A nominalistic account of mathematical truth.

A NOMINALISTIC ACCOUNT OF MATHEMATICAL TRUTH

A Dissertation Presented

by

EDWARD ARTHUR OLDFIELD

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 1977

Department of Philosophy

A NOMINALISTIC ACCOUNT OF MATHEMATICAL TRUTH

A Dissertation

By

EDWARD ARTHUR OLDFIELD

Approved as to style and content by:

Michael Jubien, Chairman of Committee

Edmund L. Gettier III, Member

Terence Parsons, Member

Frank Wattenberg, Member

Robert C. Sleigh, Department Head

Philosophy Department

September 1977

ACKNOWLEDGEMENTS

My greatest debt for help in writing this dissertation is owed to Michael Jubien. His advice and criticism taught me a great deal and led to substantial improvements in this work. In addition, his writings have been a focal point of my thought on the topics considered here. I would also like to thank Terence Parsons for his helpful criticism of this dissertation.

Portions of drafts of this dissertation were delivered at Syracuse University, the University of Illinois-Urbana Champaign and the University of Wisconsin-Milwaukee. I would like to thank the participants in those discussions for their comments. In particular, I would like to thank Peter vanInwagen and Jose Benardete.

Finally, I would like to thank my teachers, especially Jose Benardete, Fred Feldman, Edmund Gettier, Kenneth Haas, Michael Jubien, Thomas McKay and Terence Parsons for giving me a conception of what philosophy is and how it ought to be done.

TABLE OF CONTENTS

CHAPTER I

This dissertation attempts to make a contribution to the project of giving an adequate nominalistic account of mathematical truth. What should an account of mathematical truth provide? This is difficult to say. Roughly speaking, the idea would be to specify what mathematical sentences "say". This, of course, is vague. At the minimum, however, an account of mathematical truth should give a theory of truth for mathematical sentences.

As here conceived, the project of giving an adequate nominalistic account of mathematical truth would proceed as follows:

1) An informal mathematical language[1] a sentence belongs to is specified.

2) A formal language is described.

3) A translation procedure is described mapping sentences of the informal language into sentences of the formal language.

---

[1]Throughout this dissertation 'theory' and 'language' will be used to mean 'interpreted theory' and 'interpreted language', respectively. In speaking of uninterpreted theories or languages, I will make this explicit, often using the modifier 'uninterpreted'. There will also be occasion to speak of formal and informal theories and languages. A formal theory (language) will always be a theory (language) based on first order syntax for which an interpretation is explicitly provided. An informal theory (language) will be any other theory (language). Thus we will speak of the informal theory 'there are numbers', the interpretation of this theory being simply its ordinary English meaning, i.e. its meaning in the informal language English.

4) It is argued that this translation procedure gives adequate representations of the sentences of the informal language. (Of course, it will be difficult to say what an adequate representation is.)

5) 'Nominalism' is defined.

6) It is argued that the formal theories which are translations of informal theories are nominalistic.

The definition of truth would be: a sentence is true if its translation is true in some formal language which adequately represents some language to which that sentence belongs. Similarly, a sentence is false (neither true nor false) if its translation is false (neither true nor false) in some formal language which adequately represents some language to which that sentence belongs.

In (1) 'sentence' means 'interpreted sentence'. Only interpreted sentences "say" anything. Thus it is only interpreted sentences for which an account can be given. Note also in (1) we speak of finding a language, rather than the language, to which a sentence belongs. Clearly, any mathematical sentence belongs to English. Such sentences are also sentences of certain sub-languages of English, e.g. arithmetical language, so that there is no such thing as the language to which a given sentence belongs.

Throughout this dissertation (1) will present no problem. We will always be speaking of interpreting languages (not just sentences) so that whenever an individual sentence is considered, the language to which it belongs will already have been specified.

The contribution that this dissertation attempts to make to this project of giving an adequate nominalistic account of mathematical truth is as follows. Under (5), a theory is nominalistic if it is not ontologically committed to any particular abstract entities and it is not ontologically committed to the kind abstract entity. The notion of nominalism, as here understood, thus depends on the notion of ontological commitment. This latter is a notion which is discussed in Chapter II.

In that chpater criteria of ontological commitment are advanced which apply to formal theories of the kind described in later chapters. The work here is partly supplemental to and partly critical of Michael Jubien's work ([12], [13]) on this topic. Jubien proposed criteria which employ the notion of a quantificational model structure $\langle G,K,R \rangle,\Psi$ in the sense of Kripke [15]. Jubien suggests that these criteria are properly applied only in conjunction with certain model structures. He makes only fleeting comments, however, concerning what is an appropriate model structure for applying these criteria. Chapter II attempts to say what is an appropriate model structure for applying these criteria. This involves specifying what sort of "worlds" are to go into such a model structure. It is also argued that if a model structure of the sort defined in this chapter is used in applying criteria of ontological commitment, then some very simple criteria considered and rejected by Jubien are adequate. The notion of commitment used then in later chapters implies that a theory is committed to a particular if that particular exists in every world at which that theory is true; a theory is committed to a kind k if there are k's at every world at which that theory is true.

Chapters III and IV take up (2), (3) and (6) above. In Chapter III a general technique for interpreting first order languages is considered. Roughly speaking, the idea is that instead of assigning (at a world) a single range for the quantifiers and single denotations for the constants and predicates, the quantifiers, constants and predicates are given multiple interpretations. This technique will be described a bit more fully, in an introductory fashion, later in this chapter.

Employing this general technique then particular formal languages are described, procedures are given for translating sentences of some informal mathematical languages into these formal languages and it is then argued that the formal theories which represent pure mathematical theories such as Peano arithmetic and Zermelo-Fraenkel set theory are nominalistic.

Chapter IV introduces another general technique for interpreting first order languages. Again, this technique will be introduced in somewhat more detail later in this chapter, but very roughly the idea is as follows. As above the quantifiers, constants and predicates are given multiple interpretations. A sentence is true if <u>necessarily</u> it is true in every interpretation having certain properties.

As in Chapter II, particular formal languages interpreted using this general technique are described; the method for translating from some informal mathematical languages into these languages is given and finally it is argued that the formal theories which are translations of informal mathematical theories are true in the empty world (the world at which nothing exists). (That there is such a world, as the notion is used here, is a consequence of the discussion of Chapter II.) This, of course, implies that these formal theories are nominalistic.

It is to be emphasized that what are of most interest in these chapters are the general techniques introduced, not the particular languages described. First, no attempt is made to deal with all mathematical languages. Only certain examples are chosen. Also this emphasis on the general techniques becomes obvious in the discussion of set theory where in Chapter III, two importantly different interpretations are given. It is less obvious in other places, but it should be kept in mind. For example, an interpretation of a combined number theory-set theory is given which allows '$0=\phi$'

to be neither true nor false. An interpretation of another language allows (the translation of) 'all apples are sets' to be neither true nor false. Such results are not built into these techniques, however, for different results can be obtained for these sentences using these techniques. The flexibility of these techniques should be emphasized. Finally, throughout both of these chapters the techniques of interpretation are also applied to languages which mix both mathematical and non-mathematical vocabulary.

In Chapters V and VI, (4) is considered. These chapters attempt to indicate the inadequacies of a platonistic interpretation of mathematical languages. In both of these chapters, arguments are considered which may be construed as attempting to show that informal arithmetic, for example, is not committed to the kind <u>number</u> nor to any particular. If these arguments were sound, this would suggest (though, of course, not entail) that informal mathematical language is most adequately represented by a language interpreted using one of the techniques presented in the preceding chapters. The argument discussed in Chapter V is that of Paul Benacerraf [2]. In Chapter VI an argument of Jubien's is considered [14]. In both cases it is contended that these arguments are inconslusive. These arguments do, however, raise questions which are sufficient to motivate interest in the work of Chapters III and IV.

The treatment of (4) in this dissertation is thus quite incomplete. First, there is no conclusive argument that platonism is false or that if platonism is false, then some interpretation employing one of the techniques presented here is correct. Second, as emphasized above, even if it is assumed that some interpretation employing one of the techniques presented here is correct, little is said to justify the claim that one particular such interpretation is correct. For example, some finitists and constructivists have held that many of the sentences mathematicians generally hold to be true are not true. Such positions are ignored (but not ruled out) here. They would have to be considered in a complete account of mathematical truth.

The general position to be considered here holds, among other things, that although 'there are numbers' is true, this informal theory is not committed to the kind _number_ or to the kind _abstract entity_. One natural response to this is to ask for an account of this theory which does not imply that this theory has such commitments. It might well be expected that this positive project cannot be accomplished. That is, it might be suspected that no reasonable account of 'there are numbers' can be given which does not imply that this theory is ontologically committed to the kind _number_.

Chapters III and IV attempt to show that this suspicion is incorrect. That is, this theory can be given an account (which has some plausibility) according to which this theory does not have such a commitment. Platonism is seen not to be the only plausible basis for giving an account of mathematical truth. Another objection to this general position is as follows.

This objection begins by noting that a common, even paradigmatic, way that a theory can have an ontological commitment to a kind k or an individual a is simply to assert 'there are k's' or 'a exists'. From this it is concluded that 'there are numbers' and 'one exists' must have an ontological commitment to the kind number and the particular one, respectively. Thus the general position considered in this dissertation is untenable.

I agree that 'there are k's' and 'a exists' commonly do have ontological commitments. In fact, it is by using such sentences that the notion of ontological commitment is generally explained: theory T is ontologically commited to a just in case T says that a exists.

This indicates that a position which held that no theory has ontological commitments is in a serious difficulty: how is 'exists' being used if it is claimed that no theory says that anything exists? Let us grant then that at least some interpreted sentences of the form 'a exists' and 'there are k's' must have ontological commitments. Still this leaves open the possibility that this is not true in all cases. In particular, the above considerations do not show that it is untenable to hold that in mathematics sentences which seem to have ontological commitments do not have these commitments. It is just such a position which is developed in Chapters III and IV.

The technique for interpreting first order theories which will be investigated in Chapters III and IV has been more or less anticipated by Hartry Field [7], Michael Jubien ([12], [14]), Paul Benacerraf [2], Hilary Putnam [22] and (on Field's interpretation) Quine [26]. We will build to the presentation of this technique by discussing Field's and Jubien's approaches.

Hartry Field introduced the notion of partial signification (or partial reference) as a way of dealing with the question of indeterminacy which has been raised by Quine. To take the well-known case, Field suggests that if we were faced with a situation in which we could translate 'gavagai' either as 'rabbit' or 'undetached rabbit part' we should say that 'gavagai' partially signifies the set of rabbits and that it partially signifies the set of undetached rabbit parts.

Now this particular application of the concept of partial signification is of no special interest here, but the concept itself is. The idea behind it is that even if a word has not been suitably "hooked up" with reality (however, that may work) so that it has a single referent, still it may be that there is a class of things all of which are partial referents to that word. Reference then is a special case of partial reference: it is that case in which a word partially refers to only one thing. In general, however, the notion of partial reference has application in situations where, roughly, a word is used in such a way that there is more than one thing which has equal right to be called the referent of that word. A very simple example of this notion might be that if one person wrote the Iliad and another wrote the Odyssey it might be reasonable to say the 'Homer' partially referred to both of these authors.

A fuller understanding of this notion is, however, only to be gained by seeing how it is used to give an interpretation of a first order language. Field uses this notion to define the conditions under which a structure for a first order language L partially accords with the semantics for L. The idea is that L is a (formalized version of) natural language and that the structure assigns referents to names and extensions to predicates which those names and predicates partially signify.[2] More precisely:

---

[2]The "semantics for L" then should be understood as being the semantical properties of the language which L is intended to mormalize. These semantical properties would include what the partial referents of the terms of that language are.

(1) A structure[3] *M* partially accords with the semantics of *L* if and only if each term of *L* partially denotes or partially signifies the entity which *M* assigns to it. (211)

Characterizing truth for a language interpreted in this way is then easy.

(2) A sentence of *L* is true if and only if it is true-relative-to-*M* for every structure *M* that partially accords with the semantics of *L*. (212)

Also we may add

(3) A sentence of *L* is false if and only if it is false-relative-to-*M* for every structure *M* that partially accords with the semantics for *L*.

A simple example will illustrate the technique here. Consider the language *L* which has one name '*h*' and one one-place predicate '*F*' whose intended interpretation is discourse concerning the fatness of Homer. Assume as above that '*h*' partially denotes two different people *a* and *b*. Assume also that '*F*' partially denotes only one set: the set of fat people. There will then be two interestingly different structures for *L* which partially accord with the semantics for *L*. One such structure assigns *a* to '*h*', while the other assigns *b* to '*h*'. Now the truth of the sentence '*Fh*' will depend on the fatness of *a* and *b*. If both of them are fat '*Fh*' is true; if neither of them are fat '*Fh*' is false; while otherwise '*Fh*' is neither true nor false. Of course,

---

[3]Field and I both use 'structure' in the sense of Shoenfield [31].

a priori each of these cases is possible. This brings out an important point: structures which partially accord with the semantics for L need not be elementarily equivalent: they need not make the same sentences true.

Unfortunately (1) is not in general adequate, as Field points out ([7]: 214). The simplest sort of example of this is the following. Suppose we say that 'gavagai' partially signifies both the set of rabbits and the set of rabbit parts. Suppose also that another word 'gavago' also partially signifies both of these sets, but that the native equivalent of 'All gavagai are gavago' is false. (Thus, it might be said that 'gavagai'-'gavago' are paired much as 'rabbit'-'rabbit part' are paired in English.) According to (1), however, a structure which assigned the set of rabbits to both 'gavagai' and 'gavago' could partially accord with the semantics of the native language. In such a structure the equivalent of 'All gavagai are gavago' would be true and so by (3) this sentence would not be false.

Field says that the problem here is that 'gavagai' and 'gavago' are "correlatively indeterminate". Both partially signify the same sets, but roughly speaking, whenever 'gavagai' signifies one set, 'gavago' signifies another. To deal with this problem, Field introduces the concept of one term $\underline{t}_1$ being the basis of another term $\underline{t}_2$ ($\underline{t}_1=\underline{b}(\underline{t}_2)$). If $\underline{t}_1$ is the basis of $\underline{t}_2$, then the signification of $\underline{t}_2$ is a function of the signification of t_1. A term is independent if it does not have a basis. (1) is then revised as follows.

(1*) A structure M partially accords with the semantics of L if and only if (M(t) is the signification of t in M)

(a) each independent term of L partially denotes or partially signifies M(t);

(b) each dependent term t of L denotes or signifies M(t) relative to the correlation of M(b(t)) with b(t).

This presents us with two problems. First, what exactly is it for one term to be the basis of another? Second, what does (b) of (1*) assert? The second problem can be dealt with fairly easily. What Field has in mind here is that M assigns a dependent term t the correct value relative to what M assigns the basis of t. This can be made precise as follows. We introduce the notion of a function f for a language. f is a function of three arguments whose first and second domains both consist of all the terms and predicates of L and whose third domain and whose range consist of all sets. When given a term t of L, its basis and a set a (the signification of the basis of t) as arguments, f gives as value a set. This latter set would be the signification of t relative to the assignment of a to its basis. We may interpret (b) of (1*) then as saying that

(b') the function f for L is such that M(t)= f(t, b(t), M(b(t))).

This is, of course, very schematic. Field tells us nothing about which of the functions f are functions for a language, except that they are supposed to operate in such a way that the difficulty we have considered concerning (1) can be avoided; nor, to return to the first problem, does he say what it is to be the basis of a term. A more serious problem is that there is no guarantee that (1*) will work in general. It does seem to work in many cases. Thus, for example, suppose we have a simple arithmetical language L whose non-logical vocabulary consists of a one-place function letter 'S' and a name '0'. Then it would be natural to say that 'S' is the basis of '0' and that a structure M partially accords with the semantics of L if M('0')= the "zero" of what M assigns to 'S'. (That is, if 'S' is assigned a set of ordered paris, M('0') is the unique thing which is the first member of one such ordered pair, but is not the second member of any such ordered pair.) This essentially tells us what the function f is for the language L.

There is, however, a serious problem for this approach. Suppose that both 'gavagai' and 'gavago' partially signify the set of rabbits, the set of rabbit parts and the set of rabbit stages and that the native equivalent of 'All gavagai are gavago' is false. Suppose now that we attempt to interpret that fragment L of the native language whose non-logical vocabulary consists only of the predicates 'gavagai' and 'gavago'. We have seen above that if we let both 'gavagai' and 'gavago' be independent, it will turn out that 'All gavagai are gavago' is not false. So let us assume that 'gavagai' is the basis of 'gavago' and consider a structure M which assigns the set of rabbits to 'gavagai'. Intuitively M could assign either the set of rabbit parts or the set of rabbit stages to 'gavago' and still partially accord with the semantics for L. However, the function f for L must assign one of these sets (and not the other) to 'gavago' and so it seems that either the structure which assigns the set of undetached rabbit parts to 'gavago' or the structure which assigns the set of rabbit stages to 'gavago' does not partially accord with the semantics for L. This is an undesirable result.

This problem could be solved if we dropped the requirement that there be a single function for a language. If we allowed that there could be two functions for the language L above the problem would be solved. However, we are still left with the primitive notions of a function for a language and a term being the basis of another term. Until something more is said about these notions, there is no guarantee that even this further complication will be adequate in general. Furthermore, throughout all of this the notion of partial signification plays a minimal role and, as Field says, "the existence of such correspondence relations as

[this] is not a cause for much satisfaction unless we can use them in an explanation of truth and falsehood" ([7]: 209). What is unclear here is whether the notion of partial signification will actually do any work when it is finally said what a function for a language is.

The notion of partial signification was introduced to deal with situations where different things have equal right to be the signification of a term (or predicate). The problem we have run into here is, roughly, that we cannot in general form structures which partially accord with the semantics for a language by making arbitrary choices among the partial significations of the terms of that language. Intuitively, we have to look at the structure as a whole and see whether it is adequate. In the light of the difficulties raised above, I think then that it is best for our purposes to take the notion of 'partial accordance' (or as it will be called here 'acceptability') as primitive. (2) and (3) then remain intact. We no longer need (1*) although the idea behind it may be helpful in explaining in certain cases which structures are acceptable for the semantics for a language. Also, 'partial signification' can be defined: a term _t_ partially signifies _a_ in _L_ if there is a structure _M_ which is acceptable for the semantics for _L_ and _t_ signifies _a_ in _M_.

Just as partial signification is a generalization of signification, so the notion of a structure being acceptable is a generalization of the notion of a structure being intended, as, for example, when it is said that a certain model is the intended or standard structure for a language. The notion of being an intended structure is a special case of being an acceptable structure: it is the case where it is the only acceptable structure. In general, however, we will speak of acceptable structures in cases where, for one reason or another, it is not appropriate to speak of _the_ intended structure of the language.

As Field points out, one possible application of the notion of an acceptable structure is to problems of vagueness. "Red", for example, does not have a single, non-fuzzy, set as its extension. To interpret a language which has 'red' as its only non-logical vocabulary, we would not interpret it by a single structure, but by a class of acceptable structures. To be acceptable, a structure must include some things in the extension of 'red', e.g. ripe tomatoes, but acceptable structures could diverge over whether to assign certain other things to the extension of 'red'. A sentence of the language would be true if it was true in all (or perhaps "most") of the acceptable structures for the language, false if it was false in all (or perhaps "most") of the acceptable structures for the language and indeterminate otherwise. This may serve to indicate the possible applications of the notion. I believe that the idea is fairly clear. We will have more to say about the notion of an acceptable structure, as applied to the case of interpreting mathematical languages later.

We have now arrived at a generalization of a technique for interpreting first order languages which is employed by Michael Jubien as a way of formally representing informal mathematical theories which are believed not to have certain ontological commitments. Both the question of ontological commitment and this application to an account of mathematical truth will be of interest to us later, but for now we are interested only in the general technique.

Jubien's approach is as follows. In some cases, constants and predicates of a language do not have referents and extensions (for whatever reason) so that it is not possible to interpret this language (in accordance with the semantics for that language) by a single structure which assigns objects to constants and extensions to predicates. Even so, in some of these cases the language (and a theory stated in that language) may provide enough information so that it is possible to say what the intended isomorphism type of a structure of that language would be. For such languages (and Jubien thinks that mathematical languages are prime examples of this) it may be appropriate to interpret them not by a single structure but by a set of structures of the intended isomorphism type. Subject to a complication to be noted shortly, truth of a sentence is characterized as truth in all such structures. Falsehood is falsehood in all such structures. Note that the requirement that the interpreting structures be isomorphic implies that if a sentence is true (false) in one of these structures, it is true (false) in all such structures. Thus no room is left for sentences which are neither true nor false.

It is clear in what sense Jubien's technique is a special case of the technique discussed above. Jubien's technique coincides with the latter technqiue in the special case in which the acceptable structures for a language are pairwise isomorphic. It is only a special case, for in general there is no requirement that the acceptable structures of a language be isomorphic. In fact, there is no requirement that the acceptable structures for a language be elementarily equivalent and it is this which allows for the possibility that there are sentences of a language interpreted by a set of acceptable structures which are neither true nor false.

Of course, the terminology introduced above carries over to Jubien's technique. Thus we may say, for example, that the constants of the theories (if there are any) which are properly interpreted in accordance with Jubien's technique have (multiple) partial referents. Although the constants of such theories do not have referents, they do have partial referents.

The complication in Jubien's approach which was noted above is that it involves the notion of necessity. He is interested in interpreting theories which (1) for whatever reason are not appropriately interpreted by a single structure (2) do, however, provide enough information so that it is possible to tell what the intended isomorphism type of a model of that theory would be (3) but which could be true even if there actually were no model of that intended isomorphism type. The technique as thus far described encounters difficulties over (3). If there are no acceptable structures for a theory, then on the characterization of truth as truth in all acceptable structures, every sentence of such a theory would be true. Alternatively if we characterize truth as: A is true in L if there is an acceptable structure for L and A is true in all acceptable structures to L, and falsehood as: A is false in L if A is false in all acceptable structures for L, we would get the result that every sentence of such a theory is false. Both of these results are unacceptable.

Jubien suggests that to interpret such theories we consider all possible models of the intended isomorphism type. A sentence of a language L is then true if necessarily for every model of the intended isomorphism type that sentence is true in that model.

Once again, this may be generalized to the case where not all the acceptable structures for a language are isomorphic. Truth (and falsehood) are defined as before.

Our results may now be summarized (these definitions are preliminary and will be superceded in later chapters): An acceptably interpreted first order language (AIL) is an ordered pair $\langle \underline{L}, \underline{M} \rangle$ where $\underline{L}$ is a first order language and $\underline{M}$ is a set of structures of $\underline{L}$. A sentence $\underline{A}$ of an AIL $\langle \underline{L}, \underline{M} \rangle$ is true if $\underline{M} \neq \emptyset$ and $\underline{A}$ is true in all $\underline{m}$ in $\underline{M}$; it is false if it is false in all $\underline{m}$ in $\underline{M}$.

A modally acceptably interpreted first order language (MAIL) is an ordered pair $\langle \underline{L}, \$ \rangle$ where $\underline{L}$ is a first order language and \$ is a sentence of the form: a structure $\underline{S}$ is acceptable for $\underline{L}$ if __________. A sentence $\underline{A}$ is true in $\langle \underline{L}, \$ \rangle$ if necessarily for all structures $\underline{S}$ if $\underline{S}$ is acceptable for $\underline{L}$, then $\underline{A}$ is true in $\underline{S}$; it is false in $\langle \underline{L}, \underline{\$} \rangle$ if necessarily for all structures $\underline{S}$ if $\underline{S}$ is acceptable for $\underline{L}$, then $\underline{A}$ is false in $\underline{S}$.

Before proceeding to a fuller discussion of AIL's and MAIL's, other anticipations of the technique employed here may be noted. When Paul Benacerraf says, "Number theory is the elaboration of the properties of all structures of the order type of the numbers. The number words do not have single referents," ([2]: 70-1) one plausible way of interpreting him is to say that he thinks that the language of arithmetic is properly understood as an AIL and that numerals do not have single referents but that they do have partial referents.

When Hilary Putnam says, "Now the natural way to interpret set-theoretic statements in the modal-logical language is to interpret them as statements of what would necessarily be the case if there were standard models for the set theories in question," ([22]: 20) he appears to be considering the possibility of interpreting set theory as a MAIL.

Since interpretations by sets of structures may be somewhat unfamiliar and since such interpretations in general allow for sentences to be indeterminate, i.e. neither true nor false, some elementary points about the logic of languages interpreted in this way may be in order. We restrict our attention to cases where the set of structures used to interpret a language is non-empty. Thus we consider interpretations by arbitrary, non-empty sets of structures. We say that a sentence is valid* if it is true in all such interpretations.

(1) All and only classical first order logical truths are valid*. The proof of this is obvious. This differentiates the interpretation studied here from other familiar interpretation which allow a third truth value, but relative to which not all classical logical truths are valid.

(2) Obviously, -A is true (false) (indeterminate) if A is false (true) (indeterminate). The situation with respect to disjunctions is somewhat more complicated. If either (or both) A or B are true, then AvB is true. If both A and B are false, their disjunction is false. If one is indeterminate and the other is false, their disjunction is indeterminate (since it is true in the non-empty set of structures in which the indeterminate disjunct is true and false in the non-empty set of structures in which the indeterminate disjunct is false). Thus far disjunction operates on this interpretation just as it operates in Kleene's strong three-valued logic. The remaining case in which both A and B are indeterminate admits of two possibilities. On this assumption, AvB cannot be false (since it is true in any of the structures in which A is true, for example), but it can be either indeterminate or true. If A is indeterminate, then AvA illustrates the first possibility, while Av-A illustrates the second possibility. More generally, if A and B are indeterminate, the AvB is true just in case the union of the set of structures in which A is true and the set of structures in which B is true is equal to the full set of structures which interprets A and B; otherwise AvB is indeterminate. This once again differentiates interpretations by sets of structures from other interpretations which allow a third truth value.

CHAPTER II

In later chapters we will be discussing the ontological commitments of theories. It is thus important here to discuss the notion of ontological commitment. In doing this we will also be introducing certain technical notions and metaphysical presuppositions to be used in later chapters. In particular a special notion of a world will be introduced which plays an important role in later chapters. The discussion here follows the work of Michael Jubiel ([12], [13]) quite closely and is partly supplementary to and partly critical of that work.

It is taken here as established that ontological commitment is an intensional notion. In assuming this we are siding with Cartwright [3], Jubien [11], (to a certain extent) Parsons ([18], [19]), Scheffler and Chomsky [14] against Quine ([23], [27]) who claims that ontological commitment is a notion which belongs to the theory of reference. What this means is that in stating conditions for the satisfaction of '$\underline{T}$ is ontologically committed to $\underline{x}$' we need to use some notion such as entailment, analyticity or possibility.

With Jubien we distinguish between commitment to kinds and commitment to particulars. Intuitively, 'there are people' carries commitment of the first kind, while 'Nixon exists' carries commitment of the latter kind. We also distinguish between de re and de dicto commitment to particulars. Intuitively, 'Carter exists' has a commitment de re to Carter, but (given an "attributive" use of the definite description 'the President') 'the President exists' does not have a de re commitment to Carter, but does have a de dicto commitment to there being a unique President. The important difference here is that the latter theroy places no restriction on who is the President, while the former theory says that some particular individual exists.

Parsons considers (without endorsing) the following criterion of commitment to kinds: ([19]: 74).

> (*) T is ontologically committed to kind k if T entails that there are things of kind k.

The analogous criterion of de re commitment is:

> (**) T is ontologically committed de re to a if T entails that a exists.

In what follows we will largely be concerned with discussing this criterion of de re commitment. A few obvious remarks about (**) are that it is unclear just what sense of 'entails' is employed. For example, if we accept the equivalence of 'T entails p' with 'in every possible world in which T is true, p is true' then corresponding to different senses of 'possible world' we will obtain different senses of 'entails' and thus different versions of (**). It is also, of course, conceivable that 'entails' is used in some entirely different way. Further evaluation of (**) thus requires that more be said about how 'entails' is being used. It is also unclear what the ranges of the variables '<u>T</u>' and '<u>a</u>' are which are used in (**).

In later chapters we will be concerned with evaluating the ontological commitments of certain formal theories which are interpreted by sets of intensional interpretations. Thus, it is natural to discuss (**) by restricting the range of the variable '<u>T</u>' to such formal theories and then asking whether there is some sense of 'entails' in which (**), so restricted, is true. In doing this we will also be forced to say what the range of '<u>a</u>' is in (**). We could then ask whether there is some generalization of (**) which is true when applied to a more general class of theories. Thus, we proceed by introducing this formal method of intensionally interpreting theories and then ask whether various versions of (**) are plausible when applied to such theories.

Of course, primarily what we are interested in are the ontological commitments of informal theories which people actually hold. Initially at least it would appear that it is only such theories which have ontological commitments. From this point of view we regard the ontological commitments of formal theories as derived from the commitments of the informal theories they represent.

Now a given informal theory might be formalized in number of different ways and different criteria of commitment might be required to give the correct results for these different formal theories. For example, it might be that informal theory _T_ has two different formalizations _T'_ and _T''_ and that there are two different criteria of commitment _C'_ and _C''_ such that the commitments of _t'_ relative to _C'_ and the commitments of _T''_ relative to _C''_ are precisely those of _T_, but that applying either _C'_ to _T''_ or _C''_ to _T'_ gives incorrect answers. If we then think of the commitments of formal theories as relative to that of informal theories they represent, then we need to ask not whether such and such a criterion is correct absolutely, but whether it is correct relative to a certain method of representing informal theories as formal theories. This will be disucssed (and an actual example given) later in this chapter.

Thus far we have spoken as if we can speak of the ontological commitments of formal theories only relative to informal theories they represent. From such a point of view the only point of introducing formal theories is to introduce some regimentation into discussions of ontological commitment. In fact, there is another reason to introduce these formal theories. In certain cases it may be controversial just exactly what the ontological commitments of an informal theory are. (The examples to be discussed at length in this dissertation are mathematical theories.) Given this we might want to discuss the ontological commitments of formal theories independently of whether they are representations of these problematic informal theories and only later (and partly in the basis of what the commitments of the formal theories are) ask whether the formal theories adequately represent the informal theories. The assumption here which we make is that the notion of commitment applies not only to informal theories and to formal theories relative to some translation, but also to formal theories outright. The reader must judge whether or not this is true. The claim here would be that these formal theories are enough like informal theories so that we can ask of them (outright) whether they have or lack certain ontological commitments. If we can do this, then we can ask whether or not certain criteria of commitment for such theories are correct absolutely and not just relative to informal theories they represent. Throughout much of this chapter it will be this question of relative correctness which will be considered. In later chapters, however, the criteria developed here will be applied outright to formal theories.

We take the notion of an intensionally interpreted theory (IIT) essentially from Jubien, ([13]: 516-7), the major difference from Jubien's account being that we do not treat constants as rigid designators and also allow that '-(Ex)(x=a)' can be true at some worlds.

Thus, classical first order theories will be a subset of the class of theories defined here. The added breadth gained by allowing certain constants not to denote at some worlds is an advantage when we attempt to formally represent informal theories such as 'God does not exist'. In order to do this, however, a decision must be made concerning how to treat non-denoting constants semantically. We make the decision that for atomic predicates, $\underline{P}^n$, a sequence does not satisfy $\underline{P}^n(\ldots\underline{a}\ldots)$ if $\underline{a}$ is undefined. Otherwise the clauses in the definition of satisfaction are the same as in the definition of satisfaction for classical first order theories. Truth in a structure then is just satisfaction by all sequences.

Using the notion of a quantificational model structure $\langle \underline{G}, \underline{K}, \underline{R}, \Psi \rangle$, we define an intensional interpretation of a theory $\underline{T}$ to be an ordered set

$$I = \langle \Delta; \rho^1_1, \rho^1_2, \ldots; \rho^2_1, \ldots; \ldots; \alpha_1, \alpha_2, \ldots \rangle$$

which satisfies the following conditions: (1) for each $\underline{H}$ in $\underline{K}$, $\Delta(H) \subseteq \Psi(H)$; (2) for each $\underline{H}$ in $\underline{K}$ and $\underline{i}$, $\alpha_i(H) \varepsilon \Delta(H)$ if α_i is defined at $\underline{H}$ (the α_i are not in general total functions); and (3) for each $\underline{H}$ in $\underline{K}$ and $\underline{i}$, $\underline{j}$, $\rho^i_j, (\underline{H}) \subseteq [\underline{I}(\underline{H})]^i$. In general, we denote

$$\langle \Delta(\underline{H}); \rho^1_1(\underline{H}), \rho^1_2(\underline{H}), \ldots; \rho^2_1(\underline{H}), \ldots; \alpha_1(\underline{H}), \alpha_2(\underline{H}), \ldots \rangle$$

by '$\underline{I}(\underline{H})$'. We then define: a sentence is true at a world $\underline{H}$ if it is true in $\underline{I}(\underline{H})$.

$\Delta(\underline{H})$ is the range of the quantifiers at $\underline{H}$. It may be a proper subset of $\Psi(\underline{H})$ if we do not want the quantifiers to range over all the things which exist at a world (as for example is the case in Zermelo-Fraenkel set theory, the quantifiers of which range only over sets).

The justification for allowing constants not to denote was the facilitate the representation of such theories as 'God does not exist'. If this is going to be interesting, we will want '-(Ex)(x=g)' to be true at some worlds. In order to do this, a free logic should be adopted for dealing with ITT's.

The point of this definition is that instead of simply assigning extensions to predicates and denotations to constants, predicates are assigned functions from worlds to extensions and constants are assigned partial functions from worlds to denotations. Roughly speaking, an intensional interpretation could be thought of as a function from worlds to extensional interpretations.

We are then interested in seeing if some version of (**) is plausible when applied to IIT's. We might then make (**) more precise then by saying

> (1) an IIT, T, is ontologically committed de re to a if $a \in \Psi(H)$ for every H at which t is true.

The analogous version of (*) would be

> (2) an IIT, T, is ontologically committed to kind k if for every H at which T is true, the set of things which are of kind k at H is non-empty.

Whereas (1) is well-defined for any IIT, this is not so for (2). In order to apply (2) we need to ask whether there are things which are of kind k at H for worlds H at which T is true. This, however, will not make sense for ITT's which are defined relative to certain model structures.

Let me give an example before moving to the general point. Consider the model structure

$\langle 0,\{0,1\},\{0,1\}x\{0,1\}\rangle,\{\langle 0,\{\text{Nixon},\text{Ford}\}\rangle,\langle 1,\{\text{Nixon}\}\rangle\}$

Then define an intensional interpretation for a theory whose sole non-logical axiom is '(Ex)Fx' as follows. For each H, let $\Delta(H)=\Psi(H)$ and let the interpretation of 'F' also be $\Psi(H)$.

We have now specified an IIT. Note that there is no trouble applying (1) to this theory. All that we have to do is to ask whether, for example, Nixon is in $\Psi(H)$ for every H at which the theory is true. He is and so according to (1) this theory is committed de re to Nixon (though not to Ford).

Ask now whether this theory is committed to the kind politician. In order to do this we need to find out if Nixon is a politician in the world 1. Is he? There is no saying because in specifying the IIT we did not have to say whether Nixon is a politician in 1.

The moral is that we can only apply (2) to IIT's which are defined relative to model structures for which we can ask for arbitrary kinds k and worlds H whether there are things of kind k at H. That is, we need to be able to ask not only whether arbitrary sentences of the object language are true at worlds (we get that when we define the IIT) but we also need to be able to ask whether sentences of the metalanguage, e.g. 'there are things of kind k at H', are true at the worlds in the model structure used to define the IIT.

For many purposes, e.g. defining validity, we need only the technical notion of an object language sentence being true at a world in a model structure. For purposes of assessing the ontological commitments of theories in accordance with (2), we need the informal notion of a sentence of the metalanguage being true at a world. What sort of things are such that sentences of the metalanguage are true at them? It is natural to turn here to the metaphysical notion (as opposed to the purely technical notion needed for defining validity) of a possible world. Sentences of the metalanguage are true at possible worlds (in a metaphysical, but not a technical, sense of 'possible world').

In short, for the purposes of assessing ontological commitments in accordance with (2), it appears that we need a notion which is much like the metaphysical notion of a possible world used, for example, by Plantinga. It might then be proposed that we apply (1) and (2) only to IIT's which have been defined relative to the model structure the worlds of which are the metaphysically possible worlds.

If this approach is taken, there is no reason to believe that (1) is an adequate criterion of de re commitment. There may well be things which exist in every metaphysically possible world. If so, (1) will imply that every IIT is committed de re to each of these necessary existents. Note that we are making a double point here. First, this argument shows that (1), so understood, is inadequate because there is no way of interpreting informal theories as formal theories in such a way that (1) applied to the formal theory implies that that formal theory has exactly the de re commitments that, intuitively, the informal theory has (assuming there are necessary existents). Second, apart from considerations of representing informal theories, it is unacceptable that every IIT should be committed de re to, say, God.

We can then either give up (1) or look for a different model structure. In fact, there is independent reason to look for a different model structure. Even if there are abstract entities which metaphysically necessarily exist, we might still want to consider the ontological commitments of theories which imply that there are no abstract entities, e.g. 'there are no abstract entities'. Even if God necessarily exists, we might still want to consider that ontological commitments of theories which imply that there is no God.

In order to meet these desiderata, we want to define an IIT relative to a structure which embodies in some sense a broader notion of possibility that the notion of metaphysical possibility. This point is also made by Jubien ([13]: 527-8), who does not, however, say what that model structure would be. What follows then can be understood as, among other things, a supplement to Jubien's theory of commitment.

What do we want of such a model structure? At least this: for every set of entities there is a world at which all and only the members of that set exist. We also want to let there be worlds at which, to take the example just given, there are concrete entities but no abstract entities. We also want a model structure relative to which for arbitrary kinds *k* we can say whether there are things of kind *k* at a given world.

Typically, we will also want to specify an IIT by saying things like, "for every world *H*, let the interpretation of '*Bxy*' at *H* be the set of ordered pairs $\langle x, y \rangle$ such that at *H* *x* is the brother of *y*." Thus we need not only to be able to specify whether there are things of kind *K* at *H*, we also need to be able to specify the interpretations of the predicates of the language in the wholesale fashion just indicated. Note that this cannot be done within the approach taken here if we take the technical notion of a world discussed earlier. If we used that notion, essentially the most that could be done would be to give a piecemeal interpretation for the predicates, e.g., "at H, let the interpretations of 'B' be $\{\langle a,b\rangle, \langle b,a\rangle\}$, at H_2, let . . ."

A final condition is that "kind-necessities" be preserved in every world in the model structure. I follow Quine [24] and Jubien ([12]: 86-7) in thinking that the theory 'there are fish' has a commitment not only to the kind fish but also to the kind animal. Now if we are to evaluate commitment to kinds in accordance with (2), this means that if something is of the kind fish at a world H, then it must be of the kind animal at H.

The proposal to be advanced here attempts to meet these conditions by making use of the notions of a kind, a property and metaphysical possibility. One might conceive of a reduction of this rather liberal set of metaphysical notions as follows. Kinds are simply certain properties and if the only properties which are of interest in this context are those expressed by English predicates, then properties could be reduced to predicates. I have no suggestion as to how to reduce the notion of metaphysical possibility.

A world H relative to a language L is then defined as an ordered pair $\langle x, f\rangle$ where (1) x is a set and (2) f is a function which maps each kind $\underline{k}$[1] to a subset of $\underline{x}$ (the set of things of kind $\underline{k}$ at $\underline{H}$) and which maps each primitive predicate '$\underline{P}^n$' of $\underline{L}$ to an ordered pair $\langle \underline{P}^n, \underline{y} \rangle$ where $\underline{P}^n$ is an n-ary property (the meaning of '$\underline{P}^n$') and $\underline{y}$ is a subset of $\underline{x}^n$ (the things which have $\underline{P}^n$ at $\underline{H}$). $\underline{f}$ must meet the following conditions:

(a) for every member $\underline{z}$ of $\underline{x}$ there is a kind $\underline{k}$ such that $\underline{z} \varepsilon \underline{f}(\underline{k})$;

(b) for any set $\underline{S}$, $\underline{S}'$ of kinds and thing $\underline{z}$, if $\underline{z} \varepsilon \underline{f}(\underline{k})$ for each $\underline{k}$ in $\underline{S}$ and it is metaphysically necessary that if something is of each of the kinds in $\underline{S}$, then it is of at least one of the kinds in $\underline{S}'$, then for some $\underline{k}'$ in $\underline{S}'$, $\underline{z} \varepsilon \underline{f}(\underline{k}')$,

---

[1]This construction relies on the assumption that there is a set of all kinds. If there were no such set, then the functions $\underline{f}$ used to define worlds would not exist. I am unsure as to whether this assumption is justified since I am uncertain what some of the important properties of kinds are. For example, is the "negation" of a kind a kind? What about disjunctions? Is being Nixon a kind? I am not sure. If this assumption is unjustified, the definition of a world would have to be altered; perhaps along the lines of the definition of an appropriate model structure given below.

Another question is: What things are members of $\underline{x}$, where $\langle \underline{x}, \underline{f} \rangle$ is a world? Clearly anything which actually exists could be a member of such an $\underline{x}$. The issue here is whether there are non-existent entities and, if so, whether such entities are members of sets. If both of these conditions are satisfied, then we might well want such entities to be the members of some $\underline{x}$, where $\langle \underline{x}, \underline{f} \rangle$ is a world. Doing this would allow for the possibility of having de re commitments to non-existent entities. I have deliberately left the constructions in this chapter neutral between whether or not non-existent entities figure in the worlds we define.

(c) for any primitive predicate 'P^n' of $\underline{L}$, kind $\underline{k}$ and thing $\underline{z}$, if $\underline{f}('\underline{P}^n') = \langle \underline{P}^n, \underline{y} \rangle$ and $(\underline{Ex}_1)\ldots(\underline{Ex}_{n-1})(\underline{x}_1\ldots\underline{z}\ldots\underline{x}_{n-1}\ \underline{y})$ and it is metaphysically necessary that for all $\underline{x}_1\ldots\underline{x}_{n-1}$, $\underline{z}$ if $\underline{x}_1\ldots\underline{z}\ldots\underline{x}_{n-1}$ stands in the relation $\underline{P}^n$ then $\underline{z}$ is of kind $\underline{k}$, then $\underline{z}\varepsilon\underline{f}(\underline{k})$.

(a)-(c) are intended to force $\underline{f}$ to preserve kind-necessities. Intuitively (a) and (c) are supposed to force things into some kind or other and (b) is supposed to keep them in the "correct" kinds given that they are in one given kind. (c) is also added so that, for example, (a natural formalization of) '$(\underline{Ex})\ (\underline{Ey})(\underline{x}\varepsilon\underline{y})$' might be committed to the kind set.

Some consequences of (b) (not had by its simpler sub-case: if $\underline{z}\varepsilon\underline{f}(\underline{k})$ and it is metaphysically necessary that anything of kind $\underline{k}$ is of kind $\underline{k}'$, then $\underline{z}\varepsilon\underline{f}(\underline{k}')$)are:

(i) assuming that necessarily all rational animals are humans, but not necessarily all rational things are humans, if $\underline{z}\varepsilon\underline{f}$(rational) and $\underline{z}\varepsilon\underline{f}$(animal), then $\underline{z}\varepsilon\underline{f}$(human);

(ii) If $\underline{z}\varepsilon\underline{x}$, then either $\underline{z}\varepsilon\underline{f}$(concrete) or $\underline{z}\varepsilon\underline{f}$(abstract). (This result also relies on (a). (a) forces everything to be a member of some kind and (b) forces the member of any kind to be either concrete or abstract. We explicitly rely on (ii) in Chapter III where it is argued that if there are infinitely many things at $\underline{H}$, but only one concrete thing, then there are abstract things at $\underline{H}$.

It would now be natural to let $\langle \underline{G}, \underline{K}, \underline{R} \rangle, \Psi$ be defined by stipulating that $\underline{K}$=the set of all worlds; $\underline{R}=\underline{K}\mathrm{x}\underline{K}$; $\Psi(\underline{H})=\underline{x}$, where $\underline{H}=\langle \underline{x}, \underline{f} \rangle$ and $\underline{G}=\langle \underline{y}, \underline{f} \rangle$ where $\underline{y}$=the set of things which actually exist and for each kind $\underline{k}$, $\underline{f}(\underline{k})$=the set of things which actually are of kind k.

Such a natural approach will, however, lead to paradox. I assume that the set theory used in defining this model structure is Zermelo-Fraenkel set theory with individuals. Then there is no set of all sets. Therefore, there is no set y which is the set of things which actually exist. Therefore, G does not exist. Also there are kinds, e.g. set such that there is no set of all the things of that kind. Thus f (where $G=\langle y, f\rangle$) is only a partial function. Furthermore, there is a difficulty with K even apart from the difficulty with G: there will be no set of all worlds. This, of course, also affects the specification of R and Ψ.

This is quite a serious problem. On the one hand, if (1) is to be applied as a criterion of de re commitment, there must be many different kinds of worlds. This was seen above where it was noted that there is no reason to believe that (1) is adequate if we consider only the metaphysically possible worlds. On the other hand, paradoxes must be avoided.

The most elegant solution would be obtained if there was a clearly understood theory which allowed there to be totality which "contained" all sets and which allowed such totalities to be "members" of other totalities. (Thus G is a totality which is a "member" of K which is also a totality.) In the absence of such a theory a slightly complicated set-theoretical construction is required.

On this approach it is natural to define G and K in such a way that they are guaranteed to be sets, but to put enough conditions on K to guarantee that the members of K reflect the important properties of all the worlds. That is just how we proceed. Proceeding in this way forces, however, giving up speaking of the model structure (relative to a language). Instead we define the notion of an appropriate model struxture.

In an appropriate model structure $G=\langle x, f\rangle$ where x is simply some set of things which actually exist and for all z in x and kinds k, $z\varepsilon f(k)$ if z is actually of kind k. This guarantees that G will be a set. G is, roughly, a fragment of what would ordinarily be regarded as the actual world. In an appropriate model structure K is simply some set of worlds which meets the conditions below. As before $R=K\times K$ and $\Psi(H)=x$, where $H=\langle x, f\rangle$.

We require that K satisfy:

(R1) For any mutually disjoint sets a, b, it is not the case that in every H in K in which all of the members of a exist, all of the members of b exist.

(R2) For any sets S, S' of kinds, if there is a world at which all of the members of S are exemplified, but none of the members of S' are exemplified, then there is an H in K where $H=\langle x, f\rangle$ and for every k in S, $f(k)\neq\emptyset$ and for every k' in S' $f(k')=\emptyset$.

(R3) For anything x and kind k if there is a world at which x is not of kind k and there is a world H in K at which x exists, then there is a world in K at which x is not of kind k.

(R1) is designed to guarantee that there are no "ontological dependencies" in K: saying that some set of things exists will not guarantee that some other thing exists. (R2) and (R3) are designed to guarantee that the only kind dependencies that there are in K are those which exist throughout all the worlds.

It is important to verify that K in fact is a set. (It might be thought that (R1)-(R3) force K to be too large to be a set.) This is easily done. One way of showing this is to begin with enough worlds to satisfy (R2) and then add worlds to that set so as to satisfy (R1). Adding worlds cannot keep the resultant set from satisfying (R2). Finally we add worlds so as to satisfy (R3). The set achieved thereby clearly is a set and also satisfies these three conditions.

Is an appropriate model structure the general sort of model structure we seek? (We, of course, obtain a particular model structure relative to a language by assigning meanings to the predicates of that language.) First, note that we wanted a model structure relative to which we could ask whether there are things of kind k at H for arbitrary kinds k and worlds H. This notion (which we have already used in stating the conditions above) is now easily defined:

there are things of kind $\underline{k}$ at $\underline{H}$ ($\underline{H}=\langle \underline{x}, \underline{f}\rangle$) if $\underline{f}(\underline{k})\neq\emptyset$.

An appropriate model structure also embodies the broad notion of possibility we need for discussing ontological commitment. First, (R1) guarantees that there are no ontological dependencies. Second, (R2) implies that there are worlds at which there are concrete things, but no abstract things. On the other hand, in any world at which there are fish there are animals. Note also that by (R3) for no thing $\underline{x}$ and kind $\underline{k}$ is $\underline{x}$ of $\underline{k}$ at every world at which it exists (provided that it is not metaphysically necessary that everything is of $\underline{k}$). So, for example, in some world Nixon is not of the kind human, nor even of the kind concrete entity. We could summarize this by saying that an appropriate model structure preserves kind-necessities, but not individual-necessities.

This implies that 'Nixon exists' (interpreted in the obvious way as an IIT) is not committed to the kind human according to (2). Is that correct? I think that it is, but this is a difficult case. If one thinks that it is incorrect, the opposite result can be achieved by adding a new condition which must be met by the functions $\underline{f}$ used to identify the worlds. This condition would be:

(R*) For any set $\underline{S}$ of kinds and member $\underline{z}$ of $\underline{x}$, if it is metaphysically necessary that if $\underline{z}$ exists, then it is of one of the kinds in $\underline{S}$, then $\underline{z}$ is a member of $\underline{f}(\underline{k})$ for some $\underline{k}$ in $\underline{S}$.

Note also some oddities of the concept of a world which I propose we use for evaluating ontological commitments. If we suppose that being an animal and being a vegetable are kinds, then on the definition of a world given above there will be worlds at which something is both an animal and a vegetable. In fact, there is nothing to keep something from being both a vegetable and a non-vegetable at a world (assuming that the latter is a kind).

I do not think that this result has any undesirable consequences for ontological commitment. In fact, it has the desirable consequence that we can evaluate the ontological commitments of the rather strange theories which may, in fact, be held. This result simply indicates that a model structure which is appropriate for considerations of ontological commitment might not be appropriate for other purposes. In any case, this result could easily be blocked by adding a new condition (which is easily stated) which must be met by the functions _f_ used to identify the worlds. Adding this condition would not, as far as I can tell, affect our further discussion of ontological commitment.

Note finally the consequences if we have a model structure relative to a language containing a two-place predicate '_Bxy_' and if that predicate is assigned the property _brother_ by the worlds of the model structure. Then there will be worlds at which both I and my actual brother exist, but are not brothers. There will also be worlds at which I have a brother, but (assuming that having a parent is a kind) I have no parent. If, however, it is metaphysically necessary that if something has a brother then it is an animal, then (c) implies that in any world in which I have a brother I am an animal.

I do not think that we need conditons beyond (a)-(c), but it is important to see that the approach outlined here is fairly flexible. Such conditions should be added only if they are needed to give the correct results for ontological commitment. Roughly speaking, we want to impose minimal structure on these worlds so as to be able to evaluate the commitments of as many theories as possible.

Michael Jubien has raised some powerful objections to (1) ([13]: 518-9). I want now to see just how convincing Jubien's objections to (1) are if we apply (1) to IIT's defined relative to the model structure we have just described. (Of course, Jubien did not consider (1) restricted in this way, so that we will not directly be criticizing his claims.)

Jubien claims that although an IIT may "say nothing about" some thing, that thing might "happen" to exist in every world at which that theory is true. In such a case (1) would imply that the IIT is committed de re to that thing in spite of the fact that intuitively the IIT does not say enough about that thing to be committed to it. While I think that this point is correct when applied to IIT's defined relative to arbitrary model structures, it is not clearly correct when applied to the particular case we are considering. It will be useful to distinguish two sorts of this "inflated domain" counterexample to (1).

(i) Necessary beings. If there is anything which exists in every Ψ(H) that thing will exist in every Ψ(H) at which an IIT is true, regardless of whether that IIT says anything at all about that thing.

Clearly, however, there are no such necessary beings in an appropriate model structure ⟨G, K, K⟩,Ψ. Thus, there can be no such inflated domain counterexamples to (1) given our restriction.

(ii) "Popping-up" entities. Even if something does not necessarily exist, it might "pop up" (exist) in every H at which an IIT is true, regardless of whether that IIT says anything at all about that thing. Here is an example of this:

Let the sole non-logical axiom of T be '(Ex)Fx'. Define I by letting Δ(H)=ψ(H) for every H and letting 'F' be assigned the set of things which are cows at H at any world H at which Nixon exists and letting 'F' have the null extension at any world at which Nixon does not exist. Then, according to (1) ⟨T, I⟩ is committed de re to Nixon.

(1) might be rejected on the basis of such a case on the grounds that ⟨T, I⟩ lacks the syntactic means of

"picking out" Nixon and that a necessary (though not a sufficient) condition of a theory's being committed de re to, say, Nixon is that it have the syntactic means of "picking out" Nixon.[2]

I do not think this example conclusively shows that (1) is inadequate. Let me introduce a new predicate into English. We say, *x* is an *F* if *x* is a cow and Nixon exists. Now consider the informal English theory 'there are *F*'*s*'.

Given the meaning of '*F*', this theory is, intuitively, committed de re to Nixon. But furthermore, there is nothing wrong with representing this informal theory as the IIT $\langle T, I \rangle$ lately defined. If so, then it is not obviously unacceptable for $\langle T, I \rangle$ to be committed de re to Nixon.

---

[2] In fact Jubien's own criterion fails to accord with this principle. Let the sole non-logical axiom of *T* be $(Ex)(Ey)(x \neq y \ \& \ (z)(Fz \leftrightarrow z=x \vee z=y))$. Define *I* by letting $\Delta(H)=\Psi(H)$ for every *H* and letting the extension of 'F' be {Nixon, Ford} in any world at which those two men exist and $\emptyset$ otherwise. Then Jubien's criterion (5) implies that $\langle T, I \rangle$ is committed de re to both Nixon and Ford. Yet given that $\langle T, I \rangle$ lacks the syntactic means of picking out either of these men, Jubien's principle implies that $\langle T, I \rangle$ should not be committed to either of these men.

I suggest the following criterion as according with this principle. (Of course, I do not endorse this criterion.)

(5') $\langle T, I \rangle$ is committed de re to *x* if
i) there is a theorem $(E!x)A$ of *T* and *x* is the unique satisfier of *A* in every *H* at which $\langle T, I \rangle$ is true and
ii) for every *y*, if *x* is existentially dependent on *y*, there is a theorem $(E!x)B$ of *T* and *y* is the unique satisfier of *B* in every *H* at which $\langle T, I \rangle$ is true.

Clause ii) is added to deal with the problem with abstract entities which motivates Jubien to move from his (4) to (5).

In other words, $\langle \underline{T}, \underline{I} \rangle$ is an adequate (I do not claim that it is the only adequate) formal representation of an informal theory which has a de re commitment to Nixon. Thus it is not a conclusive objection to (1) that it implies that $\langle \underline{T}, \underline{I} \rangle$ is committed de re to Nixon.

Let us say that the informal predicate '$\underline{F}$ ' hides commitments. (A non-contrived example of such a predicate, which was pointed out to me by Parsons is 'is British'.) The example just considered then suggests a general answer to the inflated domain objection to (1). The answer consists in saying that any apparent inflated domain counter-example to (1) can be explained as involving a theory which is an acceptable formal rendering of an informal theory which has a predicate which hides commitments.

Note that both Jubien's criterion of de re commitment (5) and my revision of it (5') imply that $\langle \underline{T}, \underline{I} \rangle$ is not committed de re to Nixon. Thus Jubien must find some other way of representing the informal theory 'there are $\underline{F}$'$\underline{s}$'. There is no problem here (in this case, at least). Roughly speaking, in order to apply Jubien's criterion, this theory should be represented as an IIT whose axiom is '$(\underline{Ex})\underline{Cx}$ & $(\underline{Ex})(\underline{x}=\underline{n})$' with '$\underline{C}$' interpreted to mean $\underline{\text{cow}}$ and '$\underline{n}$' treated as a name of Nixon. The general maxim to follow in representing informal theories which may involve predicates which hide commitments if Jubien's criterion is to be applied is to "reveal as much structure as possible".

Thus we would represent 'Tom is British' by first expanding this theory to 'Tom bears R to Great Britain' (where R is whatever relation it is that someone bears to Great Britain if he is British) and then representing that theory just as you would think you would. As we have seen, not all of the structure of an informal theory need be revealed in order to apply (1).

Jubien's criterion and (1) thus give different results for a given IIT. In particular, if $\langle \underline{T}, \underline{I} \rangle$ is committed de re to x according to Jubien's criterion then (1) implies that it is committed de re to x, but the converse does not hold in general.

The above remarks suggest, however, that insofar as what we are interested in is the commitments of informal theories, then the differences between (1) and (5) ((5')) may be unimportant. As noted earlier, we can view the process of evaluating the commitments of an informal theory as consisting of two components. First there is the translation component whereby the informal theory is rendered as an IIT. Then there is the criterion component whereby the commitments of the formal theory are evaluated.

In the example above, we have seen that pairing (1) with one translation procedure and pairing (5) with another translation procedure give precisely the same results for the informal theory 'there are F's'.

If what is taken to be important is the evaluation of the commitments of informal theories, then is there any interesting difference between (1) and (5)? Or do the differences between (1) and (5) cancel out (in the case of informal theories) because different translation procedures can be used in applying these criteria?

It would be impossible to answer this question affirmatively until we had specified the respective translation procedures more carefully. There may, however, be reason to answer this question negatively if we assume that any IIT is a countable theory. Then it is an easy consequence of the Lowenheim-Skolem theorem that (5) implies that no IIT is committed de re to more than countable many things. (This is even more obvious in the case of (5').) However, an IIT could have a de re commitment to uncountably many things according to (1).

Thus, if there is an informal theory (someone might think that set theory is an example) which has a de re commitment to each of uncountable many things (not just a commitment to there being uncountably many things), then there will be no way of representing this theory as an IIT and then applying (5) to gain the result that this IIT is committed de re to uncountably many things. There is no a priori obstacle to gaining this result with (1).

It is my suspicion that this is the only interesting difference between (1) and (5) concerning informal theories. For the time being, however, I want to limit myself to the fairly modest claim that Jubien's objection to (1) is not conclusive if we limit our consideration to IIT's defined relative to the model structure we have defined; by regarding certain IIT's as renderings of informal theories which have predicates which hide commitments we can explain why the apparent inflated domain counterexamples to (1) need not be regarded as genuine counterexamples.

I thus propose (1) as a criterion of de re commitment. Besides the fact that it is in some sense the most obvious criterion and also that it is not vulnerable to inflated domain counterexamples, I have another reason for adopting (1). In what follows, we will be concerned with evaluating the commitments of certain formal theories. We will largely be concerned with showing that these theories do not have certain commitments. By applying (1) which is, as has been noted, a more liberal criterion than (5) is when applied outright to formal theories we make the claim that these formal theories do not have certain commitments less controversial. (I would claim that any more liberal criterion than (1) is clearly too liberal.) If one thinks that (5) (or (5')) is adequate, then it should be applied and it will be seen that if an IIT is not committed to x according to (1), then it is not committed to x according to (5) ((5')). Thus, there are purely strategic reasons for accepting (1) for the purposes of the work to be pursued here.

I also propose that (2) be accepted as the criterion of commitment to kinds. What about de dicto commitment? Given the way we have proceeded thus far, it is natural to regard de dicto commitment as a relation between an IIT and an individual property. (An individual property is a property which is exemplified by at most one thing at any world.) Following the line suggested by (1) and (2), we then say

(3) ⟨_T_, _I_⟩ is committed de dicto with respect to an individual property _P_ if _P_ is exemplified at every world at which ⟨_T_, _I_⟩ is true.

It is important to note how (3) depends on our definition of a world. On our definition, relative to a language _L_ we can speak of the kinds that are exemplified at a world and also of the properties expressed by the primitive predicates of _L_ at a world. On the basis of this we can ask, for example, whether the individual property: _being the unique thing which is not-P and of kind K_ (if _P_ is a property which is expressed by some predicate of _L_) is exemplified at a world.

Extending (1), (2) and (3) to AIL's and MAIL's. Given the intensionality of ontological commitment, in order to assess the ontological commitments of theories interpreted as AIL's or as MAIL's, we need to provide intensional versions of AIL's and MAIL's. Once this is done, (1), (2) and (3) apply immediately,

Whereas an AIL is interpreted by a set of (extensional) structures, an intensional AIL (IAIL) is interpreted by a set of intensional structures in the sense of 'intensional structure' defined at the outset of this chapter. Thus an IAIL is an ordered pair $\langle \underline{L}, \$ \rangle$ where $\underline{L}$ is a first order language and \$ is a set of intensional interpretations for $\underline{L}$. An IAIT is an ordered pair $\langle \underline{T}, \langle \underline{L}, \$ \rangle \rangle$ where $\underline{T}$ is a first order theory and $\langle \underline{L}, \$ \rangle$ is an IAIL. A sentence of $\langle \underline{L}, \$ \rangle$ is true at $\underline{H}$ if it is true in $\underline{I}(\underline{H})$ for every $\underline{I}$ in \$. Then, for example, a IAIL is committed de re to Nixon if Nixon exists at every world at which the theory is true. The definition of an IMAIL will be given in Chapter 4.

Finally, in later chapters we will be interested in asking <u>how</u> <u>many</u> things a theory is committed to existing. We give a criterion for this as follows: $\underline{T}$ is committed to there being α-many things off in every world at which $\underline{T}$ is true at least α-many things exist.

CHAPTER III

We now consider the consequences of interpreting mathematical theories as certain kinds of IAIT's. As noted in Chapter One, in so doing we will be investigating the consequences of interpreting mathematical theories in the way that Benacerraf and Field suggest. The motivating intuition behind the approach to be studied here and in Chapter Four is that a structure is acceptable for interpreting a mathematical theory if it has the correct "structural" properties. A mathematical theory is not the theory of one such structure in particular. It will be the business of Chapters Five and Six to clarify and attempt to support this intuition. Unless a formal theory of mathematical truth could be constructed which embodied this intuition, however, this intuition would be of little value. It is to such formal matters that we now turn.

(A) Consider a first order language $\underline{L}$ for arithmetic whose sole predicate is the two-place predicate '$\underline{S}$' (the successor relation). We provide this language with an interpretation according with the intuition noted above. In other words, we want to define a set \$ of acceptable intensional interpretations such that the IAIL $\langle \underline{L}, \$ \rangle$ captures the intuition in question. To do this we give necessary and sufficient conditions for an intensional interpretation of be a member of \$.

An acceptable structure for arithmetic is, roughly, an omega-progression. An acceptable intensional interpretation $\underline{I}$ for $\underline{L}$ is then one in which for any world $\underline{H}$ (1) if there are only a finite number of things which exist at $\underline{H}$, $\underline{I}(\underline{H})=\langle\Psi(H);\phi\rangle$ (2) if there are infinitely many things which exist at $\underline{H}$, $\underline{I}(\underline{H})$ is isomorphic to the "standard model" of $\underline{L}$.

The effect of (1) is that if there are not enough things at $\underline{H}$ to form the domain of an extensional interpretation of $\underline{L}$ which has the correct structural properties, then '$\underline{S}$' is given the null interpretation. Given such an interpretation, '$\underline{Sxy}$' is false for every assignment of members to the domain of '$\underline{x}$' and '$\underline{y}$'.

From (1) and (2) it follows that if $\underline{I}\in\$$, $I(H)=\langle\Psi(H);\phi\rangle$ if there are only finitely many things at $\underline{H}$. Roughly speaking, the idea is that an acceptable intensional interpretation picks out an omega-progression at $\underline{H}$ if there are enough things at $\underline{H}$.

The "standard model" of $\underline{L}$ is simply the model whose domain is the set of natural numbers and whose assignment to '$\underline{S}$' is the successor relation defined on the natural numbers.

Having now specified \$, we say that a sentence of $\langle\underline{L}, \$\rangle$ is true at a world $\underline{H}$ if it is true in every $\underline{I}(\underline{H})$ for $\underline{I}$ a member of \$. Conditions (1) and (2) together guarantee that for every $\underline{H}$, for every $\underline{I}$, $\underline{I}'$, $\underline{I}(\underline{H})$ is isomorphic to $\underline{I}'(\underline{H})$ so that if a sentence of true in some $\underline{I}(\underline{H})$ it is true in any I'(H). Thus, defining falsehood at H as falsehood in all the I(H)'s, every sentence will be either true or false at every world. (Of course some sentences could be true at one world and false at another.) There are no indeterminate sentences.

(B) This technique may also be applied to the case of set theory. Consider a formal language L whose sole predicate is the two-place predicate 'ε'. We aim to provide this language with a set $ of acceptable intensional interpretations such that the resulting IAIL captures the intuition that set theory is not the study of one particular structure, but of what all structures which meet certain conditions have in common.

Two difficulties confront the attempt to define $. First of all, for each I in $ and H, I(H) will be some set. Hence the domain of I(H) will be a set. But (in ZF) there is no set of all sets and thus it seems that the domain of I(H) cannot contain all the sets. Thus it seems that it is impossible to give an interpretation of L in this way which captures the intention of set theory to speak of all sets.

This is a quite general difficulty which confronts the attempt to provide a formal semantics for set theory. It presents no special problem for our approach. What it indicates is that formally we may not be able to give the "intended interpretation" of set theory, but only a model which reflects that intended interpretation. We proceed with that qualification on what we do.

The second difficulty concerns possible structural indeterminacy in set theory. In the case of arithmetic, by contrast, it is generally thought that we have a clear idea of the notion of a standard model of arithmetic. Such a model must be isomorphic to the vonNeumann model of arithmetic.

Given differing conceptions of sets and possible unclarities within a given conception of sets, it is not at all clear that any two acceptable extensional interpretations must be isomorphic. In one such interpretation the GCH might be true, while in another it might be false. One such interpretation might even contain another as a set in its domain.

Once again, this is a general problem which presents no special difficulty for the approach taken here. In specifying $\$$ we will in effect take a position on the question of structural indeterminacy, but other positions on this question can be accommodated within this approach. We will consider two such positions.

First, we take the radical position that there is no indeterminacy in set theory. Let $\underline{C} = \{\underline{x} \mid \text{rank } x < \text{the first inaccessible cardinal } \underline{K}'\}; \{\langle \underline{x}, \underline{y} \rangle \mid \text{rank } \underline{x}, \text{ rank } \underline{y} < \underline{K}' \ \& \ \underline{x} \varepsilon \underline{y}\}$. $\underline{C}$ thus is the model of set theory which consists of the sets of rank less than the first inaccessible cardinal. Then we say that an intensional interpretation I of L is a member of \$ if for any H (1) if there are fewer than R(K') things which exist at H, I(H)=⟨Ψ(H);∅⟩; (2) if there are not fewer than R(K') things which exist at H, I(H) is isomorphic to C.

Given this definition of \$ and our customary definitions of 'true' and 'false', every sentence of L will be true or false in every H. (As before, however, a sentence could be true at one world and false at another.)

A second, more liberal, definition of \$ allows I to be a member of \$ if for any H (1) if there are only finitely many things which exist at H, I(H)=⟨Ψ(H);∅⟩; (2) if there are infinitely many things which exist at H, then I(H) is a model of ZF. On this definition, the GCH, for example, would be neither true nor false at any world at which there are infinitely many things. In this case truth coincides with provability.

(c) An interpretation can also be provided for a combined arithmetic and set theory. The language in question has the predicates 'S' ('is a set'), 'ε', 'S^2' ('is the successor of'), and 'N' ('is a number'). An intensional interpretation I is a member of \$ if (the following condition employs the second method for interpreting set theory just noted) for any H (1) if there are only finitely many things which exist at H, then I(H)=⟨Ψ(H);∅,∅,∅,∅⟩; (2) if there are infinitely many things which exist at $\underline{H}$, then $\underline{I}(\underline{H})=\langle\Psi(\underline{H});\ \underline{A};\ \underline{B};\ \underline{C};\ \underline{D}\rangle$ where $\langle\underline{A},\ \underline{B}\rangle$ is isomorphic to the standard model of arithmetic and where $\langle\underline{C},\ \underline{D}\rangle$ is a model of ZF.

Many sentences are neither true nor false relative to this interpretation (at worlds at which there are infinitely many things). First of all, the GCH and, more generally, any sentence which is also a sentence of ZF and which is independent in ZF will be neither true nor false (at any such world). More interestingly, many sentences which "mix" arithmetical and set theoretical language will neither true nor false. '0=ø' abbreviates a sentence which is neither true nor false. The reason for this is that, to speak for the moment as though '0' and 'ø' are in the language $\underline{L}$, '0' and 'ø' get assigned the same thing by some $\underline{I}$ and \$ and different things by other $\underline{I}$ and \$. The truth value of such sentences as '0=ø' will be the major topic of discussion in Chapter Five and it will be seen that the fact that on the interpretation provided here '0=ø' is neither true nor false (once again, at worlds at which there are infinitely things) may be an important consideration in favor of viewing this interpretation as the correct interpretation of arithmetic plus set theory.

(D) This approach may also be used to interpret the languages appropriate for stating theories of impure sets. Such a language might contain, for example, the predicates

'ε', '$\underline{S}$' ('is a set') and '$\underline{A}$' (is an apple'). We say that an intensional interpretation $\underline{I}$ of this language is a member of \$ if for any H (1) if there are only finitely many things which are non-apples and which exist at $\underline{H}$, $\underline{I}(\underline{H}) = \langle \Psi(\underline{H}); \emptyset; \emptyset;$ the set of things which are apples at $\underline{H}\rangle$, (2) if there are infinitely many non-apples which exist at $\underline{H}$, then $\underline{I}(\underline{H}) = \langle \Psi(\underline{H}),; \underline{A}; \underline{B};$ the set of things which are apples at $\underline{H}\rangle$ where $\underline{I}(\underline{H})$ is a model of ZFA. ZFA is the theory obtained for ZF by adding the axiom $(\underline{Ex})(\underline{Sx}\ \&\ (\underline{y})(\underline{y}\varepsilon\underline{x} \leftrightarrow \underline{Ay}))$ and rewriting the axioms of ZF by replacing each occurrence of a universal quantifier '...$(\underline{x})$...' by '...$(\underline{x})(\underline{Sx} \rightarrow$...)'. (We assume existential quantifiers have been eliminated.) This theory allows some (even all) apples to be sets, but not all sets can be apples; hence we have the requirement that there are infinitely many non-apples.

(E) (D) also illustrates that this approach has no difficulty dealing with "mixed theories". In terms of the notion of partial signification, we can say that a theory is mixed if some of its predicates partially signify more than one extension and others partially signify more than one extension. Intuitively, a mixed theory is one which contains some mathematical and some non-mathematical vocabulary. In (D) 'ε' partially signifies more than one extension (at a world), while '$\underline{A}$' partially signifies exactly one extension (at a world).

In general, to interpret mixed theories, if a predicate '*P*' does not, according to our philosophical theory, partially signify more than one extension (in the informal theory in which it is used), then for every *I*, *I*' in $, *I*(*H*) assigns the same extension to the predicate which represents '*P*' as does *I*'(*H*). The assignments to other predicates may vary form *I* to *I*'.

It is significant that this approach handles mixed theories so easily. In giving a theory of mathematical truth, we must account not only for the truth of pure mathematical theories, but also applied mathematical theories, e.g. theories such as physics. Such mixed theories create important problems for the approach to be considered in the next chapter. Furthermore it is crucial that such theories be interpreted.

The general approach towards the problem of interpreting mathematical theories to be considered in this chapter should now be clear. It is an important question to ask whether the interpretations we have given these formal languages should be regarded as giving an account of the associated informal languages. In giving an interpretation for the language whose sole predicate is '*S*', have we given an account of the informal language whose sole predicate is "is the successor of'?

One way of beginning to answer this question is to ask whether this interpretation gives the correct truth values to sentences. The answer to this, at least in the case of arithmetic, the case of set theory being less clear given the possible problem of structural indeterminacy is, Yes.

On the other hand, there are many other ways of interpreting mathematics (e.g. interpretations in set theory) which also apparently give correct truth values. The question which then arises is: why prefer one method of interpreting theories to another (provided these methods agree on truth values)? Does it even make sense to speak of one, rather than another, being the correct account? What is there to choose between such accounts? In large part, such questions are the business of Chapters Five and Six. However, here we may note that although two methods may agree on the distribution of truth values, but differ in the ontological commitments they imply a theory has.

Thus we now turn to consider what the ontological commitments of mathematical theories are relative to interpreting them as IAIT's as discussed in this chapter.

To the IAIL defined in (A) above we add the following axioms to obtain an IAIT.

(1) (E!x)(y) - Sxy

(2) (x)(E!y) Syx

(3) (x)(y)(z)(Sxy & Sxz → y=z)

(1)-(3) together guarantee that there are infinitely many "numbers". We can now discuss the ontological commitments of this theory.

I claim that (i) this theory T has no de re commitments and (ii) T is not committed to the kind abstract entity. To take (i) first (which is, I think, intuitively, clearly true) is there some thing which exists at every word at which T is true? This amounts to the question whether there is some thing, such that in any world at which it does not exist only finitely many things exist. (The existence of infinitely many things is sufficient for there to be acceptable interpretations of this theory and T is true in all such interpretations.) It is clear that there is no such thing given the notion of a world defined in Chapter Two.

Now consider whether T is committed de re to something x. (Again, intuitively, there is no such thing.) We have just said that there is a world H at which there are infinitely many things and x does not exist. T is true in each such world. Thus by the criterion developed in Chapter Two, T has no de re commitments. The reader should note here that we are relying in a substantive way in this argument on the notion of a world which has been defined. If it is thought that in some sense the notion of a world we defined is inappropriate for evaluating commitments, this result would not follow. Note, however, that intuitively _T_ has no de re commitments. Thus any correct criterion of de re commitment should imply this. We may also view this as confirmation of the theory of commitment stated in Chapter Two.

(ii) There is a world at which there are no abstract entities and there are infinitely many concrete entities. Let _H_ be such a world. In _H_ each _I_ in $ picks out a structure isomorphic to the standard model of arithmetic (the domains of which structures consist solely of concrete entities). _T_ is true in each such structure; hence it is true in a world at which no abstract entities exist. Hence, by our criterion of commitment for IAIT's, _T_ is not committed to the kind _abstract entity_.

We could say that a theory in _nominalistic_ if (1) it is not committed de re to any entity which is abstract and (2) it is not committed to the kind _abstract entity_. (Recall that on our theory of commitment (2) does not imply (1) although we did consider a modification of our theory which would have had that result if we assume that any abstract entity is essentially abstract.) A program of nominalism in mathematics would then be successful if every mathematical theory could be interpreted so as to satisfy (1) and (2). Nominalism need not be understood in this way (in fact in Chapters Five and Six platonism (non-nominalism) will be understood in a different way as the claim that mathematical constants refer to abstract entities) but I think this characterization does capture at least part of what nominalism in mathematics has been taken to mean.

In this sense I claim that both arithmetic and set theory, as here interpreted, are nominalistic. This then raises the interesting possibility that if informal arithmetic and informal set theory (in some sense) should be interpreted in this way, that both of these theories are nominalistic.

It is relatively clear how these results are obtained. Roughly speaking, we allow anything at all to be the 'ϕ' of an acceptable structure for set theory. In other words, we interpret 'ϕ' as partially denoting everything. In such a way there is no one thing or kind of thing which must exist in order for 'ϕ exists' to be true. Something is mildly puzzling here, however, since in interpreting arithmetic and set theory as we have we have used both set theory and arithmetic in our metalanguage. A suspicion may arise here that this shows that a general program of nominalism in mathematics cannot be carried through in this way for the reason that our meta-theory has non-nominalistic commitments. This issue arises throughout this dissertation, for example, in our definition of a world in Chapter Two, but perhaps most obviously in (A) above where we "referred to" the standard model of arithmetic in specifying the interpretation of a language. This same issue arises in Chapter Four, where it is argued that this suspicion is incorrect.

A second important objection to the approach taken here is as follows. Since informal arithmetic has 'one exists' as a theorem, informal arithmetic has a de re commitment to the number one. Thus either our formal interpretation of informal arithmetic is inadequate or we are incorrect in claiming that this formal theory lacks de re commitments. Similar remarks could be made concerning commitment to the kind <u>number</u>.

This is the same objection which was discussed in the opening pages of Chapter One. Our answer to it is that it is not a priori untenable to hold that in mathematics 'there are' and 'exists' are used in special ways: ways such that 'one exists' and 'there are numbers' do not have a de re commitment and a commitment to the kind <u>number</u>, respectively.

A final, related, question concerns the applicability of this technique of interpretation. Obviously any informal theory could be represented as an IAIT in such a way that the resultant formal theory lacks de re commitments. This is disturbing in that it seems to open the door to holding that no theory has de re commitments, which is, perhaps, an absurd result. Much the same sort of thing bothered Quine, I believe, in his worry that the Lowenheim-Skolem theorem showed that every theory could be reduced to a theory of numbers.

In fact, interpreting theories as IAIT's is a general technique. Any informal theory can be represented as an IAIT. The question is, however, whether they should be so represented. Presumably some theories do have de re commitments. Such theories should not be represented by formal theories which lack such commitments. Thus what is shown in this chpater in no way indicates that arithmetic lacks de re commitments. That could only be established by philosophical argument. The semantical technique introduced here only shows how mathematics might be represented.

On the other hand, although these theories have no de re commitments and are nominalistic, in another way they do have quite heavy ontological commitments. For example, the theory T of arithmetic has a commitment to there being infinitely many things, in the sense that there are infinitely many things in any world at which it is true. Set theory, of course, may have an even heavier commitment, depending on how it is interpreted. In particular, on our first interpretation given above, it has a commitment to there being inaccessibly many things.

But even though these theories have these commitments, they are nominalistic. Can a general program of nominalism in mathematics be thus carried through by interpreting mathematical theories as IAIT's? We have already noted one suspicion that it cannot, but in fact there is an easier way of showing this program will not work.

If we consider just pure mathematical theories, I think it is clear that this program can be carried through. Difficulties arise, however, in certain cases of mixed theories. Consider the theory *T'* whose non-logical axioms are the axioms of the theory *T* of arithmetic noted above and the additional axiom (*E!x*)*Cx* (there is exactly one concrete thing). The interpretation of the language of this theory is such that *I* is in $ if for any *H* (1) if there are only finitely many things which exist at *H*, then $I(H)=\langle\Psi(H);\ \emptyset;$ the set of things which are concrete at $H\rangle$ and (2) if there are infinitely many things which exist at *H*, then $I(H)=\langle\Psi(H);\ A;$ the set of things which are concrete at $H\rangle$ where *A* is isomorphic to the standard successor relation defined on the natural numbers.

In any world *H* at which *T'* is true, there is only one thing which is concrete at *H*, but there are infinitely many things which exist at *H*. Thus *T'* is committed to the kind abstract _entity_. (Note that we are here relying on one of the requirements which we placed on worlds in Chapter Two: if something is not concrete, it is abstract.) _T_' still does not have any de re commitments to abstract entities, but it is not nominalistic (though it is nominalistic in the sense of Chapters Five and Six in that it does not construe mathematical constants as names) and this commitment is not due alone to the non-mathematical, i.e. 'there is exactly one concrete thing', part of T'.

Now this example is unrealistic in that _T_' is not a theory which anyone is likely to hold. More realistic examples are easy to find, however, If set theory is interpreted in the first way described above, then that theory plus the axiom 'there are only denumerably many concrete things', (represented in the obvious way as an IAIT) has a commitment to the kind _abstract_ _entity_.

An even simpler case is the theory ZFA if we interpret '_A_' as having as its extension at a world the set of concrete things at that world. ZFA then "says" there is a set which contains all the concrete things and in any world in which ZFA is true there are abstract things (since otherwise the domain of a model of ZFA would be a member of itself). Thus an impure set theory which asserts the existence of a set containing all concrete entities will be committed to the kind _abstract_ _entity_. Thus not every mathematical theory can be shown to be nominalistic by interpreting it as an IAIT.

This point has relevance to a claim of Field's Field claims ([7]: 220-1) that we can account for arithmetical truth while holding that there are infinitely many physical objects but no abstract objects by interpreting arithmetic in the way we have presented in this chapter. We have seen that that claim is true, We have also seen, however, that we cannot account for mathematical truth in general (counting mixed theories as mathematical theories) while holding this view. Thus the point about arithmetic is only of limited interest.

Let us summarize the results of this chapter. Mathematical theories, interpreted as IAIT's:

(1) can lack de re commitments to abstract entities (even when they are not nominalistic);

(2) can allow such problematic sentences as '$0=\emptyset$' to be neither true nor false;

(3) have commitments to there being K things, for some cardinal K;

(4) are not, in general, nominalistic.

CHAPTER IV

In this chapter we investigate the consequences of interpreting mathematical theories as IMAIT's. The primary consequences for pure mathematical theories, so interpreted, are that if such theories allow the formation of sentences which combine mathematical theories, e.g. '0=∅', then such sentences can be neither true nor false and that no such theory has an ontological commitment to there being at least one thing. In other words, it is shown how such theories can be true in the empty world. (Note that given the definition of a world given in Chapter Two there is an empty world.) Two obvious corollaries of this second consequence are that pure mathematical theories (so interpreted) are nominalistic (in the sense defined in Chapter Three) and that such theories have no de re commitments. This result should be of interest given widespread scepticism concerning whether a nominalistic account of mathematics is possible.

Most of the work of this chapter is concerned with extending this treatment of pure theories to mixed theories. Difficulties are encountered, but I believe a solution is found. The results for mixed theories could be summarized by saying that unless the non-mathematical part of a mixed theory is non-nominalistic, the entire thoery (interpreted as an IMAIT) is nominalistic. More generally, the only commitments of such a mixed theory are those of the non-mathematical part of that theory. Interpreting mathematical theories as IMAIT's thus results in importantly fewer ontological commitments than interpreting them as IAIT's.

Consider the theory T of arithmetic defined in Chapter 3. The axioms and interpretation \$ remain as before. All that we change to interpret this theory as an IMAIT are the definitions of truth and falsehood.

Instead of saying that a sentence is true at a world H if it is true in every I(H) for I in \$, we now say that it is true at H if for every world H' and I in \$, if $I(H') \neq \langle \Psi(H), \phi \rangle$ then H is true in I(H'). A sentence is false at H if for every world H' and I in \$ if $I(H') \neq \langle \Psi(H), \phi \rangle$ then it is false in I(H').

Some explanation would be helpful here. The idea is that a sentence is true just in case necessarily it is true in every structure in which 'S' has an extension with the "right" structure. Given the way we have defined \$, for every I in \$ and world H I(H) either gives 'S' the null extension or gives it an extension with the right structure, that is, it assigns it an extension isomorphic to the successor relation of the standard model of arithmetic. By looking only at I(H)'s where 'S' does not have the null extension, we guarantee that we consider only the right structures.

The reason why arithmetic, interpreted as an IMAIT is not committed to there being at least one thing is clear. Let $\underline{H}$ be the world at which nothing at all exists. The theory of arithmetic just defined is true at $\underline{H}$, because its axioms are true in every $\underline{I}(\underline{H}')$ which is not identical with $\langle \Psi(H'); \phi \rangle$. Thus this theory is true at a world at which nothing at all exists. This implies that this theory is not committed to there being at least one thing.

These results are straightforwardly applicable to both set theory and the combined set theory-arithmetic discussed above. In each case the interpretation $ remains as before and only the definitions of truth and falsehood are changed.

Clearly, neither of these theories is committed to there being at least one thing. Furthermore, for the combined set theory-arithmetic in which such sentences as '0=ϕ' can be formed, such sentences can be neither true nor false. ('0=ϕ' can also be interpreted to be true or false, also.) This method of interpreting mathematical theories as IMAIT's can be applied to any mathematical theory, so that we have now achieved what was promised in opening: interpreting pure mathematical theories as IMAIT's has the counsequences that certain sentences which combine mathematical languages can be taken to be neither true nor false and that these theories so interpreted are not ontologically committed to there being at least one thing. As we have noted, this latter consequence implies that such theories have no de re ontological commitments and that such theories are nominalistic.

Again, it could be argued, as it was in Chapters One and Three that since 'there are numbers' has a commitment to the kind number, either our representation of arithmetic is inadequate or we are incorrect in holding that this formal theory lacks a commitment to the kind number. I believe, however, that this objection was answered in Chapter One.

This modal interpretation of mathematics thus has some interesting consequences. However, if we attempt to interpret theories which mix mathematical language with language about ordinary concrete entities, difficulties arise. This difficulty concerning mixed theories might be expected. In giving a modal interpretation of mathematics, we essentially look at all worlds to see whether a sentence is true at a world. That does not create any difficulties, because a true mathematical theory is necessarily true. However, if we consider the truth value of a sentence about concrete entities we definitely do not want to look at all worlds to see if that sentence is true at a world H, for, roughly, it is facts which are peculiar to H which make this sentence true.

To see this in a particular case, let us add to our language of arithmetic the one-place predicates '$\underline{F}$' and '$\underline{N}$' (to be interpreted as meaning: is a fish and is a number, respectively.) To give a modal interpretation of the resultant language we specify \$ and give a definition of truth. An intensional interpretation $\underline{I}$ is a member of \$ if for any $\underline{H}$, (1) if there are only finitely many members of $\Psi(\underline{H})$ $\underline{I}(\underline{H})=\langle\Psi(\underline{H});\emptyset;\emptyset;$ the set of things which are fish are $\underline{H}\rangle$ and (2) if there are infinitely many members of $\Psi(\underline{H})$, $\underline{I}(\underline{H})=\langle\Psi(\underline{H});$ $\underline{A}$; $\underline{B}$; the set of things which are fish at $\underline{H}\rangle$ where $\langle\underline{A}, \underline{B}\rangle$ is isomorphic to the standard model of arithmetic.

If we were to follow the general pattern of the definition of truth for an IMAIT presented above, we would say that a sentence of this theory is true at a world $\underline{H}$ if for every $\underline{H}'$ and $\underline{I}$ in \$, if $\underline{I}(\underline{H}')\neq\langle\Psi(\underline{H});\emptyset;\emptyset;$ the set of things which are fish at $\underline{H}\rangle$ then that sentence is true in $\underline{I}(\underline{H}')$.

This, however, will not work. Suppose that at $\underline{H}$, something is a fish. Then we want '($\underline{Ex}$)$\underline{Fx}$' to be true at $\underline{H}$ on this interpretation. Unfortunately, for some $\underline{I}(\underline{H}')$ in which '$\underline{S}$' does not get assigned $\emptyset$, '$\underline{F}$' does get assigned $\emptyset$. In such an $\underline{I}(\underline{H}')$, '($\underline{Ex}$)$\underline{Fx}$' is false. Thus, on the proposed definition '($\underline{Ex}$)$\underline{Fx}$' is not true at $\underline{H}$.

To solve this problem, let us think about what sort of structure we would want to examine in defining truth of a sentence of this theory. Suppose that $\underline{I}(\underline{H})=\langle\Psi(\underline{H});$ $\underline{C}$; $\underline{D}$;

the set of things which are fish at $\underline{H}\rangle$. ($\langle\underline{C}, \underline{D}\rangle$ might either be $\langle\phi, \phi\rangle$ or isomorphic to the standard model of arithmetic.) Then it appears that the sort of structure we seek would be of the form: $\langle\underline{D}, \underline{A}, \underline{B}$, the set of things which are fish at $\underline{H}\rangle$ where $\Psi(\underline{H}) \subseteq \underline{D}$ and $\langle\underline{A}, \underline{B}\rangle$ is isomorphic to the standard model of arithmetic. Let us call such a structure an acceptable structure relative to $\underline{H}$, $\underline{I}$. Then the definition of truth at a world which we seek would be: a sentence is true at $\underline{H}$ if for every $\underline{I}$, $\underline{I}'$ and $\underline{H}'$ if $\underline{I}'(\underline{H}')$ is an acceptable structure relative to $\underline{I}$, $\underline{H}$, then that sentence is true in $\underline{I}'(\underline{H}')$.

The general idea here is to keep the extensions of some predicates fixed. Let us call such predicates _world-bound_. The extensions of other predicates (essentially, the "mathematical" ones) are allowed to vary. Such predicates we will call _other-worldly_. Then we also (in some cases) need to consider a larger domain.

Now suppose that $\underline{I}(\underline{H})=\langle\{\underline{a}, \underline{b}\}, \phi, \phi, \{\underline{a}\}\rangle$. $\underline{H}$, in other words, is a world at which two things exist, one of which is a fish. It is now easy to see that $(\underline{E}!\underline{x})\underline{Fx}$ is true at $\underline{H}$, since in any acceptable structure relative to $\underline{H}$ and $\underline{I}$, only $\underline{a}$ is in the extension of '$\underline{F}$'.

Note, however, that $(\underline{E}!\underline{x})$-$\underline{Fx}$ is false, since in any such acceptable structure there are infinitely many things not in the extension of '$\underline{F}$'. This is neither surprising, nor disturbing, since in order to make the mathematical sentences have the correct truth values at _H_, we need to have the sentence which says that there are more than two things come out true.

This fact does, however, point out a certain lack of expressive power of this language. Somehow we would like to be able to express 'there is a unique non-fish' by a true sentence of this language, since, in at least one sense, this sentence is clearly true at _H_. In fact, this is easily done by adding a new world-bound predicate '_E_' to this language. ('_E_' is supposed to suggest _existence_ for reasons that will soon be made clear.) In every _I_(_H_), the extension of '_E_' is equal to the domain of _I_(_H_). In this way it is like an existence-predicate. Since '_E_' is world-bound, however, it keeps the same extension when we move from _I_(_H_) to a structure acceptable relative to _H_, _I_.

Then it is easy to see that (_E!x_)(_Ex_ & - _Fx_) is true at the two-thing world described above. In general, if we want to speak only of what really exists at _H_ we relativize the quantifiers to '_E_'.

'_E_' then adds important expressive powers to this language. In general, it may be useful to add such a world-bound existence predicate to the language used to formalize mixed informal languages.

Note also that the sentences with only arithmetical (other-worldly) predicates get the correct truth values.

In any acceptable structure, the assignment to 'N' and 'S' guarantees that, since that assignment ⟨A, B⟩ is isomorphic to the standard model of arithmetic.

As for sentences which mix world-bound and other-worldly predicates, consider (*) (x)(Fx v Nx). As we have arranged things, (*) is neither true nor false at H. The reason for this is that in some acceptable structures the extension of 'N' will be the domain of the structure, while in other acceptable structures there are things which are neither in the extension of 'F' nor in the extension of 'N'. (*) is a natural formalization of 'everything is a fish or a number'. As such, it seems not unreasonable that (*) is neither true nor false, but note that by slight and obvious modifications of the definition of an acceptable structure different results could be obtained. For example, by requiring the domain of an acceptable structure relative to the world H above be {a, b} the extension of 'N', (*) could be true at H. Note also that on both of these definitions both (x)(Ex → Nx) and (x)(Fx → Nx) are neither true nor false. Once again, however, different results are easily obtained. This is desirable, since the truth values of sentences which mix mathematical with non-mathematical sentences are often difficult to ascertain. In this context, I am more interested in exhibiting a general technique which can be accommodated to any philosophical stand on these questions, rather than actually taking a stand on these questions.

This distinction we have made between world bound and other-worldly predicates, though useful as far as it goes, is not, however, adequate for dealing with all mixed languages. Consider 'has at least ____ brothers'. This is a predicate which, intuitively, is non-pure at its first place and pure at its second place. Let us call such predicates _mixed_ and adopt the convention of speaking of the places of mixed predicates as being either _pure_ or _non-pure_. In giving a definition of a mixed theory we will want to make certain requirements on how the mixed predicates are interpreted, just as we put conditions on how the world-bound predicates are interpreted.

Suppose for the sake of illustration that we are dealing with a language having predicates '$\underline{N}$', '$\underline{S}$', and '$\underline{B}$' to be interpreted as meaning _is a number_, _is a successor of_ and _has at least ___ brothers_, respectively. Suppose that $\underline{I}(\underline{H})=\langle\Psi(\underline{H});\ \underline{N};\ \underline{S};\ \underline{B}\rangle$ where $\underline{N}$ is the extension of '$\underline{N}$', etc. What sort of a structure do we want to consider in defining truth at $\underline{H}$ of a sentence of this language? Such a structure will be of the form $\langle\underline{D},\ \underline{N}';\ \underline{S}';\ \underline{B}'\rangle$ where $\Psi(\underline{H})\subseteq\underline{D}$ and $\langle\underline{N}';\ \underline{S}'\rangle$ is isomorphic to the standard model of arithmetic. What requirements do we want to put on $\underline{B}'$? The following appear to be minimal requirements.

First, we require that if $\langle\underline{x},\ \underline{y}\rangle\ \varepsilon\ \underline{B}'$, then $\underline{x}\varepsilon\Psi(\underline{H})$.

This is intended to capture the idea that '*B*' is non-pure at its first place.

Also, note that '*B*(*x*, 1)' abbreviates certain open sentences of this language. Roughly speaking, we would like to guarantee that if *x* satisfies '*B*(*x*, 1)' in one acceptable structure relative to *H*, *I*, then it satisfies '*B*(*x*, 1)' in every acceptable structure relative to *H*, *I*.

We implement this requirement as follows. Let us say that '*Py*' *is an M-sentence relative to a set Q of structures if* '*y*' is its only free variable, it contains only other-worldly predicates (in this case, '=', '*N*' and '*S*') and '(*E*!*y*)*Py*' holds in every member of *Q*. Then we require: if in some acceptable structure *S* relative to *H*, *I*, there is an *M*-sentence '*Py*' relative to the set of acceptable structures relative to *H*, *I*, such that $\langle x, y \rangle$ is in the extension of '*B*' in *S* and *y* satisfies '*Py*' in *S*, then there is an *M*-sentence '*P*y*' relative to the set of acceptable structures relative to *H*, *I*, such that y satisfies '*P*y*' in *S* and every acceptable structure *S*' relative to *H*, *I* there is a *z* satisfying '*P*y*' in *S*' and such that $\langle x, z \rangle$ satisfies '*B*' in *S*'.

The idea here is that if $\langle x, y \rangle$ satisfies '*B*' in some acceptable structure and *y* satisfies an *M*-predicate, then there is some *M*-predicate which picks out correlates of *y* in every acceptable structure including *y* itself in the given structure.

Let us move from these examples to a general definition of a mixed IMAIL. (In fact, pure IMAIL's can be understood essentially as special cases of mixed IMAIL's, so this is a perfectly general definition.) A mixed IMAIL is an ordered triple $\langle \underline{L}, \$, \underline{g} \rangle$ where $\underline{L}$ is a first order language and \$ is a set of intensional interpretations of $\underline{L}$. We assume an exhaustive and mutually exclusive division of predicates into the categories world-bound, mixed and otherworldly. Furthermore, we assume a specification of which places of the mixed predicates (if any) are pure and which are non-pure.

$\underline{g}$ then is a three-place function which takes an intensional interpretation $\underline{I}$, a world $\underline{H}$ and a world $\underline{H}'$ into the set of structures whose domains are a subset of $\Psi(H')$ and which are acceptable relative to $\underline{H}$, $\underline{I}$.

In general, we require $\underline{g}$ to satisfy the following conditions. If $\underline{S}$ is a member of the domain of $\underline{G}(\underline{I}, \underline{H}, \underline{H}')$ then,

(i) the domain of $\underline{I}(\underline{H})$ is a subset of the domain of $\underline{S}$;

(ii) the extension of any world-bound predicate is the same in $\underline{I}(\underline{H})$ as it is in $\underline{S}$;

(iii) for any mixed predicate '$\underline{P}^n$' and member $\underline{S}$ of $g(I, H, H')$, if $\langle a_1, \ldots, \underline{a}_n \rangle$ is a member of the extension of '$\underline{P}^n$' in $\underline{S}$, then
(a) if $\underline{i}$ a non-pure place of '$\underline{P}^n$, then $\underline{a}_i$ is

a member of the domain of $\underline{I}(\underline{H})$;

(b) If for each $\underline{i}$ which is a pure place of '$\underline{P}^n$', $\underline{a}_i$ satisfies a predicate '$\underline{P}_i$' which is an $\underline{M}$-predicate relative to $\bigcup^{I\varepsilon\$}_{H'\varepsilon K}$ $\underline{g}(\underline{I}, \underline{H}, \underline{H}')$, then there are predicates '$\underline{P}^*_i$' relative to $\bigcup^{I\varepsilon\$}_{H'\varepsilon K}$ $g(I, H, H')$ such that $\underline{a}_i$ satisfies '$\underline{P}^*_i$' in $\underline{S}$ and for any $\underline{S}'$ in $\bigcup^{I\varepsilon\$}_{H'\varepsilon K}$ $\underline{g}(\underline{I}, \underline{H}, \underline{H}')$, there is an n-tuple $\langle \underline{b}_1, \ldots, \underline{b}_n \rangle$ which satisfies '$\underline{P}n$', in $\underline{S}'$ such that if $\underline{i}$ is a non-pure place in '$\underline{P}^{n}$', then $\underline{a}_i = \underline{b}_i$ and if $\underline{i}$ is a pure place, then $\underline{b}_i$ satisfies '$\underline{P}^*_i$' in $\underline{S}'$.

A sentence then is true at $\underline{H}$ if it is true in every member of $\bigcup^{I\varepsilon\$}_{H'\varepsilon K}$ $\underline{g}(\underline{I}, \underline{H}, \underline{H}')$. That is, it is true at $\underline{H}$ if it is true in every acceptable structure relative to $\underline{H}$. A sentence is false if it is false in every member of $\bigcup^{I\varepsilon\$}_{H'\varepsilon K}$ $\underline{g}(\underline{I}, \underline{H}, \underline{H}')$.

To test this definition, we interpret an impure set theory as a mixed IMAIT. To obtain such a theory we relativize each of the axioms of ZF to a predicate '$\underline{S}$' ('is a set') and then add axioms asserting the existence of impure sets. In this case we add the sole axiom $(\underline{Ex})(\underline{Sx} \; \& \; (\underline{y})(\underline{y}\varepsilon x \leftrightarrow \underline{Cx}))$. (The set of concrete things exists.)

To interpret this theory as an IMAIT we specify the set \$ of interpreting intensional interpretations. $\underline{I}$ is in $\underline{\$}$ if for every world $\underline{H}$ (1) if there are only finitely many non-concrete things which exist at $\underline{H}$, then $\underline{I}(\underline{H})=\langle\Psi(\underline{H});\ \phi;\ \phi;$ the set of things which are concrete at $\underline{H}\rangle$; (2) if there are infinitely many non-concrete things which exist at $\underline{H}$, then $\underline{I}(\underline{H})=\langle\Psi(\underline{H});\ \underline{A};\ \underline{B};$ the set of things which are concrete at $\underline{H}\rangle$ where $\underline{A}$ is the extension of '$\underline{S}$' and where $\underline{I}(\underline{H})$ is a model of the impure set theory just stated.

We then treat '$\underline{C}$' as a world-bound predicate and 'ε' and '$\underline{S}$' as other-worldly in order to specify $\underline{g}$. In specifying $\underline{g}$ we will be saying what an acceptable structure relative to $\underline{H}$, $\underline{I}$ is. In general, if $\underline{I}(\underline{H})=\langle\underline{x};\ \underline{y};\ \underline{z};\ \underline{t}\rangle$ we say that $\underline{g}(\underline{I},\ \underline{H},\ \underline{H}')$ is the set of $\underline{S}=\underline{x}';\underline{y}';\ \underline{z}';\ \underline{t}'$ such that $\underline{S}$ is a model of this theory, $x\subseteq x'\subseteq\Psi(H')$ and $t=t'$.

For example, if $\underline{I}(\underline{H})=\langle\{\underline{a},\ \underline{b}\};\ \phi;\ \phi;\ \{\underline{a}\}\rangle$ then an acceptable structure relative to $\underline{H}$, $\underline{I}$ will be of the form $\langle\underline{D};\ \underline{A};\ \underline{B};\ \{\underline{a}\}\rangle$ where this structure is a model of this theory and $\{\underline{a},\ \underline{b}\}\subseteq\underline{D}$.

By our definition, we have guaranteed that each of the axioms of this theory will be true at the world $\underline{H}$ just described (under the given interpretation). In fact that theory will be true at every world. Another sentence which is true at $\underline{H}$ is $(\underline{E}!\underline{x})\underline{Cx}$. $(\underline{E}!\underline{x})-\underline{Cx}$ is false for the reason we have seen above. We could adequately express 'there is a unique non-concrete thing' by adding an existence predicate '$\underline{E}$' as described above. $(\underline{x})(\underline{Cx}\ \underline{v}\ \underline{Sx})$ and $(\underline{x})(\underline{Cx}\rightarrow\underline{Sx})$ are both neither true nor false.

Clearly, this theory can be true under this interpretation at a world with any number of things. Thus this theory is not committed to there being more than one thing. This technique can be applied generally to impure set theories, so that we have the result that not only can interpreting pure set theories as IMAIT's yield theories without commitments to pure sets, also interpreting impure set theories as mixes IMIAT's yield theories without commitments to any sets. To take the crucial case, if rather than adding an axiom asserting the existence of the set of concrete things to our modified ZF, we had added an axiom asserting the existence of the unit set of Frege, the resultant theory would have these commitments:

(1) a commitment de re to Frege;

(2) a commitment to there being at least one thing; but not

(3) a commitment de re to the unit set of Frege;

(4) a commitment to there being at least two things.

These points have relevance both to the scope of the method of interpretation developed here and to the work of Michael Jubien in this area. In a forthcoming paper [14] and in published work [12] Jubien presents a theory of mathematical truth which is close to that stated here. Although his treatment of combined and mixed mathematical theories is different from ours, it is not too inaccurate to say that he proposes that mathematical theories be interpreted as IMAIT's. He then correctly draws the inference that under this interpretation pure mathematical theories have no de re commitments.

On the other hand, he makes an assumption that implies that de re commitments to contain impure sets are possible. Apparently, he believes that although we can interpret pure mathematical theories in such a way that they have no de re commitments, it is not possible to interpret impure set theories in this way.

The reason he thinks this is quite clear. In defining an interpretation for impure set theory, in our metalanguage we are constantly "speaking of" impure sets, e.g. each I(H) is an impure set. It is apparently this fact which Jubien has in mind in thinking that he needs to assume that de re commitments to impure sets are possible.

This is an important problem which arises throughout this work. It arose in Chapter Three when we wondered whether interpreting mathematical theories as IAIT's could be used as the means of carrying out a general program of nominalism. At that point we defereed discussion of this problem, so that what follows is intended also to answer this objection as it arose in Chapter Three.

Where does this supposed commitment lie? It does not lie with the object language impure set theory which is to be interpreted, for we have seen that we have a way of interpreting this theory relative to which this theory has no de re commitments to sets. It can only be then that the meta-theory in which this theory is given has commitments to sets. Whether it does though depends on how this meta-theory is interpreted (in a meta-meta-language). The question then is how to interpret the impure set theory which is part of the meta-theory.

Our answer to this is clear: interpret that impure set theory as an IMAIT. On such an interpretation the meta-theory will be seen to have no commitment to sets (de re or otherwise) despite the fact that it does "speak of" sets. Just as although the object-language theory "speaks of" impure sets, but has no commitments to sets, so too for the meta-language. To sum up, the situation is this: no matter what level we are talking at, it will appear that our talk has platonistic commitments; by ascending to a meta-language (with its apparent commitments) we show how to escape this result.

This seems to underline an important aspect of our approach to nominalism. It might be thought that the only way of establishing nominalism would be to provide translations of, say, arithmetic into some chosen nominalistic language. Such has been the approach of many in this field.

For us, by contrast, nominalism is not a question of translation, but of interpretation. On our approach, the nominalist can feel perfectly free to use "platonistic" language without being able to eliminate this language. All that he needs to do is to give an interpretation of this language relative to which this theory is not committed to numbers, sets, etc. In this chapter we have seen how this can be done.

Just as in Chapter Three, the question of applicability arises here. Any informal theory could be represented as an IMAIT. Have we thus opened the door to holding that no theory is committed to there being anything at all? Once again, it is one thing to show that a theory could be represented as an IMAIT and quite another to hold that it ought to be so represented. There is some plausibility, I believe, to holding that, say, set theory could be true in the empty world. Part of this plausibility derives from the idea that set theory would be true no matter what and that there might be nothing at all. The truth of set theory, so to speak, does not depend on any particular facts about any world. Hence it is true in the empty world. The truth of some non-mathematical theories, e.g. 'there are horses' does depend on particular facts about a given world. Hence such theories are not generally true in the empty world, are not true no matter what and so should not be represented by theories which would be true in the empty world. Thus, I am quite confident that the work of this chapter cannot be used as a means of arguing that no theory has any commitments.

A bit of clarification concerning the categories of predicates would be helpful. Roughly speaking, the other-worldly predicates will be those which can appear in pure mathematical theories. The world-bound predicates are those which can appear in theories which do not contain any mathematics at all. The mixed predicates are the other predicates. They are predicates which can only appear in theories having a mathematical and a non-mathematical component. Note, however, that this is only a rough characterization since, for example, this characterization allows '=' to be both a world-bound and an other-worldly predicate.

To be more precise, consider an arbitrary place of an arbitrary primitive predicate. Consider whether we want to guarantee that any satisfier (at H) of that place must actually exist (at H). If so, and if this holds also for any other place of that predicate, then we will classify the predicate as being world-bound. If this does not hold for any place, the predicate is other-worldly. All other predicates are mixed. The places of a mixed predicate for which the above condition does hold are the non-pure places of that predicate. The other places are the pure places.

Note that on this characterization '=' is an other-worldly predicate. Note also that (as '=' shows) that the satisfier (at H) of an other-worldly predicate may exist (at H). Those predicates are other-worldly only in the sense that their satisfiers need not exist. Finally, it is possible that for certain purposes a finer set of classifications of predicates would be needed. It is clear, however, that the given classification is exhaustive and mutually exclusive.

CHAPTER V

In a well-known paper [2] Paul Benacerraf has argued, "there are no things as numbers" ([2]: 73). Benacerraf proceeds in arguing for this conclusion by first arguing that numbers are not sets ([2]: 57-8); then he "extends" this argument to show that numbers are not anything at all ([2]: 69-70).

In arguing that numbers are not sets, he asks the question, of all the set-theoretic models of number theory is there one which is identical with the "intended" model of number theory, i.e. the model whose domain consists of 0, 1, 2, etc.? For example, if we let '0' denote ϕ and let the successor of a set $\underline{x}$ be $\underline{x} \cup \{\underline{x}\}$, we get a model of arithmetic. Is 0 then identical with ϕ? Or is it one of the other sets which is the denotation of '0' in a model of arithmetic?

Benacerraf says,

> if there exists. . . a "correct" account, do there also exist arguments which will show it to be the correct one? Or does there exist a particular set of sets b, which is really the numbers, but such that there exists no argument one can give to establish that it. . . is really the numbers? It seems altogether too obvious that this latter possibility borders on the absurd. If the numbers constitute one particular set of sets, and not another, then there must be arguments to indicate which ([2]: 57-8).

Just what is the argument? Let us use '*' to abbreviate 'it can be shown that' and '$\underline{S}$' to abbreviate 'is a set'.

Now Benacerraf's premises could perhaps be expressed as follows.

(1) $S(0) \rightarrow (Ex)(Sx \,\&\, *(x=0))$

(2) $-\ (Ex)(Sx \,\&\, *(x=0))$

(3) $-\ S(0)$

(1) is a rendering of the principle that if 0 is a set, then there is a set x such that it can be shown that 0 is x. (2) simply says that there is no such x.

How does Benacerraf argue for (2)? He considers the logicist account of numbers which says that 0={ø} and argues that there is no reason to believe that it is true ([2]: 58-62). Then he says that when this is seen, it is clear that there is no reason to believe that any other set-theoretic account of 0 is true. Thus what he has argued (and what I am willing to accept is that

(4) $-*(a=0)$

is true whenever a is replaced by any standard name (e.g. 'ø') of a set. But this does not show that (2) is true. Suppose (4) is true for any such a. Even so, it could be true that

(5) $S(0) \,\&\, *(0=0)$

since '0' is not a standard name of a set. If (5) is true, (2) is false.

In arguing that both (2) and (5) and also (15) and (18) below are contradictories, I assume that the inference from *(... 0 ...) to (Ex*(... x ...) is valid. I think that that inference is valid. For example, we would accept the inference from *(0 is even) to (Ex)*(x is even). Such as inference is not valid for all numerical terms, however. Suppose, for example, that there is a non-constructive proof that there are numbers with the property P. Assume also that this non-constructive proof is the only means we have of knowing that there are such numbers. In particular, we know of no examples of numbers which have P. Then we could introduce 'a' as an abbreviation for 'the least number with property P'. Then we would have *Pa but not (Ex)*Px. If we replaced 'a' with '0', however, the inference would be valid.

Thus we might attempt another reconstruction of Benacerraf's argument.

(6) S(0) → (Ex)*(Sx & x=0)

(7) -(Ex)*(Sx & x=0)

The conclusion of course is (3). (6) again appears to be a reasonable rendering of the principle Benacerraf states. Does the evidence ((4)) Benacerraf presents show that (7) is true? Yes, for assume (7) is false. That is, assume

(8) (Ex)*(Sx & x=0)

Now given the interpretation of '*' the only way for (8) to be true is if

(9) '*(Sa & a=0)'

is true for some substitution of a constant for a. But the only substitutions for a which yield the truth of '*Sa' are precisely those substitutions which do not according to (4) yield the truth of '*(a=0)'. Thus (9) is false for every a and so (7) is true.

In fact, Benacerraf's argument can be simplified here, for by an easy argument (6) implies

(10) $S(0) \rightarrow *S(0)$

and in arguing that (9) is false, we have established

(11) $-*S(0)$

Thus in its simplest form, Benacerraf's argument can be seen as proceeding from (10) and (11) to (3). A slightly more complicated version which is closer to the text proceeds from (6) and (7) to (3). Note also that in accordance with Benacerraf's intentions both of these arguments are easily generalized to show that no number is a set.

Before turning to a consideration of whether (6) and (1) are plausible, let us see how Benacerraf "extends" his argument to show that there are no numbers. Benacerraf says,

> I therefore argue, extending the argument that led to the conclusion that numbers could not be sets, that numbers could not be objects at all; for there is no more reason to identify any individual number with any one particular object than with any other (not already known to be a number) ([2]: 69).

How is this argument being extended? Consider the simple ((10) and (11)) version of his argument that 0 is not a set. The natural way of extending this argument is to say that if 0 is anything at all, then there is something which is such that it can be shown to be 0. More formally,

(12) (Ex) (x=0) → (Ex)*(x=0)

Benacerraf then might claim

(13) -(Ex)*(x=0)

The immediate problem here is that Benacerraf has no independent way to show that the rather implausible (13) is true.

Thus far, however, we have ignored the parenthetical comment which closed the last quotation. This comment suggests that Benacerraf wants to strengthen the consequent of (12) in some way. Perhaps his claim is

(14) '(Ex) (x=0) → (Ex)*(Px & x=0)'

is true for some replacement of 'P' by a predicate which "does not already identify x as a number." Now this condition could perhaps be stated as requiring that such a predicate 'P' be such that 'for all x, if x is P, then x is a number' is not analytic. But in order to block trivial counterexamples, (e.g. 'self-identify') we should also require that 'for all x, if x is a number, then x is P' also not be analytic.

First of all, it is easy to be puzzled why Benacerraf would call (14) an extension of either (6) or (10). In fact, if Benecerraf wants to argue from (14) and

(15) '-(Ex)*(Px & x=0)'

is true for every 'P' which meets his condition, to

(16) -(Ex)(x=0)

his earlier argument actually leads us to doubt his claim (15). Having already shown (3), he has established

(17) *-S(0)

'-S' meets his condition since 'all non-sets are numbers' is false and since 'all numbers are non-sets' though true according to Benacerraf is not analytic. Thus (17) together with '*(0=0)' which is plausible and which is not something Benacerraf can assume is false implies

(18) *(-S(0) & 0=0)

which is a counterexample to (15). Thus, so far from this second argument being an extension of his first argument, the first argument actually undercuts the second.

Perhaps, however, it is not (14) which Benacerraf means to defend. Perhaps, he means to defend the stronger claim

(19) '(Ex)(x=0) → (Ex)*((y)(Py↔y=x) & x=0)'

is true for some substitution of a predicate for 'P' which meets his condition. The point of (19) is that if 0 exists it can be uniquely picked out in some way which does not already identify it as a number. His further claim then, of course, is that the consequent of (19) is not true for any predicate which meets his condition.

Once again it is difficult to see that the considerations which might motivate acceptance of (6) or (10) also would motivate acceptance of (19). Benacerraf does not explain. In any case, it is doubtful that he can show that the consequent of (19) is false for any predicate which meets his condition. Consider 'is thought about by Tom', 'is named by '0'', 'is proved by Tom to have some property', and 'is a counterexample to the claim that all universities have more than two students and all even numbers are greater than 1'. Each of these predicates meets his condition and substituting any of them for '_P_' in (19) could under the appropriate conditions make the consequent of (19) true. I think then that we must conclude that at least in this form, Benacerraf's argument for (16) is a failure.

Let us now consider (6) and (10). Since we have accepted (7) and (11), if either (6) or (10) is reasonable to believe Benacerraf would have a convincing argument that 0 is not a set. First of all, what does Benacerraf say in favor of these premises? He says that in saying what we have rendered as (6) and (10),

> I am not committing myself to the decidability by proof of every mathematical question--for I consider this neither a mathematical question nor one amenable to proof. . . In awaiting enlightenment on the true identity of 3 we are not awaiting a proof of some deep theorem. Having gotten as far as we have without settling the identity of 3, we can go no further. We do not know what a proof of that could look like. The notion of "correct account" is breaking loose from its moorings if we admit of the possible existence of unjustifiable but correct answers to questions such as this ([2]: 58).

This is not very helpful. If 0 is a set why should it be showable that it is a set? All that Benacerraf says is that he does not defend (6) or (10) because he defends a general principle that every truth can be shown to be true. He does nothing to explain why this principle though not plausible in general is plausible in this particular case. Would he also want to say that if a proposition is a state of affairs, it must be showable that it is or that if a mental state of affairs is a physical state of affairs, it must be showable that it is? It would have been helpful if Benacerraf had answered such questions.

Perhaps we can gain some insight into the plausibility of (10) (and thus (6)) by considering the following closely-related principle. Let '$\underline{S}$' abbreviate 'is a non-set'. Consider

(20) $\underline{S}(0) \quad *\underline{S}(0)$

(20) just says that if 0 is a non-set, then it can be shown that it is a non-set. The relation of (20) to (10) is quite interesting. First of all, although in the absence of a clear justification of these principles, it is difficult to be certain about this, it appears as though (10) is plausible if (20) is. The two principles are, so to speak, on the same level. It is rather difficult to imagine a justification of either of these principles which would not also be a justification of the other. But now consider.

(21) $-*\underline{S}(0)$

This is just as plausible as (11). Benacerraf agrees ([2]: 67). So we have

$$(22) \quad -\underline{\mathcal{S}}(0)$$

But then we have an interesting argument

$$(23) \quad (\underline{Ex})(\underline{x}=0) \rightarrow (\underline{S}(0) \vee \underline{\mathcal{S}}(0)).$$

Difficulties might arise for (23) if 'is a set' has in English a limited range of significance so that, perhaps, 'Nixon is a set' has a truth value gap. If this is so, then 'is a non-set' has two possible readings depending on whether things which fall out of the range of significance of 'is a set' satisfy 'is a non-set'. I intend the broader reading so that whether or not 'is a set' has a limited range of significance, (23) is true.

(23) together with (3) and (22) implies (16). In other words, Benacerraf can argue from (10) and (20) to the conclusion that 0 does not exist. Even though his own "extension" of the argument that 0 is not a set does not show that 0 does not exist, here is another extension (which in a clear way _is_ an extension) of that argument which does seem to show that 0 does not exist. That is, it shows it if (10) and (20) are reasonable to believe.

Unfortunately, (10) and (20) are too strong to be plausible. Let '_a_' be the name of, say, the proposition that Nixon exists. Then we can construct an argument exactly parallel to that which we have just constructed.

(24) $\underline{S}(\underline{a}) \rightarrow {*}\underline{S}(a)$

(25) $\underline{\not S}(a) \rightarrow {*}\underline{S}(\underline{a})$

(26) $-{*}\underline{S}(\underline{a})$

(27) $-{*}\underline{\not S}(\underline{a})$

(28) $(\underline{Ex})(\underline{x}=\underline{a}) \rightarrow (\underline{S}(\underline{a}) \vee \underline{\not S}(\underline{a}))$

(29) $-(\underline{Ex})(\underline{x}=\underline{a})$

Each of the premises of this argument is just as plausible as the principles from which they are obtained by replacing '0' with '$\underline{a}$'. Furthermore, we could replace '$\underline{a}$' with any other name of a necessary existent (except the standard name of set) and the resulting premises would be equally plausible. And as for standard names of sets, if we replace '$\underline{a}$' by such a name and '$\underline{S}$' by 'is a number' and '$\underline{\not S}$' by 'is a non-number', once again the resulting premises are plausible. (I do not think that '$\underline{a}$' could be replaced by a name of a contingently existing being if '$\underline{S}$' means 'pure set' i.e. a set which does not contain any non-sets in its transitive closure, for any such thing is a non-set because it, unlike a pure set, does not necessarily exist.) In other words, using the same sort of argument Benacerraf uses to show that 0 is not a set, it can be shown that there are no beings which exist necessarily.

I think this shows that (10) is too strong to be plausible. Let me summarize. Benacerraf has not explained why (10) is plausible, but it appears clear that (10) is plausible if (20), (24) and an infinite number of other principles are plausible. If we assume (10) and hence the other principles patterned after it, we can show that there are no beings which exist necessarily. But that such a strong conslusion could be obtained from assuming (10) makes it seem highly implausible. Of course, Benacerraf would answer this argument by showing that (10) unlike at least some of the principles patterned after it is plausible, but he certainly has not done that. I thus conclude that in the absence of such a justification of (10), it should be rejected. To put it succinctly, given that the fact that we cannot tell whether or not God is a set does not show that God does not exist, so the fact that we cannot tell whether or not 0 is a set does not show that 0 does not exist. To apply this directly to (10), given that the fact that we cannot show that God is a non-set does not show God is a set, so the fact that we cannot show 0 is a set does not show it is not a set.

I believe then that Benacerraf's argument that there are no numbers which relies on the epistemological premises (6) and (10) is a failure. In fact, however, certain things Benacerraf writes suggest that he not only had this epistemological argument in mind, but also another argument. The crucial passage here is

> For arithmetical purposes the properties of numbers which do not stem from the relations they

> bear to one another in virtue of being arranged in a progression are of no consequence whatsoever. But it would only be these properties that would single out a number as this object or that.
>
> Therefore numbers are not objects at all, because in giving the properties (that is, necessary and sufficient) of numbers you merely characterize an <u>abstract structure</u> and the distinction lies in the fact that the "elements" of the structure have no properties other than those relating them to other "elements" of the same structure . . . That a system of objects exhibits the structure of the integers implies that the elements of that system have some properties not dependent on structure. . . To <u>be</u> the number 3 is no more and no less than to be preceded by 2, 1 and possibly 0 and to be followed by 4, 5, and so forth. ([2]: 69-70).

The epistemological connotations of 'single out' suggest that in the first paragraph he is perhaps appealing to (19). If so, the argument here is now familiar. But in the following paragraph (with one exception: he speaks of the problem of "individuating" objects), there is no hint that he is appealing to epistemological premises such as (6), (10) or (19). In short, it appears as though, having stated his epistemological argument, he then turns to approach the problem from a new point of view. What then does he argue here?

The claim is, roughly, that numbers do not have enough properties to exist. This is, of course, what his epistemological argument attempted to establish from the premise that we cannot tell whether any number is a set. Here, however, the claim might be simply that for no number *x* is being a set or being a non-set included among the properties of <u>x</u>. What is Benacerraf's argument for this? Besides his epistemological argument, the only passage which is relevant is the second paragraph of the above quotation. Let us examine just what Benacerraf has said in the second paragraph. First, I will take a quite literal interpretation of this paragraph. Then I will consider a weaker and perhaps more charitable interpretation of it.

(i) <u>In giving the necessary and sufficient properties of numbers you merely characterize an abstract structure</u>. I believe that we could replace 'abstract structure' here by 'isomorphism type'. Then the claim is that, for example, 0 has only those properties which are had by every thing which is the "0" of some structure which has the isomorphism type of the numbers.

But consider the property of being distinct from Frege. It is plausible to think that 0 has that property, because 0 necessarily exists and Frege does not. Furthermore, Frege is the "0" of a structure which has the isomorphism type of the numbers.

Note also that it is plausible to think that 0 alone has the property of being the predecessor of 1. It is the only thing which has <u>that</u> property. Of course, things other than 0 satisfy the formal equivalent of 'is the predecessor of 1' in some models of arithmetic, but this does not show that these things have the property of being the predecessor of 1.

(ii) <u>The numbers only have properties relating them to other numbers</u>. The example noted above is a counter-example to this claim. Being distinct from Frege is a property 0 has which is not a property which relates 0 to other numbers.

But furthermore since all the properties relating numbers to numbers are had, if at all, necessarily, this claim would imply that 3 does not have any properties contingently. 3, however, has the property of numbering the books on my desk contingently.

(iii) <u>To be 3 is no more and no less than to be preceded by 2, 1,</u>. . . If Benacerraf is simply saying that 3 is uniquely picked out by 'the successor of 2' he is not saying anything that a platonist would deny. If he is saying that the only properties 3 has are those of being the successor of 2, being the successor of the sucdessor of 1, etc., his claim would seem to be refuted by the fact that 3 has the property of being prime.

Perhaps instead the claim is that if 3 has a property <u>P</u> then the proposition that 3 has <u>P</u> is entailed by the conjunction of the propositions that 3 is the successor of 2, that 3 is the successor the the successor of 1, etc. The difficulty here is that this conjunction would be necessarily true and since no necessarily true proposition entails any contingently true proposition, it would follow that 3 does not have any properties contingently. We have already seen that this is false.

We have reached this situation: Benacerraf attempts to argue, roughly, that numbers do not have enough properties to exist. He attempts to show this by arguing that numbers have only certain "structural" properties. We have just argued that this strong claim is false. Thus, as far as he develops it, this non-epistemological argument is a failure.

But, of course, this strong claim could be false while it is still true that numbers do not have enough properties to exist. (Perhaps, in fact, it was some such weaker claim that Benacerraf meant to endorse. What follows then may be understood as another interpretation of the above quotation.) In particular, is 0, say, $\emptyset$ or not? If one set theorist constructs arithmetic in such a way that 0 is identified with $\emptyset$ and another constructs arithmetic in such a way that 0 is identified with $\{\emptyset\}$ is at least one of them accepting a (harmless) falsehood? I have no argument to show that the answer to this question should be, No, but I do have an intuition that the answer is, No.

This may be clearer in the case of ordered pairs. Are ordered pairs sets? The facts are that ordered pairs can be "identified" with sets. Beyond that there seems to me to be no further question: '$\langle \underline{a}, \underline{b} \rangle$ is $\{\underline{a}, \{\emptyset, \underline{b}\}\}$' is neither true nor false.

It should be clear that if there is anything to this intuition, it applies generally to mathematics. 'Rationals are ordered pairs of natural numbers', 'reals are sets of rationals', 'complex numbers are ordered pairs of reals', 'n-place relations are sets of ordered n-tuples' all are neither true nor false.

In some cases I hope this will seem to be quite clear: 'ordered pairs are sets' may be a case of this. Furthermore, if it is thought that one of these sentences just listed is neither true nor false, then I think that it should also be held that they all are neither true nor false. There does not appear to be any important difference between these examples.

One final point is this: this intuition applies not only to pure mathematical entities, but also to impure ones, e.g. '$\langle$Nixon, Ford$\rangle$ = {Nixon, {$\emptyset$, Ford}}' is neither true nor false.

In short, I have a certain intuition. Benacerraf apparently does, too, and he thought he could support this intuition by two arguments, both of which I have argued are unsuccessful. I have no arguments to replace his here although a principle quite similar to (19) is discussed in Chapter Six. Simply relying on an intuition is a rather weak position to be in, but I think that this position can be strengthened by showing how a theory of mathematical truth can be constructed which does accord with this intuition. It is also interesting to see how a platonistic theory of mathematics cannot account for this intuition.

According to the platonistic theory I am interested in discussing, both '0' and $\emptyset$ have single referents and there is a class which is the extension of 'is a set' ('S'). If so, then both

(30) $0=\emptyset$

and

(31) $\underline{S}(0)$

are either true or false. According to my intuition, neither of these sentences is either true or false.

Is there then a modified platonistic position which would accord with this intuition? Consider (31). Intuitively, we could pin the truth-valuelessness of (31) on either 'S' or '0'. Either 'S' does not have a unique extension or '0' does not have a unique referent. Similar remarks apply to (30).

Thus far, however, the platonist does not appear to be forced to pin the difficulty on 'S' or to pin it on '0'. Whatever choice is made, however, forces certain other choices. For example, if we say 'S' has a unique extension, pinning the difficulty on '0', then we will have to pin the truth-valuelessness of '$\underline{S}(\langle\emptyset, \emptyset\rangle)$' on '$\langle\emptyset, \emptyset\rangle$', thereby rejecting a platonistic theory of ordered pairs. Other choices will also be forced.

The intuition I have, however, will not through considerations like these force saying that '<u>S</u>' does not have a unique extension, that '0' does not have a unique referent, etc. All that is forced is certain conditional statements such as, if '<u>S</u>' has a unique extension, then '0' does not have a unique referent.

It is thus quite consistent with my intuition that a platonistic theory of sets, say, is true. The modified-platonist who has my intuition could say '0' does not have a unique referent, but '$\emptyset$' does; there are sets although there are no numbers.

Although this position is consistent with my intuition, it is ill-motivated. What, besides a desire to preserve <u>some</u> form of platonism would lead anyone to accept such a modified view? In the absence of some argument to the contrary, it is reasonable to hold that if either '0' or '$\emptyset$' lack unique referents then both of these terms do. In short, anyone who accepts my intuition should also reject the thesis that any mathematical term (predicate) has a unique referent (extension). Such a person should reject any form of platonism with respect to mathematical entities.

In short, my intuition is incompatible with a certain natural view of mathematical truth. In the absence of another theory of mathematical truth, the most reasonable response to this would be to say: so much the worse for my intuition. As we have seen already, however, there are theories of mathematical truth which are compatible with this intuition. If a combined arithmetic and set theory is interpreted as either an IAIT or a IMAIT, then such sentences as (30) and (31) can be allowed to be neither true nor false.

This fact provides some support for this intuition, but the light shines in the other direction, too. If interpreting mathematics as either an IAIT or a IMAIT allows us to account for an intuition not to be accounted for in other ways, this provides some jusification for saying either that informal mathematics is best represented as an IAIT or that informal mathematics is best represented as an IMAIT. These considerations, however, are neutral between whether mathematics is best represented as an IAIT or an IMAIT.

CHAPTER VI

It has often been thought that platonism suffers from epistemological problems: we know certain mathematical truths, but how could we know such truths given a platonistic account of mathematical language? A related question is, if we take platonism to be the claim that mathematical constants refer to mathematical entities, then how is it that they do refer? This is the question raised by Michael Jubien in a recent paper [14].

Jubien approaches this problem by asking if there is a satisfactory platonistic account of mathematical truth. He conceives of a platonistic account of mathematical truth as proceeding by first formalizing the syntax of an informal mathematical theory and then selecting a model (whose domain consists of pure abstract entities) of that formal theory. A sentence of the informal theory is said to be true if its formalized counterpart is true in the selected model.

Jubien does not claim that this is the only way a platonist could give a theory of mathematical truth, but he does suggest that the difficulties confronting the above approach would affect any other "platonistic" approach.

Jubien argues that this platonistic account breaks down because we cannot select a model (whose domain consists of pure abstract entities) to interpret a formalized mathematical theory. Suppose, for example, that this theory has the constant '0' in its language. Then, to specify a model of this theory it is necessary that '0' be assigned some pure abstract entity as its denotation.

Now the obvious way of doing this is to say: let '0' denote zero. (For the sake of clarity I use ''0'' always as a name of the formal theory constant and ''zero'' always as a name of the informal theory constant. I follow a similar convention with other number words.) Jubien does not deal with this directly, but he does deal with a similar obvious approach when he says that if we were to assume from the outset that there are, say, sets, this would be inappropriate because, "it is tantamount to the assumption that there is a satisfactory platonistic account of mathematical truth"([14]: 2).

Apparently then the project which Jubien presents to the platonist is to select a model without using specifically mathematical language in the metalanguage to do this.

Suppose then that the platonist cannot do this. What would this show? Jubien concludes that a platonistic account of mathematics cannot be given from the fact that the platonist cannot complete this project. But why should it be supposed that even if a platonistic account of mathematics is true, that the platonist should be able to give such an account without using mathematical language?

Jubien does not directly deal with this question, but at least the outlines of the answer seem to be clear.

He takes it to be an open question whether or not a platonistic account of mathematics is true. To assume that a denotation for '0' could be established simply by saying, "let '0' denote zero," is to assume that 'zero' has a referent. And that is essentially to assume that a platonistic account of mathematics is true. The platonist who takes this obvious approach would thus be question-begging.

Suppose then that there is no non-question-begging way of selecting a model. Jubien concludes from this that a platonistic account of mathematics is false. Why does the conclusion follow?

Jubien apparently is appealing to the premise that if 'zero' has a referent, then there is a non-question-begging way of assigning a denotation to '0'. To repeat, a non-question-begging way of doing this is to do it without using mathematical language.

Why should we accept the premise that if 'zero' has a referent, then there is a non-question-begging way of assigning a denotation to '0'? Once again, I am going somewhat beyond what Jubien says, but I suggest that the reason is that if 'zero' has a referent, then someone must have fixed its reference. Furthermore, any terms or predicates used in such a reference-fixing must already (temporally) have denotations.

Now perhaps 'zero' had its reference fixed as referring to the first natural number. That would be acceptable provided 'first natural number' already (temporally) had a denotation, but not every mathematical term or predicate could have its denotation fixed by using mathematical language. The resultant circle would in this case be vicious, since the terms used in fixing a term's reference must already (temporally) have denotations.

The premise which appears to be justified by this line of reasoning is that if 'zero' has a referent, then there is non-question-begging way of assigning a denotation to some mathematical term or predicate. Jubien can then be understood as arguing that there is no such way.

It might be objected to the above argument that it is false that if 'zero' has a referent, then someone must have fixed its reference, since this seems to imply that there was some one occasion on which 'zero' (or some word from which it has been derived) was introduced as a referring expression. This might seem to be excessively unrealistic.

Granted. Let us not therefore read into the expression 'the reference of a was fixed' the implication that there was some one occasion on which a was introduced as having a specific referent. Still, if 'zero' refers to something, it is because of some facts about the people who use this word and it could not be that every mathematical term or predicate had its reference fixed by using some other mathematical term or predicate. In other words, some people brought it about (perhaps not intentionally and perhaps not at one specific time) that 'zero' has a referent (if in fact it does). We can do the same thing (in an idealized fashion) just as they did, that is, without using some other mathematical term or predicate in fixing the reference of any such term or predicate. Thus we have our conclusion: if 'zero' has a referent, then we can assign a denotation to some mathematical term or predicate without using mathematical language.

This has been said in clarification and support of Jubien's approach. He then proceeds by asking whether '0' is a term which can be assigned some pure abstract entity as its denotation without using mathematical language. He says, "at the outset there seem [to be] only two possible ways [of doing this]: by ostension and by (unique) description" ([14]: 3).

At this point we can give Jubien's argument very quickly: '0' cannot be assigned a denotation by ostension because only sensible things can be ostended and no pure abstract entity is sensible. '0' cannot be assigned a denotation by unique description, because the only descriptions which even appear to describe a pure abstract entity uniquely all use mathematical language.

This argument can easily be generalized as Jubien does: no mathematical term or predicate can have its denotation fixed by ostension and the only descriptions which even appear to describe uniquely the denotation of any such term or predicate all use mathematical language. Thus, no mathematical term or predicate has a denotation.

This is Jubien's conclusion, but it appears that he might have drawn a much stronger conclusion. There are some difficult issues here, but prima facie the same sort of argument shows that 'property' 'state of affairs' and 'proposition' all do not have denotations. (A possible difficulty here is that some people have claimed that we can actually see properties such as redness, so that such properties are sensible.)

In particular, however, in a later part of his paper, Jubien assumes that although we cannot refer to pure sets, we can refer to any non-empty set, provided that we can refer to all of its members.

Jubien does not attempt to support this assumption. Furthermore, I believe that it is unwarranted given his argument that reference to pure sets and numbers is impossible, since the same sort of argument can be used to show that 'the empty set' does not have a referent. That is, if a term which purports to refer to a set all of whose members are specifiable in fact has a referent, then some similar term can be assigned a denotation without using that sort of language. Such terms purport to refer to abstract entities. Therefore no such term can be assigned a denotation by extension. Finally, the only descriptions which appear to describe the referent of such a term uniquely all use one or more of those terms. Therefore no such term has a referent.

Finally, the restriction that any non-empty set is specifiable, provided its members are appears arbitrary. Why not say that any set is specifiable provided its members are? If so, then the empty set would be specifiable as 'the set containing no members'. Thus, Jubien's argument can be used to undercut his assumption and also the restriction on his assumption appears to be arbitrary.

Of course, in Chapter Four it was argued that in fact his approach does not require the assumption that any sets can be objects of reference. If that argument was unsound, however, we have now seen that if Jubien has undercut platonism, he has also provided grounds for rejecting the assumption he makes in constructing his own theory of mathematical truth. Does his argument show that reference to pure abstract entities is impossible?

There are some apparent counterexamples to the claim that '0' cannot have its reference fixed without using mathematical language. 'The referent of 'zero'' and 'what I am now thinking about' both might uniquely describe some pure abstract entity and neither use (though the first mentions) any mathematical language.

If we think of ourselves as attempting to recreate (in an idealized fashion) how it was that 'zero' came to have a referent[1] we cannot use the expression 'the referent of 'zero'' to do this for no such expression could have been used by anyone to pick out the referent of 'zero' until 'zero' already had a referent. The case of 'what I am now thinking about' seems to suffer from a similar fault. 'What I am now thinking about' picks out something only if I have picked out that thing sufficiently so that I can have de re beliefs about it. But then, however it was that I picked that thing out, I could have used the same method directly to fix the reference of '0'. Fixing the reference of '0' by using 'what I am now thinking about' thus only transfers our problem to asking how it was that I am able to have de re beliefs about that thing.

We have spoken of fixing the reference of '0' in a non-question-begging way. By that we meant doing it

---

[1]It is appropriate here to say why we do not discuss Quine's approach in The Roots of Reference to carry out this project of re-creation. My reason for this is that as I read "Ontological Relativity", Quine thinks that there is no such thing as reference. The project which I take him to be pursuing in The Roots of Reference is that of explaining how, for example, number words came to be used as if they referred. No reason is given for thinking that they do refer. In fact, the picture Quine paints of how number words came to be used as if they referred--essentially, a series of confusions, would, if anything, indicate that number words do not refer.

without using mathematical language. The above remarks suggest that we should expand this notion by requiring that we fix the reference of '0' by not only not using mathematical language, but also by not using semantical or intentional language. Furthermore, for the reasons indicated in the preceding paragraph, the argument that if 'zero' has a referent then '0' can be assigned a denotation in this way goes through as before.

I believe that it is clear that no mathematical term or predicate can be assigned a denotation by unique description in a non-question-begging way. If we consider the set S of mathematical terms and predicates, certain subsets of S are such that the denotation of every mathematical term or predicate can be fixed in terms of the members of S, but we cannot (without cheating by using semantical or intentional language) fix the denotation of any member of S without using some other member of S.

Further doubts about Jubien's argument might come at two other points. (a) Is it true that we cannot assign a denotation to any mathematical term by ostension? (b) Is it true that if we can assign a denotation to say, '0', then this must be done either by ostension or by unique description?

(a) Jubien argues that if we can assign a denotation to '0' by ostension, then that denoted thing is sensible. He is here thinking of what Quine calls direct ostension:

the ostended thing is actually sensed or in some way intuited ([26]: 39-40). Now some platonists, e.g. Plato ([20]: 81b-e, 85d-86b; [21]: 73c-77b, 83b), Russell ([30]: 101), Godel ([9]: 271), have thought that at some times we actually sense abstract entities. For such platonists, direct ostension would be sufficient to assign a denotation to mathematical terms or predicates. As Jubien points out, however, there seems to be no such sense ([14]:4). In the absence of a much more adequate discussion of this supposed sense, the view that we have such a sense is unsatisfactory given the widespread belief that we have no such sense.

Quine also discusses deferred ostension: this is the case where we point at something with the purpose of ostending something else. Could we assign a denotation to '0' (or some other numberal, '0' being, perhaps, an unfortunate example) by deferred ostension?

I am aware of no explicit attempt in the literature to argue that this can be done. (In fact, apart from the just-mentioned theory of a special sense propounded by Plato, Russell and Godel little has been said about how a denotation could be assigned to, say, '0'.) Perhaps the closest to such an attempt is Quine's discussion of deferred ostension, but this is undercut by Quine's general doubts as to whether any term has a referent.

The question I wish to pursue here is whether if we, so to speak, subtract Quine's doubts about ostension from his discussion of deferred ostension, we would have the outlines of a theory according to which we could, say, assign a denotation to '0' by deferred ostension.

What is involved in deferred ostension? Take a simple case. You point at large footprints in the ground and say, "Bigfoot" meaning to introduce 'Bigfoot' as a predicate true of the animals which make those sort of prints. (To make this as simple as possible, assume there are such animals.) Under what conditions are you successful in doing this? You need to know some relationship which holds between the prints and a Bigfoot. In this case the relation is: a Bigfoot makes these prints. If you did not know that that was the relation in question, you would not successfully fix the extension of 'Bigfoot'.

Take another case: you want to fix the reference of 'Leibniz' for a friend. So you point at a picture of Leibniz and say, "Leibniz". In order for you to succeed, your friend needs to know what the relation is between what you point at and Leibniz. In this case, the relation is: _is a picture of_.

These examples suggest the following necessary condition for successfully fixing the reference of a constant by deferred ostension.

(C) x fixes the reference of 'a' to be a by ostending b only if there is some relation R such that R holds uniquely between a and b and x believes that R holds uniquely between something and b.

The parallel principle for predicates is:

(P) x fixes the extension of 'P' to be a set S by ostending b only if there is a relation R such that R holds between S and b and x believes that R holds between something and b.

These do seem to be necessary conditions of fixing the reference of a term or predicate by deferred ostension. In order to fix reference by deferred ostension, I need to know what relation the referent-to-be bears to the directly ostended thing. In the case of (C) if, furthermore, x were to intend to let 'a' refer to the unique thing which bears R to b, that would be sufficient for x to fix the reference of 'a' to be a. This makes it clear that (C) and (P) are the hard cases: if (C) or (P) could be satisfied in the case of numbers, then in fact the denotation of some arithmetical terms or predicates could be fixed by deferred ostension. Thus, henceforth we will treat (C) and (P) as not only necessary, but also sufficient conditions for fixing reference by deferred ostension.

Consider now Quine's simple example of deferred ostension of linguistic types by direct ostension of tokens of that type. (For reasons to be noted later, this is a simpler case than that of deferred ostension of numbers.) Could I fix the reference of 'the type of 'a'' by ostending 'a'? According to (C), if I can do this, then there is a relation R such that I believe truly that there is something (which in fact is the type of 'a') which bears R uniquely to 'a'. In this case it is clear that the relation in question is: is the type of.

Suppose (this is the crucial case as we have seen) further that I do not as yet have in my vocabulary any synonym of 'is the type of'. I am, so to speak, just trying to "break into" talk about types. Could I then truly believe that there is something which is the type of that token? That appears to be the question, since it is to ask whether I could fix the reference of 'the type of 'a'' by deferred ostension without using talk about types.

I am arguing that if the reference of 'the type of 'a'' can be fixed by deferred ostension, then prior to using type-language I would have to believe truly that there is something which is the type of the token 'a'. Such belief would have to be language independent in the sense that I would have that belief at a time when I lacked the linguistic means of expressing it.

Furthermore in an historically interesting sense, this belief would be innate. This historical parallel is worth considering. A modern philosopher would have said that if I believe that there is a thing that is the type of 'a', then I have the concept or idea of being a type.

Furthermore, for the same reason that we have said that the extension of 'number' cannot be fixed by direct ostension or fixed by unique description without using mathematical language, a modern philosopher (who accepted our reasoning) would have said that this concept of being a type is neither derived directly from experience, nor is definable in terms of concepts which are derived directly from experience. In short, it is precisely the sort of concept which Locke and Hume say we do not have and Descartes and Leibniz say we do have. It is an innate concept or idea.

This historical parallel can even be drawn further. Leibniz says that although we do not derive the ideas of, say, mathematics from the senses, the senses provide the occasion which allows us to bring them to mind ([33]: 41-3, 53-4). Deferred ostension would seem to be a clear case of this: pointing at the token '<u>a</u>' brings to mind the idea of the type of that token even though this idea is not derived from the senses (in the sense of being derived directly from the senses or defined in terms of such ideas).

To say then that we could believe truly that there is a thing which is the type of '<u>a</u>' even though we have no type-language implies the interesting claims that (i) we can have the belief that there is something which is the unique type of token '<u>a</u>' at a time when we lack the linguistic means of expressing that belief and (ii) we have what modern philosophers would have called the innate idea of a type.

(ii) appears to be the more likely place to attack and the empiricists, of course, had such an attack. In short, Jubien's argument that mathematical terms do not have referents depends on familiar empiricist assumptions in the sense that at least in outline the rationalists had an answer to his argument and that the most likely answer to the rationalists' argument is an empiricistic theory of the origin of our ideas.

To categorize his argument in this way is not, of course, to criticize it. It is, however, to point out that his argument is not "knock-down" in the sense that there is no even slightly plausible reply to it. There is such a reply and it can be adapted from the work of Descartes and Leibniz. In order to show that numerals do not have referents, he would have to show that the details of the rationalists' theory could not be developed in a plausible fashion.

Thus far only linguistic types have been discussed. Let us briefly apply the above points to the case of natural numbers. In order to fix the reference of '2' by deferred ostension we need doubly-deferred ostension. (This is why this case is more complicated than the case of linguistic types.) If we point at two apples and say "two" meaning to fix the reference of '2', this will be successful only if two bears the relation of being the number of to something which we ostend. But the thing to which two bears this relation is not anything we can directly ostend; rather it is a concept or a set. Thus we can think of this process as involving first direct ostension of the apples, which brings about deferred ostension of a set or concept, which finally brings about deferred ostension of two.

(b) Is it true that if we can fix the reference of '0', then this must be done either by ostension or unique description? If by 'ostension' we mean 'direct ostension', then we have already discussed one possible alternative to this dichotomy. Let us consider another alternative.

The following are, I think, familiar facts about teaching, say, set theory to someone who has never heard of set theory. Such teaching often proceeds not by attempting to ostend sets or by attempting to describe uniquely any sets, but rather by simply introducing the learner to some elementary facts of set theory. We say, for example, that there is a set which has as its only members Nixon and Ford; that for any two sets there is a set which contains all and only the members of those sets; that if two sets have all the same members, they are identical, etc. At some point, the learner will stop and say, "Oh, now I get it."

This suggests that another way of fixing the reference of mathematical terms is simply to state some of the appropriate theory. Note that stating this theory in no way involves uniquely describing any mathematical entity without using mathematical terms. The question to ask then is, is it plausible to suppose that reference could be fixed in this way?

I am again inclined to think that there is at least some historical precedent to suppose that it is plausible. If we take the rationalist point of view that there are innate ideas, then, as we have seen, we can regard the direct ostension of a token as the trigger which brings to mind the idea of a type. Similarly, taking the same point of view, we could regard description of some of the fundamental facts of set theory as the trigger which brings to mind the idea of a set. We have already seen that there is some corroboration to this view in that we can get someone to have the idea of a set simply by telling that person some set theory.

Having noted this, let us now return to the topic of deferred ostension. Dropping talk of innate ideas, which may introduce unnecessary questions of historical accuracy, let us see exactly what is involved in our reply to Jubien's argument. The claim is that we could fix the reference of 'the type of '<u>a</u>'' without using any type language provided we could truly believe that there is a type of the token '<u>a</u>' even though we cannot see types, nor can we uniquely describe them without using type language. The difficult point here is, could we have such a belief? We have pointed out, first that Jubien does not argue that we could not have such a belief and second that there is historical precedent for supposing that there are such beliefs.

If we are to be driven in the direction of holding that there are such beliefs, it would be important to know more about them. What beliefs are of this kind? What more can be said about the relationship between these beliefs and the experiences which are the occasion of our bringing them to mind? Before they have been brought to mind, in what sense do we have these beliefs or the concepts which are used in these beliefs?

I am not sure, however, that we need be pushed further in the direction of developing this rationalist theory. Traditionally the argument which has been used in favor of saying that there are such innate concepts or beliefs has been an "how else" argument: we have these concepts (beliefs), they could not be derived from experience, thus how else could we have them unless they were innate. (See ([20]: 81b-81e, 85d-86b; [33]: 41-7, 63-4).)

In this case we have a double application of the "how else" argument. It is argued that since, say, Peano arithmetic is true, how else could that be unless numerals have referents. (See ([32]: 57).) And, thus, since numerals have referents, how else could that be unless we have innate ideas. In earlier chapters we have seen, however, two different ways to block the first application of the how else argument. That is, we have shown that we can account for mathematical truth without viewing, say, numerals as names of pure abstract entities by construing mathematical theories either as IAIT's or as IMAIT's.

Furthermore, when we undercut this how else argument, I think we also remove much of the motivation for accepting the rationalists' theory we have outlined. I think that the simplicity of empiricist accounts of language and knowledge are appealing and that one of the major appeals of rationalism is simply that it is felt that empiricism cannot account for the facts. If we can show (as I think we have) that mathematical truth can be accounted for without positing innate ideas of types, numbers, sets, etc., then that severely undercuts the theory that there are such ideas and, more importantly here, it also strongly supports viewing mathematics either as an IAIT or as an IMAIT.

This line of thought would be undercut if we were forced to recognize innate ideas to, say, account for language acquisition. That would open the way to supposing that we have innate mathematical concepts. My claim here is simply that mathematics need not force us in this direction.

Jubien apparently regards his argument as a knock-down argument against a platonistic theory which holds that mathematical constants have referents. I think that enough has been said to show that this argument is not as strong as that. On the other hand, the most likely platonistic account of how mathematical constants do refer has little motivation given that there are competing theories of mathematical truth.

Questions have often been raised as to whether a non-platonistic theory of mathematical truth is possible. A more subtle and equally important question has been raised by Paul Benacerraf: even assuming that a non-platonistic account of mathematical truth can be given, will this account carry over in such a way that we can still give this account when we embed mathematical talk into other kinds of theories ([1]). We have seen in Chapters Three and Four that not only can a non-platonistic account of mathematical truth be given, this account does also carry over to mixed theories. These facts should undercut much of the doubt as to the possibility of adequate non-platonistic theories of mathematical truth.

The topic discussed here connects nicely with the discussion of Chapter Five. That discussion concerned the truth value of such sentences as '$0=\emptyset$'. Now if it is true that neither '0' nor '$\emptyset$' have referents, it is very natural to expect that there might be perplexity as to the truth value of such sentences. In other words, we can view this perplexity as some corroboration of the conclusion of Jubien's argument that neither '0' nor '$\emptyset$' have referents.

It would be helpful to summarize these considerations.

(1) We have a weak argument against platonism which relies on the (unsupported) intuition that '0=∅' is neither true nor false.

(2) We have a more persuasive, but still inconclusive, argument against platonism put forward by Jubien (which we have just discussed).

(3) The most likely reply to Jubien's argument has little motivation if there are non-platonistic theories of mathematical truth.

(4) The fact that there are perplexities over '0=∅' might be expected if (and explained by the hypothesis that) neither '0' nor '∅' have referents.

(5) We have stated two theories of mathematical truth which both allow '0=∅' to be neither true nor false and also do not construe '0' and '∅' as having referents; furthermore both of these accounts carry over to mixed theories.

(6) Finally (and this is a point we have not stressed before) both of these theories of truth are simpler than the platonist's account in that on these two theories there is no need to postulate a special kind of entity (the mathematical entity) in order to account for mathematical truth.

Alone these points are quite weak; together, however, they form a fairly strong argument against platonism: by assuming platonism is false we can get the rewards of platonism (an account of mathematical truth) while assuming a simpler theory of reference, a simpler ontology and also explaining facts for which platonims does not have satisfying explanations.

(5), however, indicates that we have an embarrassment of riches in that we have two theories which are opposed to the platonistic theory which construes mathematical constants as having referents. Is there anything to choose between interpreting, say, arithmetic as an IAIT and interpreting it as an IMAIT?

The only significant difference between these two accounts which we have seen concerns ontological commitments. For example, arithmetic construed as an IAIT is committed to there being infinitely many things. Furthermore, embedding arithmetic (so interpreted) into a theory which does not have a commitment to there being abstract things may yield a theory which is committed to there being abstract things.

Now arithmetic is supposed to be not only true, but necessarily true. What reason is there to think that there are (respectively, must be) infinitely many things? So long as we maintain the view that numerals have referents, there is a straightforward answer to this: '(Ex)(x=0)' is true (necessarily true); furthermore it says that zero exists, so zero and all the other numbers do (must) exist.

In other words so long as we hold a platonistic view of mathematical language there is reason to believe that there are (must be) infinitely many things--the infinitely many mathematical entities. It is part of construing arithmetic as an IAIT, however, that there are no special mathematical entities. That is, anything can be in the domain of an acceptable structure for arithmetic. Thus, unless we can find some independent reason to think that there are (must be) infinitely many things, construing arithmetic as an IAIT will have the undesirable consequence that there will be no reason to think that arithmetic is true (necessarily true). Of course, when we turn to set theory, as was seen in Chapter 3, the ontological bill may go up even farther. This seems to be a significant problem for interpreting mathematical theories as IAIT's. In the absence of a reason for thinking that there are (must be) infinitely many things, I suggest that we construe mathematical theories as IMAIT's.

REFERENCES

[1] Benacerraf, Paul, "Mathematical Truth," Journal of Philosophy 70 (1973), pp. 661-679.

[2] ______. "What Numbers Could Not Be," Philosophical Review, LXXIV (1965), pp. 47-73.

[3] Cartwright, Richard, "Ontology and the Theory of Meaning," Philosophy of Science 21 (1954), pp. 316-25.

[4] Chowsky, Noan, and Israel Scheffler, "What is Said To Be," Proceedings of the Aristotelian Society 59 (1958-9), pp. 71-82.

[5] Descartes, Rene, Oeuvres de Descartes, ed. by Charles Adam and Paul Tannery. Paris: Leopold Cerf, 1897-1913.

[6] Eberle, Rolf, Nominalistic Systems, Dordrecht: D. Reidel, 1974.

[7] Field, Hartry, "Quine and the Correspondence Theory," Philosophical Review 83 (1974), pp. 200-28.

[8] Frege, Gottlob, The Foundations of Arithmetic, Translated by J. L. Austin. New York: Philosophical Library, 1950.

[9] Godel, Kurt, "What is Cantor's Continuum Hypothesis?" in Paul Benacerraf and Hilary Putnam, Philosophy of Mathematics: Selected Readings, Englewood Cliffs, New Jersey: Prentice Hall, 1964, pp. 258-273.

[10] Hume, David, A Treatise of Human Nature ed. by L. A. Selby-Bigge, Oxford: Clarendon Press, 1888.

[11] Jubien, Michael, "The Intensionality of Ontological Commitment," Nous 6 (1972), pp. 378-87.

[12] ______. "Ontological Commitment to Kinds," Synthese 31 (1975), pp. 85-106.

[13] ______. "Ontological Commitment to Particulars," Synthese 28 (1974), pp. 513-31.

[14] ______. "Ontolcgy and Mathematical Truth," xeroxed; forthcoming in Nous.

[15] Kripke, Saul. "Semantical Considerations on Modal Logic," Acta Philosophica Ferrica 16 (1963), pp. 83-94.

[16] Leibniz, G. W. New Essays Concerning Human Understanding translated and edited by A. G. Langley, Chicago: The Open Court Publishing Company, 1916.

[17] Locke, John. An Essay Concerning Human Understanding ed. by Peter H. Nidditch, Oxford: Clarendon Press, 1975.

[18] Parsons, Terence. "Various Extensional Notions of Ontological Commitment," Philosophical Studies 21 (1970), pp. 65-74.

[19] ________. "Extensional Theories of Ontological Commitment," Journal of Philosophy 64 (1967), pp. 44-50.

[20] Plato, Meno.

[21] ________. Phaedo.

[22] Putnam, Hilary. "Mathematics Without Foundations," Journal of Philosophy LXIV (1967), pp. 5-22.

[23] Quine, W. V. O. "Notes on the Theory of Reference," in From a Logical Point of View, Cambridge, Mass.: Harvard University Press, 1953, pp. 130-8.

[24] ________. "On Corrap's Views on Ontology," Philosophical Studies 2 (1951), pp. 65-72.

[25] ________. Ontological Reduction and the World of Numbers," Journal of Philosophy 61 (1964), pp. 209-216.

[26] ________. Ontological Relativity and Other Essays, New York: Columbia University Press, 1969.

[27] ________. "Ontology and Ideology," Philosophical Studies 2 (1951), pp. 11-15.

[28] ________. The Roots of Reference, La Salle, Ill.: Open Court, 1974.

[29] ________. Word and Object, Cambridge and New York: Technology Press of M. I. T. and Riley & Sons, 1960.

[30] Russell, Bertrand, Problems of Philosophy, New York: Oxford University Press, 1912.

[31] Shoenfield, J. R. Mathematical Logic, Menlo Park: Addison-Wesley, 1967.

[32] Steiner, Mark, "Platonism and the Causal Theory of Knowledge," Journal of Philosophy 70 (1973), pp. 57-66.

[33] Stich, Steven, ed. Innate Ideas, University of California Press, 1975.

[34] White, Nicholas. "What Numbers Are," Synthese 27 (1974), pp. 111-24.