# NEW FOUNDATIONS FOR MATHEMATICAL THEORIES

JAAKKO HINTIKKA

**§1. The motivation.** In this paper, I shall outline a new approach to the logical foundations of mathematical theories. One way of looking at its motivation is as follows (I am following here Hintikka, 1989):

In the foundational work around 1900, e.g. in Hilbert's ***Foundations of geometry***, a crucial role was played by assumptions of *extremality* (i.e., *minimality* and *maximality*). For instance, Hilbert's so-called axiom of completeness is a maximality assumption. The Archimedean axiom can be thought of as a minimality assumption, the principle of induction likewise as a minimality axiom, and Dedekind's assumption of the existence of a real for each cut as a maximality assumption. Slowly, it has become clear to everybody that such extremality assumptions cannot normally be expressed as ordinary first-order axioms. To what extent they can or cannot be expressed in other ways, e.g. as higher-order axioms or set-theoretical axioms, and to what extent we should try to express them in such ways, will not be discussed here. In any case, in spite of the tremendous *prima facie* interest and power of extremality assumptions, they have not attracted much interest lately.

The approach proposed and outlined here relies crucially on extremality assumptions but seeks to implement them in a new way on a first-order level. Instead of introducing extremality assumptions on the top of a ready-made logic as explicit axioms, I propose to build them into the very logic we are employing, thus by-passing the difficulties the earlier uses of extremality assumptions encountered.

A logic is in effect specified by a space $\Omega$ of models together with a definition of what it means for a statement (closed formula) to be true in a model $M \in \Omega$ (and for a formula to be satisfied with in $M$). I shall not modify the latter ingredient. Instead, I propose to modify the usual space of models (of a given first-order language $L$) in the simplest possible way, viz. by omitting some of its members.

Even though this kind of modification looks innocuous, it facilitates a radical new look at the prospects of mathematical and logical theories. Most importantly, the possibility of reaching completeness can be profoundly affected.

What are the different kinds of completeness relevant here? Here are four candidates, which have not always been distinguished from each other sufficiently clearly:

(1) *Descriptive completeness.* It is an attribute of a *non-logical theory.* It means that the theory has as its models only the intended (standard) ones, i.e., that it has no non-standard ones. If there is only one standard

model, a descriptively complete theory must be categorical. Here the notion of standardness has to be characterized independently.

(2) *Semantical completeness.* It is a property of an *axiomatization of* (some branch of) *logic.* It means that the theorems of the axiomatization exhaust all valid formulas, where validity means truth in all the models of the space $\Omega$. Thus semantical completeness amounts essentially to the recursive enumerability of the set of valid formulas.

(3) *Deductive completeness.* It is a property of a *non-logical axiom system together with an axiomatization of logic.* It means that, for each statement $C$, either $C$ or $\sim C$ can be derived from the (non-logical) axioms by means of the given logic.

(4) *Hilbert's* so-called *axiom of completeness* is in effect a maximality assumption in a non-logical axiom system. While Hilbert's own intentions are not clear, this axiom can be taken to say that one cannot add new individuals to a (standard or intended) model without violating other axioms.

What happens when the space of models $\Omega$ is replaced by some $\Omega^* \subset \Omega$?

(1) Descriptive completeness becomes *ceteris paribus* easier to reach since some (or all) of non-standard models in $\Omega$ may belong to $\Omega - \Omega^*$.

(2) Since there are fewer models, there are *ceteris paribus* more valid formulas (i.e., formulas true in all of them). Hence semantical completeness can become more difficult to achieve.

In other words, by moving from $\Omega$ to a suitable $\Omega^*$ ($\Omega^* \subset \Omega$) we can trade in the semantical completeness of our underlying logic in order to achieve the descriptive completeness of suitable mathematical theories. I have argued elsewhere that this would represent a major gain in philosophical and conceptual clarity. (See Hintikka 1989.)

For example, Gödel showed elementary arithmetic to be incomplete in the sense (3) (deductive incompleteness). From this it does not by itself follow that elementary arithmetic is descriptively incomplete. This does follow if the underlying logic is complete, which Gödel had proved (for first-order logic) prior to proving the the incompleteness of elementary arithmetic. However, if we are willing to change this logic (strengthen it) so as to render it semantically incomplete, we can very well hope to reach a descriptively complete first-order theory of arithmetic. This in fact turns out to be possible.

More generally, by means of suitable extremality restrictions on models, it will turn out to be possible to formulate categorical first-order axiom systems *inter alia* for elementary number theory, the theory of reals, Euclidean geometry, and the second number class (countable ordinals). (See §5 below.)

(3) Deductive completeness, being a kind of combination of descriptive and semantical completeness, is not necessarily affected by any trade-off between the other two kinds of completeness (1)–(2).

(4) The restriction of the space of models to a suitable subset serves the same purpose as the axiom of completeness, and is supposed to replace any such explicit axiom.

**§2. Model theory via constituents.** The crucial question obviously is how the ideas of minimality and maximality can be implemented in precise and general terms. Since we are speaking of the minimality and maximality of models (or parts thereof), the obvious resource here is the theory of models. I shall review some of the basic ideas of model theory, but for the sake of certain further developments I shall do so in an unfamiliar way. I shall use in the review the technique of constituents and distributive normal forms. Even though this technique may in the last analysis be dispensable in favor of more commonly employed conceptualizations (e.g. back-and-forth techniques), it offers heuristic advantages by allowing an almost geometrical (tree-theoretical) visualization of the logical relationships under scrutiny in this paper.

An approach to model theory via constituents leads us straight to the field of stability theory. However, it is in fact quicker and much more perspicuous to develop the necessary theory directly here without going by way of stability theory. (For stability theory, see e.g. Baldwin, 1988.)

In what follows, it is assumed that we are dealing with a given first-order language $L$ with a finite list of predicate constants, no function symbols, and an unspecified supply of individual constants. I shall deal only with languages without identity. It turns out that the presence or absence of identity does not matter very much for the central purposes of this work.

A constituent $L$ with the free variables $x_1, x_2, \ldots, x_k$ will be expressed as follows:

$$C_i^{(d)}[x_1, x_2, \ldots, x_k] \tag{2.1}$$

It is a well-formed formula which has a number $k$ of arguments $x_1, x_2, \ldots, x_k$ and it is also characterized by its *depth d*. The subscript $i$ serves to distinguish different constituents with the same arguments and with the same depth from each other.

Constituents can be defined recursively as follows:

$$C^{(0)}[x_1, x_2, \ldots, x_k] \tag{2.2}$$

is of the form

$$\bigwedge_{j \in J} (\pm_j) A_j[x_1, x_2, \ldots, x_k] \tag{2.3}$$

where the $A_j[x_1, x_2, \ldots, x_k]$, $j \in J$, are all the different atomic formulas that can be formed from the predicate constants of $L$ and of $x_1, x_2, \ldots, x_k$, and where each $(\pm_j)$ is either $\sim$ or nothing, depending on $j$.

$$C_i^{(d+1)}[x_1, x_2, \ldots, x_k] \tag{2.4}$$

is of the form

(2.5)

$$\bigwedge_{j \in J} (\exists y) C_j^{(d)}[y, x_1, x_2, \ldots, x_k] \ \& \ (\forall y) \bigvee_{j \in J} C_j^{(d)}[y, x_1, x_2, \ldots, x_k] \\ \& \ C_l^{(0)}[x_1, x_2, \ldots, x_k].$$

Here the last conjunct is simply some one constituent without quantifiers with $x_1, x_2, \ldots, x_k$ as its only free individual symbols. The index set $J$ is a subset of the set of the subscripts of all the different constituents

$$C^{(d)}[y, x_1, x_2, \ldots, x_k]. \tag{2.6}$$

Intuitively, a constituent like (2.5) of depth $d+1$ tells us what kinds of individuals there exist (in relation to $x_1, x_2, \ldots, x_k$) and do not exist. The latter is accomplished in the universally quantified disjunction of (2.5) by saying that each individual must be of one of the kinds listed in the first of the three conjuncts in (2.5). Here the "kinds of individuals" are in turn specified by constituents (2.6) of a lesser depth $d$.

Each constituent (2.5) thus has a tree structure where the nodes of this labeled tree are constituents of increasingly smaller depth each occurring in its predecessor. Intuitively, each branch of such a tree describes a sequence of $d+1$ individuals that you can find in a model of $L$ in which (2.5) is true. The tree structure show how the initial segments of such sequences limit their possible continuations.

In a sense, a constituent thus presents an explicit description of certain salient structural features of a model $M$ in which it is true. The constituent tells you which (ramified) sequences of individuals (up to the length $d+1$) you can hope to find in a model in which it is true.

In this work, the term "constituent" will also be applied to substitution-instances of (2.4) with respect to the individual constants of $L$, i.e., to formulas like

$$C^{(d+1)}[a_1, a_2, \ldots, a_k] \tag{2.7}$$

or

$$C^{(d)}[y, a_1, a_2, \ldots, a_k]. \tag{2.8}$$

If identity is present in $L$, the definition of a constituent can be changed as follows: (2.2) is now of the form

$$\bigwedge_{j \in J} (\pm_j) A_j[x_1, x_2, \ldots, x_k] \;\&\; \bigwedge_{m,n \leq k}^{m \neq n} (x_m \neq x_n) \tag{2.9}$$

and (2.5) is now of the form

$$\bigwedge_{j \in J} (\exists y) C_j^{(d)}[y, x_1, x_2, \ldots, x_k] \;\& \tag{2.10}$$
$$(\forall y)\Big(\bigwedge_{m=1}^{m=k} (y \neq x_m) \supset \bigvee_{j \in J} C^{(d)}[y, x_1, x_2, \ldots, x_k]\Big) \;\&\; C_l^{(0)}[x_1, x_2, \ldots, x_k].$$

What this means is that in the presence of identity constituents can be written precisely in the same way as in the identity-free case provided that an exclusive interpretation of quantifiers and free individual variables is adopted.

In the rest of this paper, I shall assume that identity is not present.

The concept of (quantificational) depth of $d(S)$ of a formula $S$ can be defined for arbitrary formulas as follows:

(d.i) If there are no quantifiers in $S$, $d(S) = 0$.
(d.ii) $d(S_1 \;\&\; S_2) = d(S_1 \vee S_2) = \max[d(S_1), d(S_2)]$
(d.iii) $d(\sim S) = d(S)$
(d.iv) $d((\exists x)S[x]) = d((\forall x)S[x]) = d(S[x]) + 1$

It is easily seen that this definition agrees with the way the notion of depth was used in connection with constituents.

In discussing the identity of constituents we shall consider (i) the order of disjuncts and conjuncts, (ii) the choice of bound variables, and (iii) possible repetitions of identical (*modulo* (i)–(ii)) members as inessential. If this idea is used in the numbering (indexing) constituents we can prove the following:

LEMMA 2.1: *If $i \neq j$,*

$$C_i^{(d)}[x_1, x_2, \ldots, x_k] \vdash \sim C_j^{(d)}[x_1, x_2, \ldots, x_k]. \tag{2.11}$$

This is easily proved by induction on $d$. It is also obvious on the basis of the intuitive meaning of constituent.

We can also prove

LEMMA 2.2: *If $S$ is a closed formula of $L$ of depth $d$, then for each $i$ either*

$$C_i^{(d)} \vdash S \tag{2.12}$$

*or*

$$C_i^{(d)} \vdash \sim S. \tag{2.13}$$

This, too, can be proved by induction on $d$.

The same can be proved for formulas $S[x_1, x_2, \ldots, x_k]$ and constituents $C^{(d)}[x_1, x_2, \ldots, x_k]$ having the same free variables $x_1, x_2, \ldots, x_k$. We shall call this result Lemma 2.3.

In particular,

LEMMA 2.4: *For each constituent of the form*

$$C_i^{(d+1)}[x_1, x_2, \ldots, x_k] \tag{2.14}$$

*there is precisely one constituent*

$$C_j^{(d)}[x_1, x_2, \ldots, x_k] \tag{2.15}$$

*such that (2.14) logically implies (2.15). For other values of $j$, (2.14) logically implies the negation of (2.15).*

In the former case (2.15) can be obtained from (2.14) by omitting it from all constituents of depth 1, together with connectives that thereby become idle, and all repetitions. It is obvious that the result is implied by (2.14).

What Lemma 2.4 says is in effect that you can omit the last layer of quantifiers from any constituent and obtain a shallower one which is implied by the original. In fact you can omit any one layer of quantifiers in a given constituent.

Together Lemmas 2.1–2.4 entail

LEMMA 2.5: *Each consistent formula* $S^{(d)}[x_1, x_2, \ldots, x_k]$ *of depth* $d$ *with the free variables* $x_1, x_2, \ldots, x_k$ *is logically equivalent with a disjunction of constituents of the form*

$$C_j^{(d)}[x_1, x_2, \ldots, x_k].$$

Not only can we omit layers of quantifiers from a constituent; we can likewise omit arguments from it.

LEMMA 2.6: *Given a consistent constituent*

$$C_i^{(d)}[x_1, x_2, \ldots, x_k], \tag{2.16}$$

*consider the constituent*

$$C_i^{(d)}[x_1, x_2, \ldots, x_{k-1}, \{x_k\}] \tag{2.17}$$

*obtained from (2.16) by omitting from it all atomic formulas containing* $x_k$*, all connectives which thereby become vacant, and all repetitions. Then (2.17) is the only constituent of the form*

$$C^{(d)}[x_1, x_2, \ldots, x_{k-1}] \tag{2.18}$$

*which is implied by (2.16).*

*Proof:* It is again obvious that (2.16) implies (2.17). If it implied any other constituent of form (2.18), it would be inconsistent by Lemma 2.1.

Several of the basic concepts of model theory are easily and naturally defined by reference to constituents.

A consistent sequence of constituents

$$C_{i(d)}^{(d)}[x_1, x_2, \ldots, x_k] \tag{2.19}$$

with a fixed $k$ ($k > 0$), but with an ever increasing $d = 1, 2, 3, \ldots$ defines a $k$-type. It is easily shown that this definition is equivalent with the usual one, according to which a $k$-type is the maximal consistent set of formulas with $x_1, x_2, \ldots, x_k$ as their only free variables.

When $k = 0$, we have a sequence of closed constituents

$$C_{i(d)}^{(d)} \qquad (d = 1, 2, \ldots). \tag{2.20}$$

From Lemma 2.5, it is seen that (2.20) defines a complete theory, and that each complete theory can be represented in this way.

Notice that each $k$-type (2.19) implies a unique complete theory (2.20). For each member of the sequence (2.19) implies a unique constituent without any individual constants in virtue of Lemma 2.6. Those types (2.19) which so imply (2.20) are the only ones consistent with (2.20).

The $k$-type (2.19) is compatible with the complete theory iff constituents (2.19) all occur in the successive members of (2.20).

The types compatible with (2.20) will be called the types of the complete theory (2.20). Each type satisfied in a model of (2.20) is a type of (2.20), but all the types of (2.20) need not be satisfied in a given model of (2.20). The question as to which of them are satisfied is one of the central ones in model theory. Different kinds of models are distinguished from each other by the types that are satisfied in them.

One particularly useful result concerning constituents is the following:

LEMMA 2.7: *Given a complete theory (2.20), a model $M$ of (2.20), and a constituent $C$ (or a substitution instance of a constituent containing names of members of dom($M$)), if $C$ is compatible with the set of sentences true in $M$, $C$ is satisfied in $M$. For constituents without names, it suffices to assume that they are compatible with Th($M$).*

We can here perhaps see some of the advantages of the use of constituents. All the lemmas of this section can be seen to be valid directly on the basis of the import of a constituent. (Cf. the explanation of the meaning of the tree structure of a constituent given above.) Lemma 2.7 is a case in point, though perhaps slightly less obvious at first than the earlier lemmas.

Other results can likewise be read off from the intuitive meaning of a constituent, albeit not equally directly. As an example of such a result, we can mention the following result off almost immediately from the intuitive meaning of a constituent in the following:

LEMMA 2.8: *Assume that*

$$C_i^{(d)}[a_1, a_2, \dots, a_k] \tag{2.21}$$

*is compatible with*

$$C_j^{(d+k)}. \tag{2.22}$$

*Then (2.22) implies*

$$(\exists x_1)(\exists x_2)\cdots(\exists x_k)C_i^{(d)}[x_1, x_2, \dots, x_k]. \tag{2.23}$$

*Proof* (informal): In exploring a world in which (2.22) is true, we can come upon $x_1 = a_1, x_2 = a_2, \dots, x_k = a_k$ in this order. If (2.21) is likewise true in the same world, as it can be if (2.21) and (2.22) are compatible, the rest of the world is described by $C_i^{(d)}[a_1, a_2, \dots, a_k]$.

This lemma holds by the same token if there are additional free variables or constant parameters in (2.21) and (2.22).

Many of the well-known results in model theory are proved easily and in a perspicuous way by means of constituents. As an example of the use of constituents to systematize old results and to obtain new ones, Rantala's monograph ***Aspects of definability*** (1977) can be mentioned.

More illustrations of the use of constituents are offered in the next few sections.

**§3. A wrong implementation of the extremality idea.** At this point, it might seem to be easy to implement extremality conditions on models. The natural way to interpret our extremality requirements is to say the following: A model is minimal iff as few *kinds of individuals* as possible are instantiated in it; a model is maximal iff as many *kinds of individuals* as possible are instantiated in it. Then the *prima facie* plausible idea is to take the concept of type defined above as the explication of the pre-theoretical idea of a kind of individuals (or a kind of $k$-tuples of individuals). Then the question raised at the end of the preceding chapter (Which types are satisfied in a model?) would become highly relevant to the extremality project.

What can we say by way of a response to this suggestion? In order to answer the question, we need a few further concepts. It may happen that the $k$-type

$$C^{(d)}_{i(d)}[x_1, x_2, \dots, x_k] \qquad (d = 1, 2, \dots) \tag{3.1}$$

compatible with the complete theory $T_j$

$$C^{(d)}_{j(d)} \qquad (d = 1, 2, \dots) \tag{3.2}$$

stops branching from some point on. Then there is an initial segment of (3.1) such that only one continuation of it is compatible with (3.2). Such a type is called *atomic* in $T_j$ = (3.2).

It may happen that each initial segment of each $k$-type (for each $k$) compatible with (3.2) is consistent with an atomic $k$-type. Then the entire complete theory (3.2) will be called *atomic*.

Clearly, each atomic $k$-type compatible with (3.2) must be satisfied in each model of (3.2). The interesting question is whether any other types need to be satisfied. This question turns out to be more complicated than one might first suspect. A model $M$ is called *atomic* iff each $k$-tuple of the elements of the domain dom($M$) of $M$ satisfies an atomic $k$-type. A partial answer to the question just posed is given by

LEMMA 3.1: *A complete theory has a countable atomic model iff it is atomic.*

In the other direction, there are models $M$ such that each $k$-type, for each $k$, compatible with the complete theory Th($M$) true in $M$, is satisfied. Such models are called *weakly saturated.*

A related requirement is the following: Suppose a $k$-type $t_1[x_1, x_2, \dots, x_k]$ is compatible with a $(k + 1)$-type $t_2[x_1, x_2, \dots, x_k, x_{k+1}]$ in Th($M$) and that $a_1, a_2, \dots, a_k \in$ dom($M$) satisfy $t_1[a_1, a_2, \dots, a_k]$. If there always exists $a_{k+1} \in$ dom($M$) such that $a_1, a_2, \dots, a_k, a_{k+1}$ satisfy $t_2[a_1, a_2, \dots, a_k, a_{k+1}]$, then a weakly saturated model $M$ is called *saturated.*

Saturated models are interesting "special models." It is not difficult to prove that any consistent complete theory has such a model.

Atomic models are—or seem to be—minimal models in some reasonable sense, and saturated models seem to be maximal models in an equally clear sense. The idea is this: It seems that 1-types constitute the finest partition of individuals into different "kinds" that can be affected by first-order means; and *mutatis mutandis* for $k$-types with $k > 1$. Hence it seems that the poorest models one can characterize by first-order means are the ones in which only those "kinds" (i.e., types) are exemplified which must be satisfied in any case, i.e., atomic models. Likewise, it appears that the richest models that can be dealt with on the first-order level are the ones in which all the different "kinds" (i.e., types) are instantiated, perhaps with the proviso that these types are instantiated so as to allow all possible steps from $k$-types to $(k+1)$-types. In other words, the richest models seem to be weakly saturated or perhaps saturated models.

In brief, the concepts of atomicity and saturation seem to be the natural explications of the ideas of minimality and maximality that are guiding my thinking. Yet they do not do this job well at all. Extremality requirements so interpreted do not allow us to capture the intended (standard) models in the interesting cases.

For instance, we cannot in this way capture naturally the intended "standard" model of Peano arithmetic. On the contrary, it is known (see, e.g., Chang and Keisler, Example 3.4.5) that any consistent complete extension of Peano arithmetic is an atomic theory and hence has atomic models. This holds also for complete theories not true in the intended structure of natural numbers. Hence the atomicity requirement does not do the job of capturing the structure of natural numbers.

Another example is offered by the (first-order) theory of dense linear order. It has a model which has the structure of the rationals. This model is at the same time an atomic model and a saturated one. But it is not really a minimal model in some intuitive sense, for you can omit elements from it and yet preserve its status as a model. It is not really a maximal model, either, in some striking sense, because it can be embedded in a richer one, viz. the structure of the reals, which is not isomorphic with it.

The notions of atomicity and saturation of course do not exhaust the resources of contemporary model theory. For instance, there is the notion of prime model.

DEFINITION: *A model $M_0$ of a theory $T$ is a **prime model** iff it can be elementarily embedded in every model $M$ of $T$.*

Prime models might look like plausible candidates for the role of a minimal model. However, on a closer look even the notion of prime model is not an adequate explication of the idea of a minimal model. For one thing, the way this notion is usually introduced is not useful to us as such. What we are looking for are some intuitive structural characterizations of minimality and maximality, and the notion of primeness does not give us such a characterization.

When I say this I mean the following: What made the idea of atomicity so appealing is that there is a clean syntactically definable notion of a kind of individual which enabled us to speak of a model in which a minimum of such "kinds" were instantiated. More generally, what we are looking for are characterizations of a minimal model $M$ in terms of the constituent representation of the complete theory $\mathrm{Th}(M)$ true in $M$. For the constituent representation is in some obvious sense an explicit description of the most easily understandable features of the structure of $M$. In a sense, therefore, we want to have a characterization of minimality whose applicability can so to speak seem directly from the theory $\mathrm{Th}(M)$. Now it surely cannot be seen directly from model $M$ itself or from the theory $\mathrm{Th}(M)$ whether $M$ can be embedded elementarily in certain other models.

Of course, another way of defining primeness might do the trick. But the most prominent alternative characterization of a prime model, viz. to characterize it as a countable atomic model, does not fare much better. From a theory it is very hard to see directly what the cardinality of its several models might be.

Moreover, the notion of prime model is subject to most of the same objections as were marshaled above against atomic models as implementations of minimality. For instance, even though the structure of natural numbers $N$ is the unique prime model of the Peano arithmetic, it is not the prime model of all the consistent extensions of this arithmetic, viz. of those which are not true in $N$.

The most flagrant source of dissatisfaction is the fact that a prime model of a complete theory might be elementary equivalent with a proper submodel of itself. An example is offered by the theory of dense linear order, where a prime model, for instance, the structure of the rationals, could obviously be elementarily equivalent with its proper submodel. Hence prime models are not always minimal models in any intuitive sense of the word.

§4. **Super models.** The explanation of the failure of special models to implement the extremality idea is not very hard to see. Types are not the right explication of the idea of "kinds of individuals" existing in a model $M$. A type, say a one-type, characterizes a kind of individual in so far as this individual is considered alone. In order to catch full the idea of a kind of individual, we have to consider them also in relation to the other individuals in the model.

This refined idea of a "kind of individual" can be captured by means of the following definition:

Let $M$ be a model and let $a_1, a_2, \ldots$ be a sequence of members of the domain $\mathrm{dom}(M)$ of $M$. Let

$$C^{(d)}_{i(d,k)}[x, a_1, a_2, \ldots, a_k] \qquad (d = 1, 2, \ldots, k = 0, 1, 2, \ldots) \tag{4.1}$$

be a (double) sequence of mutually consistent constituents compatible with the complete theory $\mathrm{Th}(M)$ true in $M$. Assume also that the constituents $C^{(d)}_{i(d,k)}[\{x\}, a_1, a_2, \ldots, a_k]$ are all true in $M$. Then (4.1) is said to define *supertype* in $A = \{a_1, a_2, \ldots\}$ relative to $M$.

The justification of formulating the definition of a supertype in this way is that the theory defined by (4.1) clearly does not depend on the order of the $a_1, a_2, \ldots$

The corresponding sequence with individual variables instead of constants, i.e.,

$$C^{(d)}_{i(d,k)}[x, y_1, y_2, \ldots y_k] \tag{4.2}$$

can be called the *structure of the supertype* (4.1), alias a *supertype structure.*

Many of the same things can be said *mutatis mutandis* of supertypes as can be said of types. For instance, the supertype (4.1) is said to be *atomic* iff it stops branching after a certain point. More explicitly, (4.1) is atomic iff it has a member

$$C^{(d)}_{i(d,k)}[x, a_1, a_2, \ldots, a_k] \tag{4.3}$$

such that for any $e$ and any $c_1, c_2, \ldots c_l \in \text{dom}(M)$ there is only one constituent of the form

$$C_j^{(d+e)}[x, a_1, a_2, \ldots, a_k, c_1, c_2, \ldots c_l] \tag{4.4}$$

compatible with (4.3).

One of the basic properties of supertypes is the following

LEMMA 4.1: *Given $M$ and a supertype (4.1) compatible with the complete theory Th($M$) true in $M$, each member of the sequence (4.1) (and hence each initial segment of (4.1)) is satisfied in $M$ by some individual $b$, i.e., there is $b \in$ dom($M$) such that*

$$M \vDash C_{i(d,k)}^{(d)}[b, a_1, a_2, \ldots, a_k]. \tag{4.5}$$

This follows clearly from Lemma 2.7.

Hence in a sense each initial segment of each supertype compatible with Th($M$) is satisfied in $M$. The only open question is whether the entire supertype is.

From what has been said it follows that each atomic supertype is satisfied.

Assume now that (4.1) defines an atomic supertype in $M$ and that

$$C_{i(e,l)}^{(e)}[x, a_1, a_2, \ldots, a_l] \tag{4.6}$$

is the member of (4.1) after which (4.1) no longer branches. We shall say that (4.6) *determines* the atomic supertype. Let us also assume that Th($M$), represented in the form (3.2), is the complete theory true in $M$.

One the assumptions just stated, we have

LEMMA 4.2: *Let*

$$C_j^{(e+f)}[x, a_1, a_2, \ldots, a_l, b_1, b_2, \ldots, b_m] \tag{4.7}$$

*be any constituent compatible with (4.6) and Th($M$). Then we must have $j = i(e+f, l+m)$, i.e., (4.7) must be a member of (4.1).*

*Proof:* This is what it means for (4.1) to stop branching at (4.6).

LEMMA 4.3: *On the same assumptions, each member of (4.1) later than (4.6) is equivalent with all of its successors, given Th($M$).*

*Proof:* Each member of (4.1) is implied by its successors by Lemmas 2.4 and 2.6. Hence what we have to prove is that it implies them, given Th($M$). For this purpose, it suffices to show that (4.6) is not compatible (together with Th($M$)) with any other constituent of the form

$$C_j^{(e+f)}[x, a_1, a_2, \ldots, a_l, y_1, y_2, \ldots, y_m]. \tag{4.8}$$

In order to see this, let (4.8) be compatible with Th($M$) and (4.6). Then by the same reasoning as in Lemma 2.7, there are $b_1, b_2, \ldots, b_m \in$ dom($M$) such that (4.7) occurs in some supertype (4.1) of $M$. But if so, by Lemma 4.2, $j = i(e+f, l+m)$, in other words, there is only one constituent of form (4.8) compatible with Th($M$) and (4.6).

LEMMA 4.4: *If Th(M) implies that, in a sequence like (4.1), each member is equivalent with its successors from (4.6) on, then (4.6) determines an atomic supertype.*

*Proof:* If (4.6) is equivalent with each if its successors, say (4.7) with $j < i(e+f, l+m)$, then by Lemma 2.1 it is incompatible with (4.7) with any other $j$. In other words, (4.1) can be continued from (4.6) in only one way, i.e., (4.6) determines an atomic supertype.

THEOREM 4.1: *On the same assumptions as in Lemma 4.2,*

(4.9)

$$Th(M) \vdash (4.6) \supset (\forall y_1)(\forall y_2)\cdots(\forall y_m)C^{(e+f)}_{i(e+f,l+m)}[x, a_1, a_2, \ldots, a_l, y_1, y_2, \ldots, y_m]$$

*Moreover, if (4.9) holds for all F, M, (4.6) determines an atomic supertype.*

*Proof:* From Lemma 4.3 we have

$$\text{Th}(M) \vdash ((4.6) \supset C^{(e+f)}_{(e+f,l+m)}[x, a_1, a_2, \ldots, a_l, b_1, b_2, \ldots, b_m]) \tag{4.10}$$

From this (4.9) follows by first-order logic.

Conversely, if (4.9), then (4.6) implies (given Th($M$)) all its successors. By Lemma 4.4, it determines an atomic supertype.

THEOREM 4.2: *On the same assumptions*

$$Th(M) \vdash (\forall z_1)(\forall z_2)\cdots(\forall z_l)(C^{(e)}_{i(e,l)}[x, z_1, z_2, \ldots, z_l] \supset (\forall y_1)(\forall y_2)\cdots(\forall y_m)C^{(e+f)}_{i(e+f,l+m)}[x, z_1, z_2, \ldots, z_l, y_1, y_2, \ldots, y_m]) \tag{4.11}$$

*Conversely, if (4.11) holds, (4.6) determines an atomic supertype, provided that (4.6) is true in M.*

*Proof:* In the same way as in Theorem 4.1.

THEOREM 4.3: *Assume that (4.6) determines an atomic supertype. Then the same supertype structure is determined by a constituent of the form*

$$C^{(e+l)}_j[x]. \tag{4.12}$$

*Proof:* Consider

$$C^{(e+l)}_{i(e+l,l)}[x, a_1, a_2, \ldots, a_l]. \tag{4.13}$$

By Lemma 4.3, (4.13) is equivalent with (4.6), given Th($M$). The formula

$$C^{(e+l)}_{i(e+l,l)}[x, \{a_1\}, \{a_2\}, \ldots, \{a_l\}] \tag{4.14}$$

is of the form (4.12), and in fact can serve as (4.12).

In order to show that this is what we want, it suffices in virtue of Lemma 4.4 that for each $f$ and $m$ there is a constituent of the form

$$C^{(e+l+f)}[x, y_1, y_2, \ldots, y_m] \tag{4.15}$$

implied by (4.14), given Th($M$).

Now, in virtue of Lemma 2.8, (4.14) implies

$$(\exists z_1)(\exists z_2)\cdots(\exists z_l)c^{(e)}_{i(e,l)}[x, z_1, z_2, \ldots z_l]. \tag{4.16}$$

In virtue of Lemma 4.3, there is a unique constituent of the form (4.15) implied by Th($M$) and

$$C^{(e)}_{i(e,l)}[x, z_1, z_2, \ldots, z_l]. \tag{4.17}$$

But there obviously is such a unique constituent implied by Th($M$) and (4.14).

Results like Theorem 4.3 are interesting in a wider perspective. At first sight, it might seem that the switch from special models to superspecial models destroys the strategic advantages offered by concepts like type and atomicity. The crucial thing about them is how they help us to read off (as it were) the structural (model-theoretical) properties of the models of a theory, especially a complete theory, from the syntactical structure of this theory. The way in which the atomicity of a model of a complete theory hangs together with the structure of types in the constituent representation of this theory is a typical example of this strategy.

It might seem that the way supertypes are defined deprives us of the use of this syntax-to-models strategy. For supertypes are defined by reference to some given model. Hence it seems circular to study supertypes and their interrelations for the purpose of gaining insights into the structure of models.

What the theorems just proved show is that this impression is mistaken. Even though supertypes are defined by relation to one particular model, some of their most crucial properties depend only on the structure of the complete theory Th($M$) true in $M$. In particular, what the atomic supertypes of $M$ are is in a certain sense completely determined by Th($M$). For instance, each atomic supertype is determined by a constituent of form (4.12) or $C^{(d)}_i[x]$ occurring in the constituent representation of the given theory. Here (4.12) does not depend on any particular member of dom($M$).

One application of these observations is that we can define the notion of *superatomicity* for a complete theory in analogy with the definition of atomicity of complete theories. the complete theory Th($M$) true in $M$ is superatomic iff each initial segment of each *supertype* structure compatible with Th($M$) is compatible with (the structure of) an atomic supertype.

Then we can also define a superatomic model. A *superatomic model* $M$ is one in which only atomic supertypes are satisfied, and each different supertype by precisely one individual. Then we can also prove easily the following:

THEOREM 4.4: *Each superatomic complete theory has a superatomic model.*

It is also easily seen that this superatomic model is uniquely determined (up to isomorphism).

THEOREM 4.5: *A model $M$ of a complete theory $T$ = Th($M$) is superatomic iff it is a prime model but none of its proper elementary submodels is prime.*

In order to prove Theorem 4.5, we can first prove

LEMMA 4.5: *A superatomic model $M$ of a complete theory $T$ = Th($M$) is prime.*

In order to show that $M$ is prime, we have to show that it can be elementarily embedded in an arbitrary model $M^*$ of $T$. Now each member $b_i$ of dom($M$) satisfies a supertype determined by a constituent of the form

$$C_i^{e(i)}[x]. \tag{4.18}$$

But, by Lemma 4.1 (4.18) must be satisfied by some individual, call it $b_i^*$ in $M^*$. The mapping of each $b_i$ on $b_i^*$ defines an embedding of $M$ into $M^*$. It is easily seen that this is an elementary embedding.

To return to the proof of Theorem 4.5 it is clear that if we try to map $M$ elementarily into itself, each $b_i \in \text{dom}(M)$ must be mapped on itself. For the image $b_i^*$ must satisfy the same constituent (4.18) as $b_i$, the mapping being an elementary embedding. But the only member of dom($M$) to satisfy this constituent is $b_i$ itself. This proves Theorem 4.5.

In this way we can also see that one of the main shortcomings of prime models as explications of minimality is overcome by superatomic models.

This is *prima facie* a major difference between atomic and superatomic models in that in an atomic model, every $k$-type, for $k = 1, 2, \ldots$, satisfied in it must be atomic, whereas the definition of a supertype involves directly only formulas with one free individual variable. These apparently correspond to one-types only.

To reassure the reader, we can prove

LEMMA 4.6: *A superatomic model is atomic.*

In order to show this, assume first that $M$ is a superatomic model of Th($M$). Then each element of dom($M$) satisfies an atomic supertype. If $\langle a_1, a_2, \ldots, a_k \rangle$ is a $k$-tuple of such elements, one can see by the same line of argument as was given fro Theorem 4.1 that there is a consistent structure

$$C_i^{(d)}[x_1, x_2, \ldots, x_k] \tag{4.19}$$

satisfied by $\langle a_1, a_2, \ldots, a_k \rangle$ (and hence compatible with Th($M$)) such that each formula

$$C_i^{(d)}[a_1, a_2, \ldots, a_{j-i}, x, a_{j+i}, \ldots, a_k] \tag{4.20}$$

defines a supertype. Then for each $e$ there is only one constituent of the form

$$C^{(d+e)}[x_1, x_2, \ldots, x_k] \tag{4.21}$$

compatible with (4.18) and with Th($M$). In other words, (4.18) determines an atomic type.

This proof illustrates how we can get along in our "supertheory" by means of supertypes with only one free variable, i.e., with what *prima facie* should be called one-supertypes. Just because in supertypes we heed the relation of the kinds of individuals characterized by them to other individuals in the models, we do not need $k$-supertypes with $k \geq 2$.

In order to extend one horizon to arbitrary theories instead of just complete ones, we must first extend our main concepts.

A supertype is said to be ***strongly atomic*** if it stops branching in all models compatible with one of its initial segments.

More explicitly, a constituent

$$C_i^{(e)}[x, a_1, a_2, \ldots, a_l] \tag{4.22}$$

determines a strongly atomic supertype iff it determines an atomic supertype in any model of (4.22).

A theory $T$ is *strongly superatomic* iff each initial segment of each supertype structure compatible with $T$ is compatible with (the structure of) a strongly atomic supertype.

A strongly superatomic model $M$ is one in which only strongly atomic supertypes are satisfied, and each of them by precisely one individual.

It is now easy to prove suitable extensions of our earlier results, for instance

THEOREM 4.6: *Each complete theory compatible with a strongly superatomic theory has a strongly superatomic model.*

The most natural generalization of the notion of saturation (of a model $M$) is not equally directly connected with the structure of the complete theory $\mathrm{Th}(M)$.

DEFINITION: *A model $M$ is* **absolutely supersaturated** *iff each supertype compatible with $Th(M)$ is satisfied.*

A model $M$ is supersaturated relative to a set of individuals $A = \{a_i\}(i \in I), A \subseteq \mathrm{dom}(M)$ iff each supertype (4.1) with $a_1, a_2, \cdots \in A$ compatible with $\mathrm{Th}(M)$ is satisfied in $M$.

These notions still rely fairly heavily on the particular model $M$. However, the insights so far reached enable us to define a somewhat less demanding characteristic of a model which will turn out to be most useful.

DEFINITION: *A model $M$ is* **atomically supersaturated** *iff*

*(i) Each atomic supertype structure is satisfied in $M$ by precisely one individual $a_1$; and*

*(ii) $M$ is supersaturated with respect to the set $A = \{a_i\}$ of the individuals satisfying the different atomic supertype structures.*

It is easily seen that the following theorem holds:

THEOREM 4.7: *Each complete theory has an atomically saturated model.*

Many results familiar from the traditional model theory have related results that can be proved for supertypes. Here I shall mention only one as an example.

THEOREM 4.8: *From a model $M_1$ of a complete theory $T$ one can omit a countable number of individuals, each satisfying only nonatomic supertypes, and obtain a model $M_2$ of $T$ which is an elementary submodel of $M_1$.*

**§5. Super special models as implementations of extremality requirements.** The concepts defined in the preceding section serve as excellent explications of the notions of minimality and maximality which are the focal ideas of this study. The minimality requirement (principle of paucity) is naturally captured by the idea of strong superatomicity, which literally amounts to imposing on models the least possible qualitative variety in so far as the relevant qualitative differences are understood by references to supertypes. This procedure is vindicated by the fact that superatomic models of complete theories turn out to be the minimal prime models, i.e., models elementarily embeddable into any model of the given complete theory, but not into any of their own proper submodels.

Strongly superatomic models turn out to be capable of doing the kind of job they were cast to do. For one thing, a Peano-type axiomatization of elementary arithmetic turns out to be categorical and have the structure $N$ of natural numbers as its sole model, if the space of models is restricted to strongly superatomic ones. The only strongly superatomic model of the Peano axiomatization of elementary number theory can be shown to be the structure $N$ of natural numbers, if the space of models is restricted to strongly superatomic models. This does not conflict with Gödel's incompleteness result, because the new "paucity logic" is not axiomatizable. Hence the new perspective on elementary arithmetic does not automatically create new avenues of actually establishing new number-theoretical results. What it nevertheless can in principle do is to facilitate the discovery of stronger and stronger proof methods. For the search for such methods can now be guided by clear-cut semantical considerations.

But how do I know that the structure of natural numbers $N$ is the only superatomic model of the Peano axioms? These axioms are compatible with a number of different complete theories, only one of which is true in $N$. We can call it Th($N$). It is easily seen to be strongly superatomic. It has different non-isomorphic models, of which $N$ is one. It is easily seen that $N$ is in fact the only superatomic model of Th($N$).

But what about the other complete theories compatible with Peano arithmetic? How do we know that they do not have strongly superatomic models, too?

Perhaps the quickest way of seeing that they do not is to note a trivial-looking property of strongly superatomic models. Let each member of such a model $M$, say $b$, be correlated one-to-one with one of the constituents $C_1^{(d)}[x]$ which determines the strongly atomic supertype that $b_1$ satisfies. Given two such individuals $b_1$ and $b_2$, the second one being correlated with $C_2^{(e)}[x]$, one can construct effectively the formula that determines the strongly atomic supertypes satisfied by their sum, likewise for their product. This means that sum and product are recursive relations in a strongly superatomic model.

The details of this argument are given in an appendix below.

Now Tennenbaum has shown (see Tennenbaum 1959; Feferman 1958; Scott 1959; Kaye 1991, p. 153) that sum and product are not recursive in any non-standard model of Peano arithmetic (in the sense of relative recursivity just explained). From this it follows, together with the observations just made, that non-standard models of Peano arithmetic cannot be superatomic, just as was claimed.

Things are somewhat more complicated with respect to the notion of maximality (principle of plenitude). Hilbert's completeness axiom amounted to requiring that any attempted adjunction of a new individual to the intended model must lead to a violation of the other axioms. But such requirements cannot be satisfied in first-order theories in view of the upwards Skolem–Löwenheim theorem. In Hilbert's axiomatization of geometry, his completeness axiom has the intended effect only because he had also assumed the Archimedean axiom and also tacitly interpreted the notion of natural number involved in the Archimedean axiom in the standard sense. Fortunately, the intended maximality conditions can typically be interpreted so as to require only maximal qualitative richness, not necessarily the presence of the maximal selection of individuals in the intended model or models. Hence the natural course for us here is, if we want to keep our conceptualizations generally applicable, to require only maximal qualitative richness but not completeness in Hilbert's strong sense. But this does not really mean giving up Hilbert's original ideas. For even geometry, the function of the completeness axiom is to enforce continuity, not to restrict the "size" of the universe of discourse. Indeed, Hilbert's completeness axiom can be replaced by a pair of assumptions that can be roughly expressed as follows:

(H.1) If two points have the same relations to all other points and lines, they are identical.

(H.2) If $M$ is a model of the other axioms and if there is a set of relations between an unspecified individual $x$ and the members of $M$ which is compatible with the other axioms and with the diagram of $M$, there exists in $M$ an individual with these relations.

As you can easily see, (H.1) follows from other axioms. (Axioms of incidence and order suffice for the purpose.) Hence the import of the completeness axiom is essentially (H.2), which is an assumption of maximal qualitative richness rather than of maximal size as far as individuals are concerned. In fact, the force (H.2) is easily seen to amount to requiring that maximal number of supertypes be instantiated, compatible with the other assumptions.

Hence we can safely think of the maximality idea as being captured by a requirement of maximal qualitative richness. But in the preceding section we found that we have a genuine choice here. We can require either absolute supersaturation or atomic supersaturation of our models. The difference between the two appears nevertheless to be relatively unimportant. For one thing, the former implies the latter. Furthermore it will turn out that even atomic supersaturation is quite a strong assumption.

More has to be done here, however, than to explicate the twin notions of minimality and maximality. In the most interesting mathematical theories beyond elementary number theory, such as the theory of reals, axiomatic geometry, and set theory, the crucial thing turns out to be neither minimality assumptions nor maximality assumptions, but their interaction. Typically, we can assume that we are dealing with a theory which contains a one-place predicate, say $N(x)$, for natural numbers, and suitable axioms for natural numbers. (That is, when the axioms of the theory are relativized to $N(x)$, they must yield as consequences a reasonable axiomatization of natural numbers.) Notice that $N(x)$ does not necessarily have to be a primitive predicate. Then, we obviously have to assume that part of a model of the theory which corresponds to natural numbers is superatomic but that the rest of the model is maximal. But maximal in what sense? The crucial fact here is that we cannot simply assume that the individuals satisfying $\sim N(x)$ form a atomically supersaturated model, for that may be incompatible with the requirement that $\{x : M \models N(x)\}$ is superatomic. We can only require that the model realizes a maximal number of supertypes (either absolutely or reelative to the set of individuals satisfying superatomic types) compatible with the requirement that $\{x : M \models N(x)\}$ be superatomic.

We have thus motivated the following definitions:

Let us assume that we are given a model $M$ of a complete theory $\mathrm{Th}(M)$ which contains a one-place predicate $N(x)$ for natural numbers. Let us assume further that the theory $\mathrm{Th}(M)$ as restricted to $\{x : M \models N(x)\}$ is superatomic. Then the model $M$ is *absolutely Hilbertian* iff the following requirements are also satisfied:

(i) $M$ restricted to $\{x : M \models N(x)\}$ is superatomic.

(ii) A maximal subset of supertypes compatible with (i) are instantiated in $M$.

$M$ is an *atomically Hilbertian* model iff the following conditions are satisfied:

(i) As before.

(ii)* A maximal number of supertypes relative to the set of individuals satisfying a superatomic type are instantiated in $M$.

For instance, consider a set of axioms for real numbers which includes a predicate $N(x)$ for natural numbers. Then (i) becomes essentially the Archimedean axiom. By the usual Dedekind-type line of thought, one can then show that the structure of the (actual) reals is the only one which is also atomically Hilbertian.

Essentially, the same also happens in Hilbert's axiomatization of geometry. Hilbert needs the Archimedean axiom (utilizing the standard concept of natural number) to force as it were the multiples of the unit line to match the structure of natural numbers, and the axiom of completeness to ensure continuity. The latter point is especially clear in Hilbert (1900), where the axiom of completeness first made its appearance (as an axiom for the theory of reals rather than as an axiom of geometry). Hilbert's formulation of his axiom there also makes it clear that he thought of it as a maximality assumption.

If we give up requirement (i), we can obtain sundry non-standard models of reals. If we give up (ii), we need not any longer have all "real reals" in our model. Depending on the axiomatization, it may be sufficient, e.g., for the model to contain only all algebraic numbers.

Thus, we can again reach one of our main objectives. If we restrict the models of (the first-order language of) a theory of reals to atomically Hilbertian ones, then any reasonable theory of reals is categorical and yields the intended structure of reals as its only model (up to isomorphism). This descriptive completeness is not due to the requirement of minimality (superatomicity) alone, nor to the requirement of maximality (atomic supersaturation), but to the combination of the two in the requirement of the (atomically) Hilbertian character of the models.

An interesting pitfall here is that a complete theory need not have a unique richest model of a given cardinality, either in the sense of being absolutely Hilbertian or atomically Hilbertian. If we think of the supertypes compatible with the given theory satisfied one by one, then the ultimate outcome can so to speak depend on the order in which they are satisfied.

An interesting situation arises when the same ideas are extended to a suitable axiomatization of set theory. In order to apply the ideas sketched here, we have to assume that a predicate $N(x)$ for natural numbers is included in the language of the set theory or can be defined as the basis of the axiomatization. Then we can again stipulate that the models be restricted to atomically or absolutely Hilbertian ones, and see what happens.

I cannot here try to answer this question in general. Certain things are nevertheless relatively easy to see. Perhaps the most interesting perspective offered by our observation, is that in set theory, too, the greatest subtlety is due to the interplay of minimality and maximality requirements. On the one hand, one can construct poor (small) models of, say, ZF set theory which have a clear-cut structure but which clearly are not what is intended. On the other hand, attempts to enlarge the universe of set theory have not yielded any ultimate clarity either. It seems to me that the real source of difficulties in set theory is that the requirements of poverty and plenitude have to be balanced against each other. For another example, we can construct a theory of finite types as a many-sorted first-order theory. We might, e.g., assume that there is a primitive predicate $N(x)$ in the language for natural numbers which are among the individuals. If we then require that the models of a suitable axiomatization of such a type theory are automatically Hilbertian, the resulting theory has all sorts of nice features. For instance, the Denumerable Axiom of Choice is valid and so is the Principle of Dependent Choices for subsets of natural numbers. Furthermore, it will be easy to give a descriptively complete and indeed categorical axiomatization for a theory of the second number class (countable ordinals).

In this kind of many-sorted first-order reconstruction of type theory we can even start from an axiomatization of a discrete linear order (with an initial element) for natural numbers. Its only superatomic model is clearly $\{0, 1, 2, \dots\}$ with successor as the only relation. It is easily seen, however, that functions for addition, multiplication, etc., all necessarily exist in all the models of the full axiomatization, as indeed do all recursive functions. Thus the existence of the usual arithmetical functions does not even have to be assumed; it follows logically from the axioms. This would speak for a partial reducibility of mathematics to logic, if it were not for the fact that certain mathematical assumptions were built right into our concept of model and hence into our concept of logic.

A more sweeping philosophical perspective which opens here takes up an issue which was mooted by Hilbert and Kronecker. For Kronecker, natural numbers were the be-all if not the end-all of mathematics. In contrast, one of Hilbert's acknowledged aims in his ***Grundlagen der Geometrie*** was to show that there can be important mathematical theories which do not involve the concept of natural numbers at all. (See here Blumenthal 1922, p. 68.)

If the approach advocated here is right, there is more to be said for Kronecker and less for Hilbert than has been generally acknowledged. If the subtlety of advanced mathematical theories lies in the interplay of superatomicity requirement for the natural numbers with suitable maximality assumptions, then the concept of natural numbers is after all essentially involved in these mathematical theories, via the requirement of superatomicity. Mathematics looks more like a science of (natural) numbers than it has in a long time.

On a more technical level, there does not seem to be any obstacles in principle to use the time-honored strategy of using set theory to speak of its own semantics. In this way, e.g., Gödel captured his own metalogical construction of a constructible model in an explicit axiom. If this strategy works here, the requirement that only atomically Hilbertian models are considered would be expressible by an explicit set-theoretical axiom of the old style (without restrictions on the usual set of models). This axiom would be eminently acceptable, for (i) merely spells out the nature of natural numbers while (ii) follows from the idea that set theory is the theory of *all* sets, that in the world of set theory what *can* exist *does* exist.

Whether or not we can along these lines solve the outstanding problems of set theory remains to be seen. I do not seem to be the only one who thinks that they can be so solved. Gödel once wrote to Ulam, *apropos* John von Neumann's axiomatization of set theory:

> The great interest which this axiom [in von Neumann's axiomatization of set theory] has lies in the fact that it is a maximum principle somewhat similar to Hilbert's axiom of completeness in geometry. For, roughly speaking, it says that any set which does not, in a certain defined way, imply an inconsistency exists. Its being a maximum principle also explains the fact that this axiom implies the axiom of choice. I believe that the basic problems of abstract set theory, such as Cantor's continuum problem, will be solved satisfactorily only with the help of stronger axioms of *this* kind, which in a sense are opposite or complementary to the constructivistic interpretation of mathematics. (See Ulam 1958.)

What Gödel misses here is the crucial interplay between maximality *and* minimality assumptions, though his remark on von Neumann type axioms being complementary to constructivistic ideas perhaps suggests some degree of awareness of this fact.

**Appendix.** Let us assume that a strongly superatomic model $M$ of Peano arithmetic has been given and that a one-to-one correlation has been established between all natural numbers $n$ and all the strongly atomic supertypes of $M$. More explicitly, let the correlate $\varphi(n)$ of each $n$ be one of the constituents with one free individual variable that determine a strongly atomic supertype in $M$, different supertypes for different values of $n$. (Equivalently, the correlate of $n$ could be the Gödel number of this constituent.) This correlation establishes an isomorphism between $M$ and a certain relational structure. (Cf. here Kaye 1991, especially sec. 11.3.) In this isomorphism, a certain numerical relation will correspond to the relation of being the sum of in $M$, i.e., the relation which holds between three individuals in dom($M$) say $a, b, c$, when $S(a, b, c)$ is true in $M$, where $S$ is the expression of the sum of in Peano arithmetic. The question is whether this relation is recursive.

In order to show that it is, let us suppose that we are given three constituents each of which is correlated with some natural number by $\varphi$ and each of which therefore determines a strongly atomic supertype. Let these constituents be

(1) $$C_1^{(d)}[z]$$

(2) $$C_2^{(e)}[y]$$

(3) $$C_3^{(f)}[x].$$

From (1) and (2) we can form the formulas:

(4) $$C_1^{(d)}[z] \ \& \ C_2^{(e)}[y] \ \& \ S(z, y, x)$$

(5) $$(\exists z)(\exists y)(C_1^{(d)}[z] \ \& \ C_2^{(e)}[y] \ \& \ S(z, y, x))$$

By assumption, (1) and (2) each determines a strongly atomic supertype. Clearly (5) is satisfied by the sum of the two individuals in dom($M$) which satisfy (1) and (2). What we are interested in here is whether it is possible to determine recursively (effectively) whether (3) is also satisfied by this sum.

For the purpose, we shall first show that (5) determines a strongly atomic supertype. In order to prove this, assume that $M^*$ is a model in which (5) is satisfied. Hence (4) is also satisfied in $M^*$. What has to be shown is that, given $a_1, a_2, \ldots, a_k \in$ dom($M^*$) and $g \geq \max(d, e) + 2$, there is only one constituent of the form

(6) $$C_i^{(g)}[x, a_1, a_2, \ldots, a_k]$$

compatible with (5) and Th($M^*$).

In order for (6) to be compatible with (5), it must be compatible with (1) and (2). Since (1) and (2) both determine strongly atomic supertypes, there is a unique constituent of the form

(7) $$C^{(g+1)}[y, z, a_1, a_2, \ldots, a_k]$$

compatible with (5) and Th($M^*$). But since it follows from the axioms of Peano arithmetic that the sum of two numbers is uniquely determined, there is in (7) one and only one constituent of the form

(8) $$C_j^{(g)}[x, y, z, a_1, a_2, \ldots, a_k]$$

compatible with (4). Hence the constituent

(9) $$C_j^{(g)}[x, \{y\}, \{z\}, a_1, a_2, \ldots, a_k]$$

is the only constituent with the parameters $a_1, a_2, \ldots, a_k \in \mathrm{dom}(M^*)$ and $g$ compatible with (5) and $\mathrm{Th}(M^*)$. In other words, it is the only constituent which can serve as (6), which is therefore uniquely determined. This is just what was to be proved.

It is important to realize that this part of the overall proof is not supposed to be effective.

Consider now the unique constituent

$$C_m^{(\max(d,e))} \tag{10}$$

compatible with (4). It is true in $M$, and it can be obtained effectively from (4).

Assume first that $f < \max(d,e)$. Because (3) determines a strongly atomic supertype in $M$, there is in (9) a unique constituent of the form

$$C_n^{(\max(d,e)-1)}[x] \tag{11}$$

compatible with (3). It can be found effectively, given (3) and (14), and it clearly determines a strongly atomic supertype.

Assume then that $f \geq \max(d,e)$. Then there is a unique constituent of the form (11) in (10) compatible with (3). In this case it can be found simply by omitting layers of quantifiers from (3), hence effectively.

In either case, since (1) and (2) both determine a strongly atomic supertype in $M$, there is a unique constituent of the form

$$C_0^{(\max(d,e))}[z,y] \tag{12}$$

compatible with (4). It can be found effectively as follows: First we convert (4) into its distributive normal form

$$\bigvee_i C_i^{(\max(d,e))}[z,y] \tag{13}$$

This can be done effectively. However, we cannot in general know which disjuncts in (13) are consistent and which ones are not. This uncertainty can be eliminated simply by grinding out the logical consequences of (13) jointly with (4) one after the other until only one survives undisproved. But it follows from the axioms of Peano arithmetic that the sum of two individuals is uniquely determined. Hence there is in (12) a unique constituent of the form

$$C_p^{(\max(d,e)-1)}[x,z,y] \tag{14}$$

But since (14) is uniquely determined, then so is

$$C_p^{(\max(d,e)-1)}[x,\{y\},\{z\}] \tag{15}$$

Now this constituent can be compared with (11) effectively for identity. If the two are identical, (3) and (5) determine the same strongly atomic supertype, if not, they do not.

By reviewing the argument, it is easily seen that this determination can be made effectively. By Church's thesis, sum will therefore be a recursive relation in $M$, which was to be proved.

**Acknowledgments.** In working on this paper, I have profited from the comments, suggestions and criticisms by Professor Jouko Väänänen, by several members of his research group in Helsinki, and by Professors David McCarty and Philip Ehrlich. They are not responsible for any errors, however. My participation in the ASL European Summer Conference in 1990 was facilitated by a travel grant from the Academy of Finland, and my work by research support from Boston University.

## REFERENCES

JOHN D. BALDWIN, ***Fundamentals of Stability Theory***, Springer-Verlag, Berlin, 1988.

OTTO BLUMENTHAL, *David Hilbert*, ***Die Naturwissenschaften***, vol. 10 (1922), pp. 67–72.

C. C. CHANG and H. J. KEISLER, ***Model Theory***, North-Holland, Amsterdam, 1973.

SOLOMON FEFERMAN, *Arithmetically Definable Models of Formalized Arithmetic*, ***Notices of the American Mathematical Society***, vol. 5 (1958), pp. 679–680.

DAVID HILBERT, ***Foundations of Geometry***, tr. by Leo Unger, tenth ed., Open Court, La Salle, 1971. German original ***Grundlagen der Geometrie***, 1899.

DAVID HILBERT, *Über den Zahlbegriff*, ***Jahresberichte der Deutschen Mathematiker-Vereinigung***, vol. 8 (1900), pp. 180–184.

JAAKKO HINTIKKA, *Is there Completeness in Mathematics after Gödel?* ***Philosophical Topics***, vol. 17, no. 2 (1989) pp. 69–90.

RICHARD KAYE, ***Models of Peano Arithmetic***, Clarendon Press, Oxford, 1991.

VEIKKO RANTALA, ***Aspects of Definability*** (Acta Philosophica Fennica, vol. 29, nos. 2–3), North-Holland, Amsterdam, 1977.

HARTLEY ROGERS, Jr., ***Theory of Recursive Functions and Effective Computability***, McGraw-Hill, New York, 1967.

DANA SCOTT, *On Constructing Models for Arithmetic*, in ***Infinitistic Methods***, Pergamon Press, Oxford, 1959, pp. 235–255.

S. TENNENBAUM, *Non-archimedean Models for Arithmetic*, ***Notices of the American Mathematical Society***, vol. 6 (1959), p. 270.

STANISLAW ULAM, *John von Neumann, 1903–1957*, ***Bulletin of the American Mathematical Society***, vol. 64 (1958, May Supplement), pp. 1–49.

Department of Philosophy
Boston University
Boston, MA 02215, USA