to Princeton from war-time work, and I had gotten nowhere. I began increasingly to worry about how and when people would react to my continuing non-productive work. The National Research Council was supporting me with a fellowship; perhaps they would ask for a progress report before long. The Chairman of the Mathematics Department, Solomon Lefschetz, and my teacher, Church, must be expecting to hear from me about some concrete discoveries or directions for my dissertation, but I had not spoken to them about that subject for several months. And when my father asked how my work was proceeding during my visits to New York every few weeks, I had only one answer to repeat, "Well, I'm working hard!"

One night I lay in my bed in the Graduate College going over these worries. They became increasingly intense; I did not see how I could deal with them. Suddenly I noticed that my arms and legs were rigid, my throat constricted, and I had the impression that I was on the verge of screaming! At once I realized that I could not continue as I had been doing. I decided that to

[28] Of course $D_0^n = D_0$ and $D_1^n = D_1$, so we may as well set $D_0^{n*} = D_0$ and $D_1^{n*} = D_1$, too.

[29] To be technically correct we should complicate our description of the domains D_a^{n*} by using recursion over type symbols, but for present purposes we overlook this.

avoid a breakdown I would discontinue my graduate study and try to find some sort of routine work. Lying in bed I composed letters to Professor Lefschetz and to my father, explaining why I was leaving Princeton, and I determined to write and send these in the morning.[30] This process relaxed me, and I was able to fall asleep.

The next morning I remembered clearly my decision taken the night before, and fully intended to carry it out. However, I saw that there was no *immediate* need to write and send off those letters, and as I had a new idea about nameable functions which looked rather interesting I thought I might as well check that out first, so I put the letters off for a day or two.

A few weeks later I was sitting in the armchair in my study at the Graduate College, trying to "see" more clearly, for the hundredth time, the structure of the functions in the hierarchy of type-domains D_a^{n*}. I recall that I was sitting in an unusual position, with my right leg thrown over the arm of the chair, and my head bent over the other arm of the chair. I thought that if I could only get a clearer *picture* of the interaction of the functions in the hierarchy that that might help me toward my goal of seeing that there cannot be absolute choice functions for non-empty sets of real numbers.

Since each function in one of the domains D_{ab}^{n*} has a name among the wffs of type ab, I would try to visualize one of these functions, f, by picturing, in my mind, a generic wff $\mathbf{F}_{ab}$ that denotes it. Then, to visualize how f acts on some argument m from D_b^{n*}, I would take a formula $\mathbf{M}_b$ denoting m, and get the formula $\mathbf{F}_{ab}\mathbf{M}_b$ with which to visualize the value of f at m, in D_a^{n*}. But to see how *this* element of D_a^{n*} was related to *others*, I would suppose that $\mathbf{F}_{ab}$ had the form $\lambda x_b \mathbf{N}_a$, and then apply the formal rule of λ-conversion to express $\mathbf{F}_{ab}\mathbf{M}_b$ by substituting $\mathbf{M}_b$ for free occurrences of x_b in $\mathbf{N}_a$.

As I struggled to see the action of functions more clearly in this way, I was struck by the realization that I had used λ-conversion, one of the formal rules of inference in Church's deductive system for the language of the theory $\mathcal{T}$. All of my efforts had been directed toward *interpretations* of the formal language, and now my attention was suddenly drawn to the fact that these were related to the *formal deductive system* for that language. In particular, I saw that using the symbol $\vdash$ for formal provability (or derivability) as usual, we can define for each type symbol a, a domain D'_a satisfying the following conditions: (i) Each cwff (closed wff, without free variables) $\mathbf{M}_a$ denotes an element $\mathbf{M}'_a$ of D'_a, and each element of D'_a is denoted by some cwff $\mathbf{M}_a$; (ii)

[30]The letter to my father was especially difficult to compose, for he had praised my little school successes extravagantly ever since I was in first grade, and would bore friends and family members by repeating a roster of my "accomplishments" on every occasion. In fact, he had shown his high expectations for me at the time of my birth by choosing my middle name to be "Albert." He once told me that at that time (April 1921) the New York *Times* had run a series of articles publicizing Einstein's revolutionary theory of relativity, so my father decided to borrow Einstein's first name for his newborn son.

for any cwff $\mathbf{F}_{ab}$, $\mathbf{F}'_{ab}$ is a function mapping D'_b into D'_a; and (iii) for any cwffs $\mathbf{M}_a$ and $\mathbf{N}_a$, $\mathbf{M}'_a = \mathbf{N}'_a$ if, and only if, $\vdash (\mathbf{M}_a = \mathbf{N}_a)$.

To accomplish this, we begin by defining D'_a for the case where $a = 0$ or $a = 1$, so that its elements are equivalence classes $[\mathbf{M}_a]$ consisting of all cwffs $\mathbf{N}_a$ such that $\vdash \mathbf{M}_a = \mathbf{N}_a$. Then, proceeding by induction and supposing that D'_a and D'_b have been defined, we put into D'_{ab}, for each cwff $\mathbf{F}_{ab}$, the function $\mathbf{F}'_{ab}$ (from D'_b to D'_a) such that, for any cwff $\mathbf{M}_b$, $\mathbf{F}'_{ab}\mathbf{M}'_b = (\mathbf{F}_{ab}\mathbf{M}_b)'$. Assurance that $\mathbf{F}'_{ab}$ is well defined by this equation comes from the induction hypothesis that D'_b satisfies (iii). Assurance that D'_{ab} satisfies (iii) comes from the fact that if $\mathbf{F}_{ab}$ and $\mathbf{G}_{ab}$ are cwffs such that $\mathbf{F}'_{ab} = \mathbf{G}'_{ab}$, then

$$(\mathbf{F}_{ab}(\iota_{b(0b)}(\lambda x_b \sim (\mathbf{F}_{ab}x_b = \mathbf{G}_{ab}x_b))))' = (\mathbf{G}_{ab}(\iota_{b(0b)}(\lambda x_b \sim (\mathbf{F}_{ab}x_b = \mathbf{G}_{ab}x_b))))'.$$

Since D'_a satisfies (iii) by induction hypothesis, this gives

$$\vdash \mathbf{F}_{ab}(\iota_{b(0b)}(\lambda x_b \sim (\mathbf{F}_{ab}x_b = \mathbf{G}_{ab}x_b))) = \mathbf{G}_{ab}(\iota_{b(0b)}(\lambda x_b \sim (\mathbf{F}_{ab}x_b = \mathbf{G}_{ab}x_b))).$$

From this, by Axiom Schema 11^b, we get $\vdash (\forall x_b)(\mathbf{F}_{ab}x_b = \mathbf{G}_{ab}x_b)$, and so by Axiom Schema 10^b, $\vdash \mathbf{F}_{ab} = \mathbf{G}_{ab}$, as desired.

Notice how this last proof parallels the earlier proof that for f and g in D^n_{ab}, we have $f = g$ if, and only if, $f^* = g^*$. I had put some effort into finding the earlier proof, so now I saw the facts about the domains D'_a very swiftly.

The actions of the functions in D'_{ab} and D^{n*}_{ab} are so similar, that at first I thought that the two hierarchies might be identical. But as soon as I compared D'_0 with D^{n*}_0, I saw that these were very different, and that this would produce differences in the two hierarchies at every level. The reason is that D^{n*}_0 is simply D_0 (by Footnote 29), so has only two elements, while D'_0 has many elements. (In particular, if $\mathbf{M}_0$ is a Gödel sentence such that neither $\vdash \mathbf{M}_0$ nor $\vdash\sim \mathbf{M}_0$, then $(0_1 = 0_1)'$, $(\sim 0_1 = 0_1)'$, and $\mathbf{M}'_0$ are three distinct elements of D'_0.)

As soon as I observed this, it occurred to me that if we were to add further cwffs of type 0 to the list of formal axioms, this would have the effect of reducing the number of elements in D'_0 and that ultimately, by taking a maximal consistent set of axioms, the number of elements in D'_0 would be two. At that point, if we were to start with the resulting hierarchy of domains D'_a and create corresponding hierarchies of domains D^n_α, and D^{n*}_α, the three hierarchies would all be identical.

In short, I had simultaneously formulated Theorem VI of my dissertation, and discovered its proof, in a period of less than half an hour, while trying to "see more clearly" the hierarchy of nameable type domains D^{n*}_a with which I had been struggling for a year.

Immediately I realized that my discovery provided a kind of completeness proof for a system very much like the system PM of type theory which Gödel had proved incomplete. The fear of having nothing to show for my year-long dissertation work was lifted from my spirit, and for a couple of days I was euphoric. Then I came back to look at my new-found completeness theorem and see if I could find something else to put into my dissertation with it.

The very first question I asked myself was whether I could use the method that gave me completeness for type theory, to get a new proof of Gödel's completeness for first-order logic. It seemed, at first, that there was no possibility to do so. The reason is that the axiom of choice, and in particular Church's neat formulation of it via the constants $\iota_{a(0a)}$, played a crucial role in the proof for type theory, while in first order logic there is no axiom of choice, and no way to formulate one.

I decided to analyze carefully the role of the axiom of choice in the completeness proof, to see whether there was some other way of accomplishing it in first-order logic. The role that I saw first was performed in the proof by induction, sketched above, that domains D'_α satisfying conditions (i)–(iii), could be constructed. But when I wrote down details of the proof that the resulting hierarchy of domains D'_α satisfy the maximal consistent set of (added) axioms, I saw that the axiom of choice is needed *there* in a more general way, of which the earlier use is just a special case. The more general need is to show that whenever we have a wff $\mathbf{M}_0$ such that $\vdash (\exists x_b)\mathbf{M}_0$, then we also have $\vdash (\lambda x_b \mathbf{M}_0)(\iota_{b(0b)}(\lambda x_\beta \mathbf{M}_0))$. The fact that this condition holds is a direct consequence of having Axiom Schema 11^b in the deductive system that Church had set up for the theory $\mathcal{T}$, as that schema is trivially equivalent to $(\exists x_b f_{0b} x_b) \supset f_{0b}(\iota_{b(0b)} f_{0b})$.

It did not take me very long to notice that in fact, the form of the wff following $(\lambda x_b \mathbf{M}_0)$ played no role in the completeness proof; it is only necessary to have *some* cwff $\mathbf{N}_b$ such that $\vdash (\lambda x_b \mathbf{M}_0)\mathbf{N}_b$ holds if $\vdash (\exists x_b)\mathbf{M}_0$ holds. That immediately suggested to me the adjunction of new constants u_b to the language of $\mathcal{T}$ to play the role of these needed cwffs $\mathbf{N}_b$, and it was obvious that *that* process could be carried over to first-order logic. So I had my proof of Theorem I of my dissertation.

The proof of Theorem I shows how, starting with a consistent set S of formal sentences of a first-order language, one can obtain a model $\mathcal{M}$ satisfying S by using newly adjoined individual constants for the elements of $\mathcal{M}$. Conversely, starting with a structure $\mathcal{M}$ for a first-order language, the set of all sentences true of $\mathcal{M}$ is a (maximal) consistent set; and if we adjoin individual constants to the language to serve as names for the elements of $\mathcal{M}$, then for each true sentence of form $\exists x Fx$ there will be a constant u such that Fu is true of $\mathcal{M}$. Observing how one can go back and forth between consistent sets S of sentences and models of S, led me to my next discoveries—Theorem II of my dissertation (extending Theorem I to non-denumerable languages), its important Corollary (compactness), and my proof of Theorem III (representing Boolean algebras).

I had never encountered the concept of a Boolean algebra in courses on logic or algebra, but I heard about them at dinner one evening (early in 1947) from a fellow graduate student, Gilbert Hunt, who later became a distinguished probabilist. He was very excited about Stone's representation theorem, which he'd just found, so I got some background about it by browsing in [Birkhoff, 1948] and then read [Stone, 1936]. It seems to me most likely that Stone's construction of maximal ideals in Boolean algebras was the inspiration for my construction of maximal consistent sets of cwffs, when I wrote up the completeness for $\mathcal{T}$. However, it was not (consciously) in my mind at the moment when I thought of adding new formal axioms to $\mathcal{T}$ in order to reduce the number of equivalence classes in D_0' to two; *that* thought came to me as part of a visualizing process, rather than a reasoning one.

At any rate, soon after finding my proof of Theorem I and playing back-and-forth between maximal consistent sets (with witness-constants) and models, I noticed two facts. First, I saw that it would be natural to deal with languages having non-denumerably many constants. And second, I saw the compactness principle. However, it was only after I had found that I could accomplish something by using these facts, that I decided to incorporate them in my thesis by formulating Theorem II and its Corollary. The accomplishment consisted in my putting together the metamathematical proof of Stone's representation theorem that is given for Theorem III of the dissertation.

After finding the idea of the proof of Theorem III, I went back and stated Theorem II and its corollary in order to justify some of the steps in the proof. Once I had written up Theorem II and its Corollary, I saw that I might as well use them to formulate Theorem IV, which gives a generalized condition for embedding one structure into another that possesses a given elementary property.

The last application of compactness in my dissertation is Theorem V and its Corollaries I and II. These relate fields of characteristic 0 to fields of prime characteristic. The subject of characteristics was being considered in a course on rings and fields given by Emil Artin which I was attending in Spring 1947, which explains how I came to try out my new toy, compactness, in that direction.

This completes my account of how I discovered a new proof of completeness and began to use it to obtain early results in model theory. To conclude this paper I shall set down, below, three observations.

A. In 1951 a second printing was made of [Gödel, 1940], and the author took that occasion to append ten notes to the original text. Note 1 is to the effect that from the axiom that all sets are constructible ($V = L$), which he showed to be consistent with the other axioms of set theory, it follows that there is a *projective* well-ordering of the real numbers. It follows that the existence of a nameable choice function for non-empty sets of real numbers, which I was trying to prove *false*, is in fact *consistent* with what we now call the Gödel–Bernays axioms for set theory. Had I realized this in 1946, I would probably never have started to work on the problem that led to my discoveries.

After Paul Cohen proved the independence of the axiom of choice by introducing the method of forcing, that method was used by Solomon Feferman to show that it is consistent with the axioms of set theory, *including* the axiom of choice, that *there is no formula of the Gödel–Bernays set theory theory which defines a well-ordering of the real numbers*. From this it follows that the conjecture that I fruitlessly tried to prove true, is at least consistent. See [Feferman, 1965].

B. Writing this account of the origin of my completeness proofs has made me wonder about the task of historians of mathematics. In part they have to describe faithfully the order of mathematical discoveries and how these were, or were not, propagated, but in other part they must make hypotheses by means of which the observed facts are in some sense *explained*. I see that it would be exceedingly difficult for an historian who did not learn of the story I have told above, to formulate an accurate hypothesis of how I found my proof of completeness of first-order logic. And I recognize that part of the difficulty arises from the fact that my method of writing up the dissertation *hides* the process of discovery.

For one thing, completeness for first-order logic comes in Part I of the dissertation, and completeness for the theory of types is given in Part III. It would be hard enough to guess that the latter proof was discovered first, and led me to the former proof—but I made it even harder, as I shall explain in a moment. But let me first say that my *reason* for treating first-order logic first, in the dissertation, was in part because the logical calculus was simpler and much more widely known. In other part, I felt that the *result* of completeness for type theory would be of much greater interest (insofar as it gave a semantical characterization of the formally undecidable sentences of that theory), so I wanted to make it the climax of the dissertation rather than put it first.

As indicated, in addition to putting the first-discovered proof into Part III of the dissertation, I hid the discovery process in *another* way. Namely, in setting forth the formal language for type theory, in Part III, I *deleted* from Church's system the symbols $\iota_{b(0b)}$ and the axiom of choice for which they were used! Those symbols, and that axiom, which played such a quintessential role in the discovery process, were omitted in formulating Theorem VI, and the role of the symbols $\iota_{b(0b)}$ in forming "witnesses" for cwffs of the form $\exists \mathbf{x}_b \mathbf{M}_0$ was taken over by the adjunction of constants $u_b, u'_b, u''_b, \ldots$ for each type symbol b. Again, there was a perfectly sound reason for making this change, namely, it *strengthened* the scope of Theorem VI, for in the altered form it implies that if the *negation* of the axiom of choice is consistent, then there is a general model for which the axiom of choice is false; and this would not follow from the original version of Theorem VI that I discovered.

I did not *altogether* hide the symbols $\iota_{b(0b)}$ from the reader of my dissertation, for in passing from Theorem VI to Theorem VII, in which the generalized completeness theorem is extended to languages that can be non-denumerable or have additional constants, I cited Church's system in which the axiom schemas of description and choice are formulated with the symbols $\iota_{b(0b)}$, as an example. With that example is a brief note to the effect that when this formulation of the axiom of choice is made, the adjunction of special constants $u_b, u'_b, u''_b, \ldots$ for each type symbol b becomes unnecessary in the completeness proof, as their role can be taken over by the symbols $\iota_{b(0b)}$; still, it would be a very sharp-eyed historian who could detect in that brief note appended to Theorem VII, the origin of the proof of Theorem I!

Church's use of the symbols $\iota_{b(0b)}$ led me to formulate my *conjecture* (about the non-existence of an absolutely definable choice function for non-empty sets of real numbers), which can really be formulated without special symbols as is done in [Feferman, 1965]. It was my year-long and fruitless effort to prove that conjecture which led me to my completeness proof, yet the conjecture is not mentioned anywhere in the dissertation.

I do not believe that my crimes against historical discovery are isolated. It is my impression that *many* mathematical papers are written in a fashion that tends to obscure the process of discovery. Tendencies to find and exhibit *neat* proofs often result in the suppression of *first* proofs, and the thirst for *general* results can squeeze out special cases which may have led the way to discovery.

What are historians of mathematics to do about mathematical discoveries for which the process of discovery is not revealed by the discoverer? I would like to suggest that historians take such discovery processes as challenges!

Many papers dealing with the history of mathematics attempt to describe the development of some subject in a way that can be schematized by the diagram of a directed graph. The nodes, or vertices, of this graph, are the publications mentioned in the historical paper. A directed line–segment of the graph leads from vertex u to vertex w in case the author of publication w is thought to have been influenced by publication u in the course of writing

w. Normally, the *way* in which the author of publication w arrived *at its ideas* by combining those of u with those of other prior publications—that is, the *process of discovery* of the ideas of w—is not treated.

What I am suggesting is that in the diagram of a work on history of mathematics, each vertex be expanded to a "black box" of the kind mentioned in the introduction of this paper. The directed line–segments leading into the node become the inputs of the box, those leading out of the node become the outputs. This proposal, if adopted, would signify a recognition that a complete historical account of some mathematical development should include not only a set of publications and their interrelations, but some account of the process of discovery (inside the "black box") that resulted in each of the publications cited.

The existence of the publications and their interrelations can be confirmed by empirical evidence of the kind we now find in papers on the history of mathematics, but usually there is no (or very little) such evidence concerning the discovery process by which the findings of a mathematical publication were generated from its input ideas. How, then, is the historian to fill these gaps?

It seems to me that the situation is analogous to what we find in various sciences where "black boxes" appear as components of physical, biological, or social processes. Perhaps the most familiar example is found in atomic physics, where observed facts are accounted for by a theory involving hypothesized objects entering hypothesized interactions that are subject to specified constraints. Such hypotheses are retained as long as they explain confirmed observations, and they are suspended when new observations lead to a search for new hypotheses.

By seeking hypothetical explanations of the discovery *process* for important discoveries observable in publications, letters, and other accounts now employed, the history of mathematics would deepen its scientific character.

C. It seems that a distinctive feature of my completeness proof for first-order logic, which distinguishes it from Gödel's, is that when a consistent set of cwffs is given in one language, I proceed to an extended language in which new individual constants are adjoined. But in fact, something like that is implicitly present in Gödel's proof, because he begins by reducing the problem of showing that an *arbitrary* cwff is either satisfiable or refutable, to the case of an arbitrary cwff *that is in Skolem normal form*. However, in a first-order language with some fixed finite set of predicate symbols, one cannot reduce every cwff to one in Skolem normal form *without adding new predicate symbols*.

What is necessary in my proof is to start with a consistent set S of cwffs of a given language $\mathcal{L}$, and to extend S to a larger set S' having two properties: (i) S' is maximal consistent, and (ii) whenever $S' \vdash \exists x \mathbf{F}x$ for some wff $\mathbf{F}x$, we also have $S' \vdash \mathbf{F}u$ for some individual constant u. In my dissertation I accomplished this by an infinite sequence of extensions of $\mathcal{L}$, each such extension involving the adjunction of an infinite sequence of new constants. However, in teaching logic courses I discovered that it sufficed to adjoin a single infinite sequence of new constants to $\mathcal{L}$, and then to form S' by starting with S and making a sequence of enlargements by a single cwff, interweaving cwffs of the form $(\exists x\mathbf{F}x) \supset \mathbf{F}u$ with cwffs selected from each pair $(\mathbf{M}, \sim \mathbf{M})$. This method has found its way into the literature through [Hasenjaeger, 1953]; see Footnote 3 of that work, and Footnote 513 in [Church, 1956], p. 311.

In fact, in my later logic courses I found that the process of adding cwffs to S in order to achieve an S' satisfying condition (ii) can be described still more simply. Namely, it suffices to put into S' an element $\mathbf{G}u$ (for some constant u) whenever $\mathbf{G}x$ is a wff such that $\vdash (\exists x)\mathbf{G}x$. This is because, for any wff $\mathbf{F}x$, we may take $\mathbf{G}x$ to be the formula $(\exists x\mathbf{F}x) \supset \mathbf{F}x$, and easily show that $\vdash (\exists x)\mathbf{G}x$ holds.

Added January 3, 1996: A manuscript of this paper was prepared in Fall, 1993, and was sent to Professor Church. Arrangements were made to include the paper in two books, a Proceedings of the symposium at which it was presented (see Footnote 1), and a Festschrift for Church under preparation by two of his former students. Neither of these books materialized, for lack of timely agreements between editors and publishers. Church died in August 1995, and a few months later I submitted the paper to the Bulletin of Symbolic Logic.

REFERENCES

[1948] G. Birkhoff, ***Lattice theory***, American Mathematical Society, New York, 1948.

[1940] A. Church, *A formulation of the simple theory of types*, ***The Journal of Symbolic Logic***, vol. 5 (1940), pp. 56–68.

[1951] ———, *A formulation of the logic of sense and denotation*, ***Structure, method and meaning; essays in honor of Henry M. Sheffer***, New York, 1951, pp. 3–24, revisions of the system of this paper were published in ***NOUS***, vol. 7 (1973), pp. 24–33; vol. 8 (1974), pp. 135–156; and vol. 27 (1993), pp. 141–157.

[1956] ———, ***Introduction to mathematical logic, I***, Princeton University Press, Princeton, 1956.

[1965] S. Feferman, *Some applications of the notions of forcing and generic sets*, ***Fundamenta Mathematicae***, vol. 56 (1965), pp. 325–245.

[1986] S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort (editors), ***Kurt Gödel collected works***, vol. I, Oxford University Press, Oxford, 1986.

[1992] S. Feferman, J. W. Dawson, Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort (editors), ***Kurt Gödel collected works***, vol. II, Oxford University Press, Oxford, 1992.

[1929] K. GÖDEL, *Über die Vollständigkeit des Logikkalküls*, ***Doctoral dissertation***, University of Vienna, 1929, reprinted in ***Kurt Gödel Collected Works***, vol. I, pp. 60–100.

[1930] ———, *Die Vollständigkeit der Axiome des logischen Funktionenkalküle*, ***Monatshefte für Mathematik und Physik***, vol. 37 (1930), pp. 349–360.

[1931] ———, *Über formal unentscheidbare Sätze der* Principia mathematica *und verwandter Systeme I*, ***Monatshefte für Mathematik und Physik***, vol. 38 (1931), pp. 173–198.

[1940] ———, *The consistency of the axiom of choice and the generalized continuum-hypothesis with the axioms of set theory*, ***Annals of Mathematics Studies No. 3***, Princeton University Press, Princeton, 1940.

[1953] G. HASENJAEGER, *Eine Bemerkung zu Henkin's Beweis für die Vollständigkeit des Prädikaten-kalküls der ersten Stufe*, ***The Journal of Symbolic Logic***, vol. 18 (1953), pp. 42–48.

[1949] L. HENKIN, *The completeness of the first-order functional calculus*, ***The Journal of Symbolic Logic***, vol. 14 (1949), pp. 159–166.

[1950] ———, *Completeness in the theory of types*, ***The Journal of Symbolic Logic***, vol. 15 (1950), pp. 81–91.

[1953] ———, *Some interconnections between modern algebra and mathematical logic*, ***Transactions of the American Mathematical Society***, vol. 74 (1953), pp. 410–427.

[1928] D. HILBERT and W. ACKERMANN, ***Grundzüge der theoretischen Logik***, Springer-Verlag, Berlin, 1928.

[1958] E. NAGEL and J. R. NEWMAN, ***Gödel's proof***, New York University Press, New York, 1958.

[1938] W. V. QUINE, *Completeness of the propositional calculus*, ***The Journal of Symbolic Logic***, vol. 3 (1938), pp. 37–40.

[1903] B. RUSSELL, ***Principles of mathematics***, Cambridge University Press, Cambridge, 1903.

[1936] M. H. STONE, *The theory of representation for Boolean algebras*, ***Transactions of the American Mathematical Society***, vol. 40 (1936), pp. 37–111.

[1954] A. TARSKI, *Contributions to the theory of models, I, II, III*, ***Indagationes Mathematicae***, vol. 16 (1954), pp. 572–581, 582–588.

[1955] ———, *Contributions to the theory of models, I, II, III*, ***Indagationes Mathematicae***, vol. 17 (1955), pp. 56–64.

[1910] O. VEBLEN and J. W. YOUNG, ***Projective geometry***, vol. I, Ginn and Company, New York, 1910.

[1918] ———, ***Projective geometry***, vol. II, Ginn and Company, New York, 1918.

[1910] A. N. WHITEHEAD and B. RUSSELL, ***Principia mathematica***, vol. 1, Cambridge University Press, Cambridge, 1910.

[1912] ———, ***Principia mathematica***, vol. 2, Cambridge University Press, Cambridge, 1912.

[1913] ———, ***Principia mathematica***, vol. 3, Cambridge University Press, Cambridge, 1913.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CA 94720, USA

E-mail: henkin@math.berkeley.edu