THE UNIVERSITY OF KEELE

SET THEORY AND TRUTH

being a Thesis submitted for the Degree of

Doctor of Philosophy

in the University of Keele

by

A.J. Dale, M.A.

April 1969

CONTENTS

Introduction

This thesis covers a number of related topics in the foundations of set theory. Section 1 contains an analysis of the paradoxes which suggests a way of looking at the axioms of any formalised set theory as to a certain extent legislating how 'is a member of' and 'set' are to be used.

In section 2, two other points of view are examined and rejected. There are doubtless other views which directly contradict the view put forward in section 1, but clearly one cannot examine all the different solutions of the paradoxes in any one thesis. The positions examined in section 2 seem to represent views that are most opposed to the general tendency of this thesis.

Section 3 consists of a critical examination of Tarski's work on the semantic conception of truth for a formalised calculus of classes. This section has two aspects. The first is a defence of Tarski against some of his critics; the second an attack on certain of Tarski's conclusions.

Section 4 begins where section 3 leaves off. That is, it examines Tarski's conclusions about a definition of truth for set theory. It contains an analysis of Gödel's results on the incompleteness of formal systems and is particularly concerned with the transference of certain inferences made from Gödel's theorem for a formalised arithmetic to a formalised set theory. The conclusion of section 4 returns to the theme of section 1 and utilises certain metatheorems recently proved about formalised set theories.

Throughout the thesis no one axiom system of set theory is presupposed though reference will be made to several. The discussion is of a general nature and can be taken to be about any axiom system of set theory rather than some particular one.

Section 1

I

In this section I shall be concerned with the set-theoretic paradoxes. I wish to propose a way of looking at them which, if accepted, should alleviate much of the discomforture felt by philosophers, logicians and mathematicians when confronted by them. I speak of 'a way of looking at the paradoxes' rather than of 'a solution of the paradoxes' because, as will be made clear in the sequel, the way of looking at the paradoxes presented here allows of a multiplicity of 'solutions'.

It should be said here that, although much of the philosophical perplexity caused by the existence of the paradoxes may be dispelled, there will remain several problems for the mathematician and the philosopher, but these problems are not caused by the existence of contradictions. They would have arisen even without the discovery of the paradoxes. Indeed, there are two distinct problems which will remain to be solved: one of them a purely technical problem of direct concern to mathematicians only, the other a general philosophical problem analogous to the philosophical problems aroused by other mathematical and scientific theories.

The former problem is to construct a set-theory which is consistent and adequate for the needs of mathematicians working in theories which employ the concept of set. The problem of consistency has itself engendered a body of literature and it now seems that a proof of consistency for set-theory is unlikely to be forthcoming. But this is a technical question; it is sufficient for my present purposes to show that the problem of consistency would still be present even if there were no paradoxes. Certainly, the paradoxes have made the problem of consistency more urgent, because they have shown that inconsistencies can occur in the least suspected places. But the problem of consistency, at least for formal axiomatics, exists not because inconsistencies have occurred but because they might occur. The adequacy of a set-theory, referred to at the beginning of the paragraph, is needed because the mathematical theories which employ the concept of set, for example, Lebesgue measure and integration theory and the theory of real and complex functions which depends on the theory of sets of points, employ theorems of set theory. The mathematician working in such fields requires the theorems of set theory and, therefore, a set-theory which will provide him with these theorems. The related problems of the consistency and adequacy of set-theory are not then directly caused by the existence of the paradoxes and they will remain whatever philosophical solution of the paradoxes is offered They are essentially mathematical problems and can be solved only by mathematicians; they are only of indirect interest to the philosopher because they are not philosophical problems.

The general philosophical problem, mentioned on the previous page relates to the existence of sets. It is the problem aroused by the question 'Do sets exist?' or 'in what sense can sets be said to exist?' rather than the question 'What sets exist?' This is indeed a philosophical question but not one which a solution of the paradoxes will answer. The paradoxes help to give a partial answer to the question of which sets exist but not of whether sets exist.

II

In this chapter I shall state some of those paradoxes with which I shall deal. The paradoxes are all from set-theory and, although I shall speak of the other paradoxes, 'heterological' for example, in section 2, I shall not deal with them directly.

The list is not intended to be exhaustive. I have picked out those that I shall discuss in later chapters, but the treatment I propose should be capable of extension to other paradoxes of set-theory with which I shall not deal in detail, the paradox of all grounded classes, for example.

There remains the difficulty of characterising the paradoxes of set-theory. I think that Ramsey's[1] division of the paradoxes into two groups, the logical paradoxes and the 'epistemological' paradoxes (now generally referred to as the 'semantical paradoxes') will be adequate. Perhaps a more precise distinction may be made in the light of more recent work on semantics and also the distinction, now universally accepted, between object-language and meta-language. The set-theoretic paradoxes may then be characterised as those paradoxes which may be stated in the object-language of set-theory. With this characterisation the Berry, Richard

---

1. Ramsey, F.P. 'The Foundations of Mathematics', in The Foundations of Mathematics and Other Essays, London 1931

and Zermelo-König paradoxes fall within the domain of the semantical paradoxes because they each refer to an object-language as well as to sets. The Skolem-Löwenheim 'paradox', although a theorem belonging to the meta-theory of formal languages, is sometimes listed as a paradox along with the above.[1] Even if this important theorem is regarded as a paradox, it will still fall outside the scope of this section because it also belongs to the semantic category.[2]

Throughout this section and section2, then, I shall be discussing in some detail the following four paradoxes, bearing in mind that the procedure I shall advocate may be extended to the other paradoxes in the same category.

1. The Russell Paradox. Consider the set R of all those sets that are not members of themselves. If R is a member of R then R is not a member of R; if R is not a member of R then R is a member of R. Therefore, R is a member of R if and only if R is not a member of R. Assuming the law of excluded middle it follows that R is a member of R and R is not a member of R.

2. The Cantor Paradox. Consider C the set of all sets,

1. See, for example: E. Beth, The Foundations of Mathematics Amsterdam, 1959. pp.448-450

2. This will be discussed further in section 4 of this thesis.

and the set UC of all subsets of C. It follows from a general theorem of set theory, namely, for any set the cardinal number of Us (the set of all subsets of s) is greater than the cardinal number of the set s, that the cardinal number of UC is greater than the cardinal of C. Since for all x if x is a member of UC x is a member of C, UC is a subset of C, it follows from another theorem of set-theory that the cardinal number of UC is less than or equal to the cardinal number of C. Therefore C has a cardinal number greater than or equal to the cardinal number of C and the cardinal number of C is also less than the cardinal number of UC. This is a contradiction.

3. <u>The Set of all Cardinals</u>. Consider the set of all cardinals. One theorem in set theory states that there is no greatest cardinal and another theorem that for any set of cardinals among which there is no greatest member the sum of the cardinals of the set is greater than any cardinal in the set. Therefore, the set of all cardinals, which has no greatest member, gives a sum which is greater than any of the cardinals in the set, i.e. a cardinal greater than any cardinal.

4. <u>The Burali-Forti Paradox</u>. Consider the set of all ordinal numbers arranged in order of magnitude. This set is well-ordered. Suppose its ordinal number is Ω . Consider

the set of all ordinals up to and including Ω arranged in order of magnitude. The ordinal of this set will be $\Omega + 1$. Now Ω is less than $\Omega + 1$. Since the set of all ordinals up to and including Ω is an initial segment of the set of all ordinals, $\Omega + 1$ is less than or equal to Ω. This is a contradiction.

III

To understand how the paradoxes have prevented us from seeing their solution, it will be necessary to pay more attention to the way they are stated in important works on logic or the foundations of mathematics and in many of the text-books dealing with these subjects. For example, to quote only a few of the many different statements of the paradoxes to be found in such books:

"Let w be the class of all those classes which are not members of themselves. Then, whatever class x may be, 'x is a w' is equivalent to 'x is not an x'. Hence, giving to x the value w, 'w is a w' is equivalent to 'w is not a w'.[1]

"Consider the set of all sets; call it M."[2]

"Let us suppose that S is the set of all sets"[3]

".... the set of all subsets of a set M has a cardinal number higher than that of M. This is a contradiction if M is the set of all sets."[4]

These four statements or partial statements of the paradoxes as well as the statements of the paradoxes as I gave them in the previous chapter have helped to conceal, behind the words "consider" or "let us suppose" a 'hidden' existential proposition. (A notable exception to this indirect concealment occurs in Fraenkel and

---

1. B. Russell and A. Whitehead, Principia Mathematica, 2nd ed. Cambridge 1927, p.60
2. S. Kleene, Introduction to Metamathhematics, Amsterdam, 1952, p.36
3. W & M Kneale, The Development of Logic, Oxford, 1962, p.652
4. H. Curry, Foundations of Mathematical Logic, New York, 1963 p.5

Bar-Hillel's discussions of the Russell paradox.)[1]

A proof that there does not exist a last prime number can, if it is formulated in an analogous way, be turned into a 'proof' that there are inconsistencies in number theory. For example, instead of the phrase 'suppose there exists a last prime number' the 'proof' would start 'consider the last prime number, call it P'. From 'consider the number P such that P is prime and, for all n, if n is greater than P then there exists an x such that $x \neq 1$, $x \neq n$ and x divides n' it may be deduced that there is and there is not a number which is prime and greater than P. This is a contradiction.

It can be seen that such a proof would never be accepted by mathematicians because the proof has concealed the existential assumption that there exists a last prime number. It is valid only if there does exist such a number. But that there does not exist such a number only follows from the fact that a contradiction has been derived from the supposition that it does exist.

Now if the same reasoning is applied to the set of all sets or the set of all cardinals, it can be seen that by rewriting the offending phrases 'consider the set ...' or 'let the set R be ...' in the proper existential form 'suppose there exists a

---

1. A. Fraenkel and Y. Barr-Hillel, Foundations of Set-Theory, Amsterdam, 1958, p.6

set, S say, such that ...' what might be said to follow from the resulting inconsistency is not that there is some paradox that must be removed but that there is no such set as the set S. In the following chapters I shall be considering the merits of this argument and, also, what qualifications have to be put on it, since in the form given above there is much oversimplification.

IV

In this chapter I shall show that there are analogies to the Russell paradox in established fields of mathematics. I shall consider three examples of existence theorems.

1. There is no last prime number.

2. There exists a non-enumerable set. (To put this in a form more analogous to the paradoxes, there does not exist a one-to-one correspondence between the set of all sets of natural numbers and the set of all natural numbers or a subset of them).

3. There do not exist natural numbers p and q such that p/q is equal to the square root of 2.

Each of these theorems bears a resemblence to the solution put forward here to the Russell paradox. In the case of 1. there is no difficulty; it is an accepted theorem of number theory and has been so at least from the time of Euclid. In the case of 2. opinion is still divided. Arguments have been proposed, notably by the intuitionist school, for its rejection. In the case of 3., although it is an accepted theorem of analysis all outstanding difficulties have been cleared up only in the last century.

All are analogous to the statement 'there does not exist the class of all classes which contain themselves as members.' All <u>could</u> be regarded as paradoxes if we refuse to accept that they are results established by the use of reductio ad absurdum proofs. There seems little more than prejudice which would account for the attitude taken with regard to the Russell paradox on the one hand and 1., 2. and 3. on the other. Admittedly, that there is no such class as the Russell class may be surprising but this should be no criterion for rejecting that result. To some it may be just as surprising that there does exist a class which cannot be put into one-one corrdspondence with a subset of the natural numbers. The discovery that there exist irrational numbers must have surprised the Pythagoreans. In these latter cases, however, a new fruitful mathematics has come into being. In the first, the theory of transfinite cardinals and ordinals; in the second the theory of irrational numbers.

In other words, the discovery of 2. and 3. have altered fundamental assumptions held about numbers and sets. We have not been content to say here is a paradox but we have been prepared to alter our concept of number.[1] It would seem then that we should do the same for the set-theoretical results. We should not say here are some paradoxes, but say rather our concept of set must be altered according to the results we have.

1. The Pythagoreans held that lines were made up of an integral number of units. This, however, was found to be incompatible with the consequences of Pythagoras's theorem. Instead of introducing the notion of an irrational number, Greek mathematicians were forced to abandon the attempt to identify the realm of number with continuous magnitudes . C. Boyer The Concepts of the Calculus 1949 p.20

We may view the paradoxes not as inconsistencies in set-theory but as part of unfinished proofs that certain existential assumptions are false. The argument which, according to this theory, should be applied in the case where two contradictory propositions are both derivable from some assumption, q say, is that not-q is provable. This argument, frequently employed in mathematics, is simply an example of _reductio ad absurdum_. This view has some precursers. Solutions along these lines have been proposed by D. Bochvar[1], J.F. Thomson[2] and G.H. Von Wright[3].

Bochvar contends that the set-theoretic paradoxes result from definitions which include or presuppose existential assumptions of an extra-logical character. In particular, the axiom schema

$$\text{(A)} \quad (Ex_{p+1})(x_1)(x_2)\ldots\ldots(x_p)\{x_{p+1}(x_1,x_2,\ldots\ldots,x_p) \equiv U(x_1,x_2,\ldots x_p)\}$$

where U is any expression containing the free variables $x_1,x_2,\ldots\ldots,x_p$ is responsible for the existence of the Russell paradox. The logical system Bochvar constructs is a version of elementary logic with variables, not subjected to a type hierarchy

---

1. D. Bochvar, 'To the Question of Paradoxes of the Mathematical Logic and Theory of Sets', _Mat.Sbornik_ 15, 365-384. Known to me through the review by Wanda Sxmielew, 1946, _Journal of Symbolic Logic_, 11, p.129 and E. Beth, _The Foundations of Mathematics_, Amseterdam 1959, p.506
2. J.F. Thomson. 'On Some Paradoxes' pp.104-119
3. G.H. von Wright, 'The Heterological Paradox', _Societas Scientiarum Fennica Commentationes Physico-Mathematicae_ XXIV 5, 1960 pp.1-28

x_1, x_2,...., atoms $x_n(x_{n_1}, x_{n_2}, \ldots, x_{n_p})$ and excluding every application of the schema (A). This system is shown by Bochvar to be consistent. Instead of the Russell paradox being derivable from this system, it is provable that there does not exist the set of all sets which do not belong to themselves. More precisely, the sentence $\sim(Ex_2)(x_1)(x_2(x_1) \equiv \sim x_1(x_1))$ is provable. The solution that Bochvar proposes seems to depend upon the difference between logical assumptions and extra-logical assumptions. According to Bochvar the schema (A) is an extra-logical assumption which is responsible for the appearance of the paradoxes. Clearly, if he is right in his contention that the paradoxes do result from such extra-logical assumptions he needs some criterion by which to determine which sentences are of logical nature and which extra-logical. The question that arises from this is which assumptions are of a purely logical character. If there is to be a set-theory at all, there needs to be certain axioms in any set-theory from which set-theoretic theorems follow. Are these axioms of a purely logical nature?

It is difficult to see[1] the set-theory which Bochvar has proved to be consistent and what theorems of a generally accepted set-theory remain in such a system as Bochvar's. What is certain, however, is that Bochvar proved, in his system of set-theory at least, that there does not exist the set of all sets which are members of themselves. There are in various standard works on logic and set-theory similar results. For example, Quine's system, referred to as M.L. contains the theorem that there is no such set as the set of all sets which do not contain themselves as members,[2] and also Fraenkel in Abstract Set Theory in connection with the Burali-forti

---

1. The fault is not Bochvar's but mine, because I am dependent upon the review by Sxmielew (see note 1 p. 13 of this chapter). It is clear from that review that Bochvar excludes all existential assumptions since he regards them as not belonging to the province of pure logic The calculus he constructs, K_o, is a form of the first-order functional calculus with identity. Thus he proves K_o to be consistent whereas the extended functional calculus without a theory of types is known to be inconsistent. K_o is also a form of the extended predicate calculus without the whole of the 'extra-logical' part. It is not clear on what grounds he rejects these existential assumptions other than that they give rise to paradoxes, nor why he labels them extra-logical. If, as Sxmielew implies, he believes that theorems of existential character and the study of the relations of this character between things is not proper to logic even when they are expressible in logical terms, then what does he say of the existential theorems of the first-order functional calculus, $(Ex)Px \vee \sim Px$? If these are theorems of K_o then it remains to show that quantifying with the existential quantifier over individual variable is a part of 'pure' logic but quantifying over predicates does not belong to 'pure' logic.

2. W. Quine, Mathematical Logic, 1940, pp.128-9

antinomy states:

'the totality of all ordinals does not constitute a set'.[1]

But these latter theorems are not regarded by their authors as offering a philosophical explanation of the paradoxes.[2] Bochvar, on the other hand, does not limit himself to obtaining a theorem which is a consequence of the axioms and rules of derivation of a formal calculus only, but contends that the paradoxes are contradictions resulting from the intrusion of extra-logical existential assumptions which are instances of the schema (A)1

J.F. Thomson[3] argues that the 'Barber' paradox, the heterological paradox, the Richard paradox and the Russell paradox have a common form. He proves the theorem that if S is any set and R any relation defined at least on S then no element of S has R to all and only those S-elements which do not have R to themselves. In itself this is not paradoxical but 'a plain and simple logical truth'[4] which, however, provides a foundation on which many of the paradoxes are built.

The answer to the Barber paradox, based upon the theorem is that no man exists who shaves all and only those men who do

---

1. A. Fraenkel, Abstract Set Theory, Amsterdam 1961, pp.201-2
2. Fraenkel does not consider that this explains the paradox but uses the concept of classes which are not eligible for membership rather as an expedient. See his discussion op.cit. p.202. Quine also uses membership-eligibility, op.cit. p.131
3. J.F. Thomson, op.cit.
4. ibid. p.104

not shave themselves. Indeed, this is the answer which is accepted. Yet, the heterological paradox and the Russell paradox although based upon the same theorem have not had the same conclusion drawn from them. Thomson argues that the heterological and Russell paradoxes should be regarded in the same light as the Barber and that:

'a contradiction arises on supposing that there could be an adjective which is true of absolutely every adjective which is false of itself. That is, this supposition is absurd and must be given up.'[1]

'Formally, this the reasoning that the Russell set is a member of itself if and only if it is not a member of itself is just the same argument as that of the Barber: so why should we not deal with it in just the same way, and say it just shows that there is no such set as R?

The answer .... is that we should deny that there is such a set as R the set which produces the Russell paradox '.[2]

Whilst I agree, with qualifications, with Thomson on these conslusions, the schema he gives upon which the paradoxes can be based tends to conceal that <u>all</u> the set-theoretic paradoxes have a common structure, though not that of his schema. Certainly the great similarity between the Barber paradox and the

1. ibid. p.112
2. ibid. p.117

Russell paradox is brought out very clearly, but the schema is no help when we come to the Cantor paradox or the Burali-forti paradox. Here it would be necessary to go even deeper to find the common proof schema which would show that there is no such set as the set of all sets or the set of all ordinals. The common schema in all these cases is the theorem from the propositional calculus: $\{p \supset (q.\sim q)\} \supset \sim p$. Incidentally, it is this schema which is used in Thomson's proof of his theorem. The main criticism that I have to make of Thomson is that he does not go far enough.[1]

G.H. von Wright discusses the heterological paradox[2] and comes to a similar conclusion, namely, that 'heterological' does not name a property which a thing has if and only if it is not autological, or, to reformulate this proposition, heterological is not a property. Since heterological is not a property and because the definition of 'heterological' states that x is heterological if and only if it is not the case that x has a property of which x is a name, it follows that 'heterological' is heterological. It does not follow that because 'heterological' is heterological, 'heterological' is not heterological, since 'heterological' is heterological because it is not a property,

---

1. Basically, Thomson is showing, though he does not say so, that Barber, Russell and Grelling paradox have the schema $(\exists x)\ (y)\ f(x,y) \equiv \sim f(y,y)$ whilst the negation of this formula is provable in the predicate calculus.
2. G.H. von Wright, op.cit.

not because it does not have the property which it names. The principal conclusion which von Wright says should be drawn is that 'heterological' is not a property.

I shall now mention two of his arguments which I shall refer to later. Firstly, he considers that objections may be made that 'heterological' must still be a property even though the modo tollente proof $\{p \supset (q \equiv \sim q)\} \supset \sim p$ shows that it is not. He maintains that it is necessary to clarify the concept of property understood by the objector. For example, it might be said that a property is anything which can function as a predicate in a true proposition of subject-predicate form. Since "'hexasyllabic' is heterological" expresses a true proposition of the subject-predicate form 'heterological' must be a property. But the sense of 'property' which von Wright understands is the sense implicitly defined by the predicate calculus which states $\sim \{f(x) . \sim f(x)\}$ as a theorem. It is in this sense that 'heterological' is not a property. If it is maintained that 'property' should be understood in the objector's sense, then there is no contradiction or paradox because the predications involved would not be predications in the sense of the predicate calculus.

Secondly, he gives an analogy between the heterological paradox and the division by 0 in arithmetic. If for any real numbers m, k and l, ml=mk implied l=k, it could be proved that 5=7, since 0.5=0.7 This contradiction could lead to the conclusion that 0 is not a real number. It is more useful, however, to admit 0 as a real number than to reject it. Instead of rejecting 0 from the class of real numbers, it is preferred to say that there is an exception to the proposition that for all real number m,k,l, if mk=ml then k=l and the proposition is modified accordingly. The proposition becomes, for all real numbers m,k,l, if m≠0 and mk=ml then k=l.

An analogy exists between the case of division by 0 in arithmetic and the case of heterologicality. Von Wright contends that if the evidence in favour of calling heterological a property outweighs the evidence against it (in this case, the derivation of a contradiction from the supposition that it is a property), then it could be said that 'for any word x, if x is not a name of the property of heterologicality itself, then x is heterological if and only if it is not the case that x has got a property, of which x is a name.'[1]

I shall return to these two arguments later in this section. In the next chapter I shall show how similar reasoning may be applied to the paradoxes of set-theory, a subject only mentioned in passing by von Wright.

1. G.H. von Wright, op.cit. p.27

V

In this chapter I shall be looking at the paradoxes in more detail and showing the implications of viewing them as part of reductio ad absurdum proofs.

In chapter III of this section I maintained that the apparent paradoxes were only partial proofs of set-theoretic theorems. They served the same function as a contradiction in any reductio ad absurdum argument, namely, to negate the premise from which the contradiction was derived. Thus, from the assumption that there exists a class of all classes which are not members of themselves, the Russell paradox proves that there does not exist such a class. In other words, the paradoxes are the penultimate inference steps of theorems. In such a manner the paradoxes are removed and 'new' theorems take their place in set-theory. The new theorems, which replace the four paradoxes taken as examples of the paradoxes in general in chapter I of the present section, are: 1) there exists no class R such that, for all x, x is a member of R if and only if x is not a member of x; 2) there is no class C such that, for all x, x is a member of C if and only if x is a class; 3) there is no class S such that, for all x, x is a member of S if and only if x is a cardinal number; 4) there is no ordered class T of all ordinal numbers, ordered according to magnitude.

Such theorems as these represent gross simplifications of what would happen in the case of such a 'solution' being applied; Above, it is applied directly to the 'hidden' premises. In practice, however, the implications are far more complex. In the case of the Russell paradox, I believe a good case can be maintained that no such class exists, reasoning on the reductio ad absurdum argument outlined. The other paradoxes require a more subtle treatment because they are embeddied rather deeper in set-theory. Although I have not been concerned with any axiomatic or formal system of set-theory in this present section and have treated the paradoxes and purported solutions as informally as possible, it will be necessary to give a more detailed analysis of the Cantor paradox and to give the proof of the theorem on which it depends. The proof is informal and is not derived from any axiom set in particular.[1] The proof, and the remarks that I shall make on it, should help to answer the question of why from the contradiction involved in the Cantor paradox, for example, it does not necessarily follow immediately that no such class as the class of all classes exists. This would seem to contradict what I have said above, but it will

1. For a treatment of set-theory in an informal manner see Sierpinski <u>Cardinal and Ordinal Numbers</u>, Warsaw, 1958

be seen that there may be more than one premise involved and that the Russell and Cantor paradoxes may not be as independent as they seem.

Cantor's paradox follows from the theorem that for all sets X, the cardinal number of X, denoted by "$\bar{\bar{X}}$", is less than the cardinal number of $\breve{U}X$, the set of all subsets of the set X. I shall give one proof of this important theorem. For the definitions of equality and order amongst the cardinals I shall take the following:

1) $\bar{\bar{X}} = \bar{\bar{Y}}$ if and only if $X \sim Y_0 \subseteq Y$ and $Y \sim X_0 \subseteq X$[1].

2) $\bar{\bar{X}} < \bar{\bar{Y}}$ if and only if $X \sim Y_0 \subseteq Y$ and for all X_0 if $X_0 \subseteq X$ then $Y \nsim X_0$

<u>Proof</u>: Each x of X can be associated with $\{x\}$ of UX (where "$\{x\}$" denotes the set of which x is the sole member).

Hence $X \sim X_0 \subseteq UX$

(X_0 being the set of unit subsets of X)

By 1) and 2) $\bar{\bar{X}} < \bar{\bar{UX}}$ or $\bar{\bar{X}} = \bar{\bar{UX}}$

Suppose $\bar{\bar{X}} = \bar{\bar{UX}}$ i.e. $UX \sim X_1 \subseteq X$, for some X_1

$\therefore \exists \psi$, a 1-1 correspondence, such that to each $x \in X_1$, $\psi(x) = X_s$, where X_s is a certain subset of X.

3) Let R' be the set of all x which are members of X_1 and are not members of $\psi(x)$

i.e. $R' = \{x;\ x \in X_1 \,.\, x \notin \psi(x)\}$

---

1. 1) is not the usual definition of equality between cardinals, which is $\bar{\bar{X}}=\bar{\bar{Y}}$ if and only if $X \sim Y$, but due to the equivalence theorem of set-theory, 1) is equi-pollent with it. See A. Fraenkel, <u>Abstract Set Theory</u> Amsterdam, 1961, pp.58-78.

Now R' is a sub-class of X

$\therefore R \in UX$

$\therefore (\exists y)\,[\psi(y) = R' \,.\, y \in X_1]$

Now, $y \in \psi(y)$ if and only if $y \in R'$

" " " " " $y \in \{x;\ x \in X_1 .\ x \notin \psi(x)\}$

" " " " " $y \in \{X_1 \,.\, y \notin \psi(y)$

since $y \in X_1$,

$y \in \psi(y)$ if and only if $y \notin \psi(y)$

which is a contradiction.

Hence, by reductio ad absurdum, $\bar{\bar{X}} \neq \bar{\bar{UX}}$

$\therefore$ (4) $\bar{\bar{X}} < \bar{\bar{UX}}$

This theorem provides half the basis of the Cantor paradox where the set in question is the class of all classes, C.

i.e. (5) $\bar{\bar{C}} < \bar{\bar{UC}}$

The other half is provided by the fact that, in the case of C, $UC \subseteq C$. (since all members of UC are sets, all members of UC are members of C, $UC \subseteq C$ follows from the definition of subset). By (1) and (2) $\bar{\bar{UC}} \leq \bar{\bar{C}}$ which contradicts (5).

A less precise statement of Cantor's paradox is that <u>obviously</u> the cardinal number of the set of all sets is the highest that can exist, yet the theorem proved above shows that the set of all subset of this set must be greater still.[1]

1. See, for example, E. Beth, <u>The Foundations of Mathematics</u> Amsterdam, 1959, p.484.

If, for the moment, the paradoxes and the above theorem are forgotten and fresh attention paid to the class of all classes, a new relationship may be found between it and the class of all its subslasses. On an intuitive level, then, the class of all classes does have the largest cardinal number, but, as yet, nothing is known of the cardinal number of the class of all its subclasses. One would expect that, as its cardinal number cannot be higher than the cardinal number of the class of all classes, and, on the other hand, as the number of its members cannot be less than the number of members of the class of all classes, its cardinal number should equal the cardinal number of the class of all classes. Continuing to disregard the above theorem, this can be 'proved' as follows.

Let C be the class of all classes,

Let UC be the class of all subclasses of C

(6) A 1-1 correspondence, η, can be set up between C and a subclass of C_o of UC, in this way:

For each $x \in C$ let $\eta(x) = \{x\}$ i.e. the unit class consisting of x alone.

(7) Also, a 1-1 correspondnece, θ, can be set up between UC and C_* a sub-class of C, in this way:

For each $x \in UC$ associate $x \in C$

i.e. $\theta(x) = x$

(6) and (7) together imply that $C \sim C_0 \subseteq UC$ and $UC \sim C_* \subseteq C$ which implies, by definition (1), that $\bar{\bar{C}} = \bar{\bar{UC}}$ (8)

The last result (8) is not surprising. It is very much as one would expect, if the theorem (4) was ignored. Yet, (8) is in direct contradiction to (5). In other words, (8) and (5) restate Cantor's paradox. To prove (8) a 1-1 correspondence was established between C_* and UC, but according to the proof of (4) no such 1-1 correspondence can be established. If, for the moment, one accepts (8) then there must be a fallacy in the proof of (4). This proof will be more thoroughly examined to see how it comes into conflict with (8).

That part of the proof which used reductio ad absurdum reasoning began with the supposition that for some X_1, $UX \sim X_1 \subseteq X$. In the proof of (8) a 1-1 correspondence, θ, was established by which $UC \sim C_* \subseteq C$. It is no longer just a supposition that there exists such an X related to UX under a 1-1 correspondence, for C_* and UC are related by θ in exactly this way. The contradiction which followed in the proof of (4) should, therefore, follow when C is substituted for X and θ for ψ in that proof.

The first step (3) was to let R' be the class of all x which are members of X_1 and not members of $\psi(x)$. In the case of the 1-1 correspondence θ and the class C_*, this becomes; let R' be the class of all x which are members of C_* and not members of $\theta(x)$. I.e. $R' = \{x;\ x \epsilon C_*.\ x \notin \theta(x)\}$.

Using the definition of θ (x), this becomes: let $R' = \{x;\ x \epsilon C_*.\ x \notin x\}$. Now, this class R' is a subclass of the class of all classes which are not members of themselves. In other words, R' is a suspect class already since it is a certain sub-class of the Russell class R. The contradiction which follows in (4) is transformed into an argument analogous to the argument leading to the Russell paradox by the substitutions of C and θ:

Since R' is a certain sub-class of C, $R' \epsilon$ UC.

Hence, there exists y such that $y \epsilon C_*$ and θ $(y) = R'$.

From the definition of θ (y), $y = R'$.

Therefore, $y \epsilon \theta$ (y) if and only if $y \epsilon R'$, hence, if and only if $y \notin \theta(y)$.

i.e. $R' \epsilon R'$ if and only if $R' \epsilon R'$.

In the proof of (4) the contradiction led to the rejection of the supposition that there could be an equivalence between X_1 and UX, but it was taken for granted that R' would exist in the formulation of (3). If, however, the existence of R' is not assumed, the contradiction could equally well prove that R' does not exist. In the case of the class of all classes and the 1-1 correspondence set up between UC and C_*, the supposition that there could exist such a correspondence between a subclass of a class and the class of all its subclasses is no longer just a supposition. The supposition that there exists such an R' must be rejected if the 1-1 correspondence θ is accepted.

The point I have been making in the detailed analysis of the proof of (4) is that the existence of Cantor's paradox does not prove that there does not exist a class of all classes.

In the case of the Russell paradox, I have applied the method of reductio ad absurdum directly to the Russell class. (The reason for so doing I shall explain later). But with Cantor's paradox, the situation is different. For the paradox to occur, there must be two classes, the existence of which is assumed, namely, the class of all classes (together with the class of all its sub-classes) and the class R', the class of all those classes which do not belong to themselves and also belong to C_*. As I have shown, the Cantor paradox occurs because the existence of C is incompatible with the existence of R'. Therefore, the reductio method could be used to show that R' does not exist. As R' was a suspect class in any case (since it was a sub-class of the Russell class) this would not be so surprising. Two 'new' theorems would then be established. Firstly, $\bar{\bar{C}}=\overline{\overline{UC}}$ and, secondly, for all classes X <u>except the class of all classes</u> $\bar{\bar{X}} < \overline{\overline{UX}}$, analogous to the treatment given to division by zero in arithmetic. Again, this leads to complications. In the first place, $\bar{\bar{C}}=\bar{\bar{UC}}$ may still be inconsistent with some other result of set-theory; secondly, the supposition that C exists (and hence $\bar{\bar{C}}=\bar{\bar{UC}}$) implies the rejection of an infinite number of classes. This latter implication follows from the fact that there are an infinite number of 1-1 correspondences between UC and subsets of C. For example, θ can be taken to be the 1-1 correspondence which associates each x belonging to UC with $\{x\}$ belonging to C. This correspondence θ_1, say, thus establishes an equivalence between UC and C_1 a subclass of C. The class which then corresponds to R' will be the class of all those classes which belong to C_1 and do not belong to their only member. i.e. $R_1=\{x; \{x\} \in C_1 . \{x\} \notin x\}$. By the same reasoning as was used previously R_1 does not exist if C does. Similarly, by establishing the correspondence between each x of UC and $\{\{x\}\}$, it can be shown that $R_2=\{x; \{\{x\}\} \in C_2 . \{\{x\}\} \notin x\}$ does not exist if C does. By the 1-1 correspondence associating $\{\{\{x\}\}\}$ with x and $\{\{\{\{x\}\}\}\}$ with x etc., the classes R_3, R_4, etc. formed analogously to R', can be shown not to exist. If so many classes have to be rejected, it may be felt that it is the class of all classes which is the root of all the trouble and that it should be rejected rather than the classes R', R_1, R_2, etc. Nevertheless, the point is that R', R_1, ... <u>could</u> be rejected.

To summarise this chapter: although I said in chapter III that the paradoxes could be regarded as implying the non-existence of the sets that give rise to them, to do so would be to oversimplify the situation. For the rejection of one set may remove the necessity of rejecting another.

V

It might be said that it is a mistake to talk of preferring to reject one set rather than another. It might be thought that either there is such a set as the set of all sets or there is not. We are not free to choose whether a particular set exists. We can only discover that such a set exists. One might say that the paradoxes show that no such set as the Russell set exists, in the sense that we discover that the set does not exist. To talk in this way is to talk as though set-theory is a science investigating objects open to our inspection, rather as the physical sciences investigate the nature and behaviour of physical objects.

Now abstract set-theory, as opposed to theories of point sets, set of natural numbers etc., is, as its name suggests, a theory of abstract sets. Its universe of discourse is limited to sets. We are speaking in abstract set-theory of 'sets' rather than of 'sets of'. It is this change from the familiar to the unfamiliar which should make us look askance at the view that we are discovering laws of how sets behave.

Certainly, the familiar talk of sets of points or natural numbers guides us in how we shall talk of abstract sets. For what we want from a set-theory is a ready-made apparatus which will be applicable when we want to discuss sets of some particular kind.

The 'arithmetisation' of analysis is, as Wang[1] says, a misnomer. For, besides the theory of natural numbers, Cauchy convergent sequences and Dedekind cuts, in terms of either of which the real numbers can be defined, need infinite sets of natural numbers for the 'arithmetisation' to be carried out completely. What is needed, then, is a theory of sets which can be applied to natural numbers. i.e. a theory which will give the theorems which we need for the 'arithmetisation' of analysis when the 'sets' of the abstract theory are identified with sets of natural numbers, sets of sets of natural numbers, etc.

Perhaps Wang is wrong in saying that since real numbers can be regarded as sets of rational numbers the 'arithmetisation' logically calls for a general theory of sets.[2] What it does call for is only a theory of sets of rational numbers or sets of sets of such. I can see no logical reason why a general theory of abstract sets is needed, although we may feel more intellectually satisfied if we have such a theory. Wang's theory Σ with a bottom 'layer' of rational numbers would provide

---

1. H. Wang, 'The Formalisation of Mathematics', in <u>A Survey of Mathematical Logic</u>, Peking 1963, p.560
2. ibid. p.561

such a theory but even this theory goes beyond what is needed for founding analysis on the theory of natural numbers and sets of them etc.

Although we are not compelled to construct an abstract set-theory even for a successful reduction of analysis to the theory of natural numbers, nevertheless there is no reason why we should not do so in order to satisfy our intellectual curiosity. Besides, we do not want to construct a new theory each time we want to consider sets of another sort; sets of points, for example, in measure theory. To do so would be wasteful if we could find a theory which would be applicable in each case.

The abstract set-theory that we create as a result is founded on what we know from considering such sets as sets of natural numbers, sets of points, etc. This knowledge guides us in our choice of axioms. It does not, however, force us to adopt any particular axiom. As long as we choose axioms from which we can derive all the theorems that we need for the application of set-theory to some universe of discourse, we are free to choose what other axioms we like. (Consistency of these other axioms being the only limitation, since, otherwise the set-theory would have no applications.) This will be made clear in section 4.

If we look at the paradoxes in this light we should not be so puzzled by them. It should be remembered that the syntactic paradoxes have occurred only in abstract set-theory and no paradoxes have been found when considering sets of natural numbers, points etc. In such contexts the known paradoxes do not threaten, as Quine[1] and others[2] have noticed, since problems raised by '$x \epsilon x$' do not come up.

What Cantor tried to take as an axiom for this new theory of abstract sets is the axiom of comprehension in its naive form:

$$(Ex)(y)\ y \epsilon x \equiv F(y) \ ,$$

where 'F' is any condition whatever. When the variable y ranges over the members of some set, this axiom will not, taken as the only axiom of set-theory, give rise to any contradictions;[3] in other words, as Kreisel[4] remarks, when we think of the axiom as giving the existence of a set whose members are of a particular kind.

It is when the variable is not so restricted that trouble occurs. The Russell paradox follows immediately the moment we put in the specific condition '$\sim y \epsilon y$'. The problems arise when and only when we do not restrict the range of the variable toaa particular kind of object.

1. W. Quine, Set-Theory and its Logic, Cambridge, Mass., 1963 p.5
2. K. Gödel, 'What is Cantor's Continuum Problem?' American Mathematical Monthly, vol.54 1947
3. W. Quine, Set Theory and its logic p.37
4. G.Kreisel, 'Informal Rigour and Completeness Proofs' in Problems in the Philosophy of Mathematics, ed. I.Lakatos, Amsterdam, 1967, p.14 3

In abstract set-theory we are only interested in giving axiom for sets, considered as sets and not of sets <u>of</u> something, though the axioms must be interpretable for application as sets of something. In abstract set-theory there is only one primitive predicate, the membership relation. The axioms that we choose will determine the properties of this relation. The axioms determine how we are to use the phrase 'is a member of'. Initially we try to make these axioms tally with accepted uses of the phrase 'is a member of' as when we say that the number 5 is a member of the set of odd numbers or Jones ia a member of the class of unemployed.

In the field of abstract sets we first meet counter-instances of the axiom of comprehension. It appears that not every condition determines a set. For example '$x \notin x$' does not determine a set. There is no set which consists of those sets which do not belong to themselves. We cannot carry over, without inconsistency, the assumption that every condition determines a class of objects that satisfies it from, say, the universe of natural numbers or of human beings to the universe of abstract sets.

All that the paradoxes show is that the axiom of comprehension cannot be taken as one of the axioms of abstract set-theory. It is a situation which is analogous to other situations that have occurred in mathematics. The creation of imaginary numbers needed axioms which would be in accord with the axioms of the real numbers and such that the existence of a root of any integer was guaranteed by those axioms. It turns out that we cannot keep all the axioms of the real numbers, for it could be 'proved' that the square root of -1 is both less than and greater than 0. The axioms of the real numbers do not carry over to the universe of complex numbers. The axioms of order are rejected.

We are assured by the predicate calculus that there is nothing which bears the relation f to everything which does not bear the relation f to itself, whatever relation f may be. i.e. $\sim(\exists x)(y)\,[f(y,x) \equiv \sim f(y,y)]$ is a valid theorem of the calculus. So there can be no barber who shaves all and only those who do not shave themselves; there can be no set which contains all and only those sets which do not contain themselves. We discover that we cannot use the phrase 'is a member of' in the way we would have liked. The axioms for the universe of abstract sets cannot include the comprehension axiom.

It could be argued that we are not forced to give up the axiom of comprehension. Instead we could object to the phrase '$x \epsilon x$'. We could argue that this phrase is not meaningful. This is what Russell did. The axiom of comprehension is retained in the form of the axiom of reducibility and the phrase 'member but we can note here that Wang, who, perhaps more than any other logician, sympathises with some theory of types, says that Russell went needlessly far in maintaining that such expressions were meaningless.[1] Wang's system Σ presents an example of a theory where the objects are stratified in types but where it turns out that '$x \epsilon x$' is false and not meaningless.

If it should be found that Russell's reasons are not sufficient for the conclusion that certain phrases are meaningless, there remain powerful reasons for saying that the axiom of comprehension does not hold in the field of abstract sets. For whatever relation f may be, we can be sure from the predicate calculus alone that there can be nothing in the universe of discourse which has that relation to all and only those things that do not bear that relation to themselves. In chapter IV I mentioned that von Wright regards the predicate calculus as defining what a predicate is. To go back to his discussion of 'heterological' if we say analogously that there must be such a predicate as 'member of' and, at the same time, maintain that this predicate is such that there does exist an x such that $y \epsilon x$ if and only if $y \not\epsilon y$, then we must be understanding the term 'predicate' in a way which is in need of explanation.

We cannot in constructing a set-theory use the axiom of

1. H. Wang, 'The Formalisation of Mathematics', p.577

comprehension (if we regard '$x \epsilon x$' as meaningful). How are we to replace this axiom? Axioms of set existence are needed and the obvious candidate has failed. At the beginning of this section I said that it would be misleading to speak of there being just one solution of the paradoxes; it would be more correct to say that there are a multiplicity of different solutions.

If we were to ask two (classical) mathematicians what a real number is we might receive different answers. One might say that it was a set of rational numbers, theother that it was a set of sequences of rational numbers. It would depend on whether they accepted the Dedekind cut construction of the real numbers or the Cauchy construction. We might say that both answered the question of what a real number is, that both provided a solution to the problem.

Similarly, there are many axiomatic set-theories, differing greatly in the sets that the theories are committed to. Each can be regarded as substituting a number of existential axioms to replace the axiom of comprehension. There is no sense in asking which theory is the 'correct' theory, although we might worry that some system seemed inadequate for the applications we wish to make of it. By the adequacy of a set-theory I mean only that it guarantees the existence of any set which we need when we interpret the objects of the set-theory as 'sets of'. Thus, for example, we would want the intersection of two sets to exist and so any set-theory which failed to provide such a set would be inadequate.

I shall return to the question of the choice of axioms in section 4, where I shall discuss further the 'freedom' we have in choosing the axioms of set-theory.

Section 2

I

As mentioned in the previous section, I shall discuss Russell's reasons for denying that the phrase '$x \epsilon x$' is meaningful. I do not intend to discuss all of Russell's philosophy of mathematics nor even the whole of the doctrine of ramified type theory. I shall be concerned only with that part of his doctrine which touches on the above problem.

Russell's first thoughts on the discovery of his paradox seemed to be that the axiom of comprehension had to be given up. In his letter to Frege he writes,

'... there is no class of those classes which, each taken as a totality, do not belong to themselves. From this I conclude that under certain circumstances a definable collection does not form a totality.'[1]

This view is also indicated in his first paper on the subject some four years later.[2]

By 1908 his view had changed. A paper published that year outlined the theory of types in which it became nonsense to talk of a class being a member of itself.[3] The doctrine was embodied

1. B. Russell, 'Letter to Frege' (1902) in From Frege to Godel ed. J. van Heijenoort, Cambridge, Mass. 1967, p.125.
2. B. Russell, 'On some difficulties in the theory of transfinite numbers and order types' Proceedings of the London Mathematical Society, 1907, pp.29-53.
3. B. Russell, 'Mathematical Logic as based on the Theory of Types', (1908) in Logic and Knowledge, ed. R. March, 1956, pp.59-102

in Principia Mathematica[1] and his reasons for adopting the theory of types were given in greater detail.

In Principia Mathematica classes were considered to be logical fictions. Statements about classes could be translated into statements about propositional functions. The explanation of what a propositional function is remains very obscure. As Quine[2] has noticed, quantification over propositional functions which Russell allows (the axiom of reducibility asserts the existence of certain propositional functions) implies that Russell has not rid mathematics of abstract entitiew. As Quine says, Russell has replaced a clearer notion by one that is more obscure. Russell himself, in his discussion of propositional functions, veers from thinking of the function as what Quine would call an open sentence tothinking that it is some kind of entity over which we may quantify.

He speaks of a propositional function being an ambiguity. 'A function, in fact, is not a definite object ...., it is a mere ambiguity awaiting determination, and in order that it may occur significantly it must receive the necessary determination'.[3] But if propositional functions are not definite objects but mere ambiguities how can one apply existential and universal quantifiers to them?

---

1. B. Russell and A. Whitehead, Principia Mathematica, Cambridge, 1913
2. W. Quine, 'Whitehead and the Rise of Mathematical Logic' (1941) in Selected Logical Papers, New York, 1966, p.19-22.
3. B. Russell, Principia Mathematica, p.48

Russell's analysis of classes in terms of propositional functions provides his solution of the Russell paradox. For to ask of a class whether it belongs to itself is to ask whether a propositional function is satisfied by the class determined by that function.[1] The problem reduces to the problem of whether a propositional function can satisfy itself.

Russell has two arguments to show that it cannot. One rests on the simple theory of types and the other on the vicious-circle principle. (That the simple theory of types is logically independent of the vicious-circle principle was pointed out by Gödel)[2]

The former argument relies on the essential ambiguity of the propositional function. In his discussion Russell considers the possibility of substituting a propositional function for an individual in an elementary proposition. The argument is general and its conclusion is that propositional functions are divided into ranges of significance or types. The reason he gives for saying that a propositional function cannot meaningfully be an argument to an elementary propositional function is that a function is not a definite thing but an ambiguity awaiting determination. Consequently it is nonsense to say that

1. ibid. p.63
2. K. Gödel, 'Russell's Mathematical Logic', in The Philosophy of Bertrand Russell, ed. P. Schilpp, La Salle, 1944, p.147.

$\phi\hat{x}$ satisfies $\psi\hat{x}$ where both ψ and ϕ are elementary propositional functions.

This would seem too strong an argument and establishes more than Russell desires. If a function is 'a mere ambiguity' it is difficult to understand how a function can ever be an argument. That there are functions of higher type Russell does not doubt and does talk of functions as arguments to other functions. It is not clear how this can come about if functions are not 'definite things'. Nor is it clear, as I have already indicated, how these mere ambiguities can be quantified.

Russell makes the distinction between the symbol 'ϕx' and '$\phi\hat{x}$'. The first is what is ambiguously denoted, the second that which denotes (ambiguously) its many values. We may paraphrase Russell's talk of 'ambiguously denoting' and 'ambiguously denoted' in more modern terms. It is clear, I think, that Russell's use of 'ϕx' corresponds closely to the idea of an open sentence or sentence frame (Quine). For Russell says that 'By a "propositional function" we mean something which contains a variable, and expresses a proposition as soon as a value is assigned to x.'[1] There is difficulty, however, in trying to make the notion expressed by $\phi\hat{x}$ clearer.

1. B. Russell, Principia Mathematica, p.38.

For, in one place Russell says that '$\phi \hat{x}$' is 'a single thing'[1] and, further on in the text, he says that '$\phi \hat{x}$' is 'not a definite object'.[2] Any interpretation of Russell's use of '$\phi \hat{x}$' is almost certain to contradict one of these characteristics. Quine asserts that Russell's propositional functions are attributes, in the sense that the open sentence 'x has fins' determines the attribute of finnedness.[3] But this seems to contradict Russell's claim that a propositional function $\phi \hat{x}$ denotes its values.[4] For the values of a propositional function (according to Russell) are propositions. If Quine were right then an attribute would denote a set of propositions. In speaking of propositional functions denoting, Russell implies that they are linguistic entities. It is clear that attributes in Quine's sense are not linguistic entities. But if $\phi \hat{x}$ is not an attribute but a linguistic entity which is different from the open sentence $\phi \hat{x}$, which denotes and which is a single, though not definite, thing, there would seem to be no possible interpretation which would fit.

As a result of Russell's obscurity at this point, it is hard to evaluate his argument for the conclusion that '$\phi \hat{x}$ is a man' is nonsense. The argument he does give, that in '$\phi \hat{x}$ is a man' nothing

1. ibid. p.40 2. ibid. p.48
3. W. Quine, 'On Frege's Way Out', Mind 64, 1955 p.146
4. B. Russell, Principia Mathematica p.40

definite is said to be a man can be applied to another of Russell's examples which he gives to illustrate the axiom of reducibility. In this example Russell considers '$\phi ! \hat{z}$ is a predicate required in a great general' which is a function of a function. Since $\phi ! \hat{z}$ is nothing definite, it could be argued that nothing definite has been said to be a predicate required in a great general. Russell needs more argument to show that the first case is meaningless but the second meaningful.

Since functions are divided into different types in such a way that function is of a higher type than its arguments it turns out that a propositional function cannot be meaningfully said to satisfy or not to satisfy itself. It is a special case of the more general thesis of the simple theory of types.[1]

The second argument that Russell gives for denying that a propositional function cannot be meaningfully said to satisfy

---

1. Convincing arguments against the theory of types in general have been presented by M. Black, 'Russell's Philosophy of Language' in <u>The Philosophy of Bertrand Russell</u> pp.232-240 who points out a new contradiction and suggests ways of modifying the theory, though he regards the modifications as unsatisfactory. That the theory of types cannot be presented without contradiction has been argued by P. Weiss, 'The Theory of Types', <u>Mind</u> 37, 1928 pp.338-348 and F. Fitch, 'Self-Reference in Philosophy' <u>Mind</u>, 55, 1946 pp.64-73. Black's problem has been examined by F. Sommers, 'Types and Ontology' <u>Philosophical Review</u> 72 1963, pp.327-263.

itself depends on the vicious-circle principle. A propositional function Russell claims presupposes its values. Again, neither open sentences nor attributes will fit, for neither presuppose a totality of propositions which could be called the totality of their values. (Expressions of the form $\phi(\phi a)$ are not excluded either by the vicious-circle principle or the theory of types. Yet another principle has to be invoked to ensure the meaninglessness of this expression, namely Russell's theory of the proposition.) Russell claims that expressions of the form $\phi(\phi\hat{x})$ are meaningless since ϕx presupposes ϕa, ϕb, ϕc, etc. Consequently the vicious-circle principle does not allow $\phi(\phi\hat{x})$ to be a value of $\phi\hat{x}$ since $\phi\hat{x}$ would then presuppose one of its values, i.e. $\phi(\phi\hat{x})$. This argument cannot be properly evaluated until an explication of the expression '$\phi\hat{x}$' is given and in what way it can be said to presuppose its values.

The vicious-circle principle is variously phrased by Russell. 'Given any set of objects such that, if we suppose the set to have a total, it will contain members which presuppose this total, then such a set cannot have a total.'[1] 'Whatever involves all of a collection must not be one of the collection.'[2]

---

1. B. Russell, Principia Mathematica, p.37
2. ibid.

'If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total.'[1] Gödel has shown that these three statements are not equivalent to each other.[2] There is also a vagueness about the first two on account of the words 'involve' and 'presuppose' which receive no elaboration. Although Gödel considers that the first two are more plausible than the third - he adopts a realist attitude to classes - it is not clear in what sense an object can be said to involve all of a collection, (though, as Gödel points out, a description of that object can be said to involve all of a collection). The third form of the principle Gödel considers to be false if classes are considered to be independent of our description or construction of them.[3]

Hintikka has a proof that at least one interpretation of the principle is insufficient to keep out the contradictions. His interpretation is that no definition of a set y should include a bound variable which admits y as an argument.[4] Originally Hintikka proposed an interpretation of variables occurring in formulae of the predicate calculus in which the variables would

---

1. ibid.
2. K. Gödel, 'Russell's Mathematical Logic', p.133.
3. ibid. p.136 (a further discussion of this point will be found in Chapter IV of this section)
4. K. Hintikka, 'Identity, Variables, and Impredicative Definition', Journal of Symbolic Logic, vol. 21, 1956 p.242. Also K. Hintikka, 'Vicious Circle Principle and the Paradoxes', Journal of Symbolic Logic vol.22. 1957 p.245.

exclude each other. An example he gives is the geometric axiom:

a) Any two points determine a straight line.

If this is interpreted as allowing the points to coincide then a) is false. If, on the other hand, the phrase 'any two points' is interpreted as 'any two distinct points' then a) is true. As applied to set theory distinctions are made between the interpretations that may be given for the quantifiers occurring in the axiom of comprehension. The quantifiers may be interpreted with various degrees of exclusiveness. The axiom of comprehension

1) $(Ey)(x)(x \epsilon y \equiv F(x))$

may be interpreted to mean

2) $(Ey)\ (x)[x \neq y \supset (x \epsilon y \equiv F(x))]$

or

3) $(Ey)(x)[x \neq y \supset (x \epsilon y \equiv F'(x))$

where $F'(x)$ is the same as $F(x)$ except that all expressions of the form $(Ez)K$ and $(z)K$ occurring in $F(x)$ are transformed into $(Ez)(z \neq y.K)$ and $(z)(z \neq y \supset K)$ respectively. 2) represents Frege's suggestion which Quine has shown to be inconsistent.[1] Hintikka regards 3) as being the simplest way of carrying Russell's

---

1. More strictly Frege's way out is represented by $(Ey)(x)[x \epsilon y \equiv (x \neq y.F(x$
Geach has found this to be inconsistent with $(Ex)(Ey)\ x \neq y$.
See W. Quine, 'On Frege's Way Out'.

vicious-circle principle into set theory since the variable y cannot be included in the range of any bound variable in F(x).[1]

It turned out, however, that a set-theory based upon 3) would be inconsistent with $(Ex)(Ey)x \neq y$.[2] Hintikka suggests that the quantifiers could receive a still more exclusive interpretation whereby the variable x in the axiom of comprehension is prevented from coinciding with any of the free variables in F'(x) as well as the variable y. Such a course would be suicidal for set theory as the definition of unit sets, couples etc. would be impossible. The vicious-circle principle in one interpretation is insufficient to stop the derivation of the paradoxes and, in the more exclusive interpretation, is too restrictive to be a basis of set theory.

Wang has questioned Hintikka's approach to the vicious-circle principle, claiming that it is based on 'a strenuous misunderstanding'.[3] Wang points out that although the range of the variable x does not include y in 3) yet it may include sets definable only in terms of y, e.g. the unit class whose only member is y. Hintikka's inconsistency proof demonstrates this

---

1. K. Hintikka, 'Identity, Variables and Impredicative Definitions' p.242.
2. K. Hintikka, 'Vicious Circle Principle and the Paradoxes'.
3. H. Wang, 'Ordinal Numbers and Predicative Set Theory', in <u>A Survey of Mathematical Logic</u>, Peking 1963, p.640

point nicely, for the contradiction is produced by considering a set c defined in terms of two sets a and b - in fact c is the set consisting of a and b - as a possible member of both a and b.[1] Whether or not Wang is right in saying that Hintikka's approach is based on a misinterpretation of Russell cannot be known because Russell's use of 'involves' and 'presupposes' is not made clear. It is more correct to say that both Hintikka and Wang have given possible interpretations of the vicious-circle principle as formulated by Russell. Wang may convince us that his constructivist interpretation is more philosophically justifiable but this is not to say that it is what Russell intended.

There are, then, many difficulties in Russell's thesis that '$x \epsilon x$' is meaningless. It depends on a chain of reasoning the links of which are each open to dispute, even the 'safest' of these, the vicious-circle principle itself. Wang, himself in sympathy with the constructivist version of the vicious-circle principle, claims that Russell has merely stipulated that '$x \epsilon x$' shall be meaningless.[2]

---

1. K. Hintikka, 'Vicious-Circle and The Paradoxes', p.245.
2. H. Wang, op.cit. p.641

II

I shall now consider a more recent solution of the paradoxes. This solution is presented by J. Tucker in two papers[1] containing major attacks on formalism and formalists. These attacks and the solution proposed have to be examined, for if Tucker is correct then the arguments of section I of this thesis are invalid and several points that I shall make in section 4 contradicted.

The first of Tucker's papers that I shall consider attacks the formalist doctrine and holds that formalism is untenable because formal languages cannot be entirely separated from informal discourse and because its formal concepts are dependent on informal concepts. By a formalist Tucker means any logician or mathematician who sees any special virtue in formal languages. I shall deal with certain points raised by Tucker in his general attack on formalism in part III of this section but for the moment I shall concentrate on his solution of the paradoxes.

He contends that formalists have been led astray by the paradoxes because they ignore the fact that the paradoxes

---

1. J. Tucker, 'The Formalisation of Set-Theory', Mind 1963, pp.500-518
J. Tucker, 'Constructivity, Consistency and Natural Languages', Proceedings of the Aristotelian Society 1967, pp.145-168.

occur by the breaking of informal language rules. Throughout his paper he relies on an intuitive or naive notion of set, nowhere defining what a set is or giving any postulates or axioms for it. Accordingly my criticisms of his arguments will also be on the pre-formal, pre-axiomatic level. Although I do not believe that we have a consistent intuitive notion of set I shall pretend throughout this part of the thesis that we do have such a notion. Tucker's solution may be regarded as a defence of the naive concept of set. For if the paradoxes arise solely through the breaking of informal language rules then the notion of a set does not have to be revised in the light of the paradoxes (as I have suggested in the first section of this thesis) since it is a consistent notion and the paradoxes arise only when extraneous language rules are broken. I hope to show, however, that he has not demonstrated that each paradox arises from the breaking of an informal language rule.

Firstly, it may be noticed that he deals not with every paradox but with only a few of them. Although any proposed solution of the paradoxes can be illustrated only by a selection and not by all of them, there is a difference between Tucker's solution and a solution which says, for example, that each paradox is caused by violating the vicious-circle principle. In the latter case, there is some guide enabling us to see for each paradox as it turns up whether it does depend upon a violation of the vicious-circle principle. In the former case, however, there is no such guide. One is told that each paradox depends upon breaking an informal language rule, but is given no guide to discover which informal rule is broken nor how it is broken in those cases he does not discuss. If Tucker contends that each paradox is caused by the neglect of some informal rule then it is his job to show the rule in each particular case and not just those that he chooses to illustrate. This does not imply that his contention is wrong but only that he has supported it inadequately.

I shall now consider the paradoxes he does deal with and show that in each case he has failed to show that they depend upon the violation of some rule.

The first paradox with which he deals is the Epimenides. Although it is not set-theoretic it is appropriate to consider Tucker's solution here because of the connection he makes between this paradox and the Russell paradox. 'This is false' Tucker says is applied only to statements which could be false. This is the rule which is followed in informal language. The paradox occurs when this requirement is ignored and it is pretended that 'this is false' taken by itself could be either true or false. Now it is one thing to give a rule for informal language and another to show that this rule has been broken. Tucker has given a rule but has not shown that the paradox is obtained by breaking it. He says that it is pretended that 'this if false' could be either true or false when considered by itself and not in conjunction with some other statement. But it is up to him to show why it can only be pretended that 'this is false' could either be true or false. In other words he has to show that 'this is false' cannot be either true or false when taken by itself. This he does not do; he merely asserts it. I do not mean to imply that it cannot be done, but only that he has not shown it. Certainly attempts have been made to show that 'this is false' cannot be either true or false. For example, Ryle's analysis of the paradox in terms of an infinite regress shows just this. Ryle[1] claims that 'the statement I am now making is false' is analysable into 'the statement I am now making, namely, the statement I am now making, namely .... is false'. (In the form Tucker chooses 'this is false' would become 'this, namely, this, namely, ... is false'.)

---

1. G. Ryle 'Heterologicality', Analysis, 1961

This analysis, if accepted, would show that the sentence 'this is false' does not express a true or false proposition because it does not express a proposition at all. Such an analysis would imply that 'this is false' when taken by itself could not be either true or false. It would then be breaking an informal language rule to pretend that it could be either true or false. But if we accept the analysis we are not tempted to break the rule. Tucker's argument that the paradox arises by neglecting an informal rule is simply not valid, for the rule is not broken if 'this is false' is considered to be either true or false. All that has been done is to assume a false proposition, namely, the proposition that 'this is false' is either true or false. Once it has been shown that 'this is false' is neither true nor false then there is no temptation to go on and form the paradox. It seems clear that, whereas Ryle's analysis does offer a possible solution to the paradoxes, Tucker's proposed solution is, at best, a hint at where a solution might be found.

The first of the set-theoretic paradoxes with which Tucker deals is the Russell paradox. He treats it analogously to the Epimenides; his discussion of it is even more brief. I shall quote it in full:

'Russell's paradox is obtained by breaking the rule that we say of a class that it is not a member of itself only if there is some way of establishing that it is not a member of itself.'[1]

There are two objections to this argument. Firstly, that as an informal rule it is imprecise and, on one interpretation of it, questionable. Secondly, even if it is accepted as an informal rule Tucker has not shown that it is broken in formulating the paradoxes.

In one interpretation of the above rule, which seems to me to be vague because of the imprecise word 'say', it becomes: we assert of a class that it is not a member of itself only if there is some way of establishing that it is not a member of itself. But this interpretation is not a rule which is necessarily broken when the paradox is formulated. For it ignores the fact that the paradox in question is the outcome, not of asserting that the class of all classes which are not members of themselves is not a member of itself, but of <u>supposing</u> that it is not a member of itself. The rule is inapplicable in this interpretation. If the rule is extended to cover asserting and supposing then the rule is questionable. The rule now reads in this new interpretation: we assert or suppose that a class is not a member of itself only if there

---

1. op.cit. p.510

is some way of establishing that it is not a member of itself. But this is a rule which is not followed in mathematics. In some mathematical arguments, for example, a proposition is established by supposing that its negation holds. To establish that there is no greatest prime number it is supposed or assumed that there is a greatest prime number. Supposing a proposition is one of the methods of establishing the negation of that proposition. To bring the argument closer to the Russell paradox I shall consider the class of all empty classes, i.e., the class y such that

$$(x)\ x \epsilon y \text{ if and only if } (z)\ z \notin x.$$

One way of proving that $y \notin y$ is to show that y is not empty. This can be shown by the fact that the null class $\emptyset$ belongs to y. But there is another method of proving that $y \notin y$ closely analogous to the method used in the demonstration of a large class of set-theoretic paradoxes. This method is to suppose that $y \epsilon y$. Hence, $(Ez)\ z \epsilon y$ and therefore $y \notin y$. It is concluded that $y \notin y$ since if $y \epsilon y$ then $y \notin y$.

In other words, a proposition is sometimes established, and perhaps can be established only by supposing its negation. If a proposition of mathematics can be established then there is no method of establishing its negation (assuming consistency). Its negation, however, is supposed even though there is no method of establishing this negation. Clearly, this interpretation of the rule is unsatisfactory. Yet what other interpretation is possible? Possibly an interpretation could be that we assert or suppose that a class is not a member of itself only if there is some way of establishing whether it is a member of itself or it is not a member of itself. This does not rescue the rule. For the general rule which seems to be behind this particular one is that we assert or suppose a proposition only if there is some way of establishing that proposition or its negation. Again there are mathematical proofs that involve the supposition of a proposition p in order to establish ~p. But if p can be supposed only if ~p can be established or p can be established, and ~p can be established only if p is supposed then we have come full circle and cannot answer the question of whether p can be supposed. Furthermore, it is common mathematical practice to examine the consequences of some supposition, e.g., Cantor's continuum hypothesis, when there is the logical possibility that neither the proposition supposed nor its negation can be established. It may be that Tucker would disallow such suppositions but he has produced no arguments for such a bar.

Turning now to the second objection that even if the rule is accepted then it is not clear how a formulation of the paradox breaks the rule, I shall argue that Tucker has not shown that the paradox does break the rule. As with the previous paradox, the Epimedides, it is one thing to state a rule, another to show that that rule has been broken. In the case of the Russell paradox, it is not the case that there is no way of establishing that the class of all classes is not a member of itself. (Here I am taking 'established' to mean the same as 'proved'. As Tucker does not elaborate on what he means by this vague word I may be misinterpreting his argument, but it is difficult to see what else could be meant by 'established' in such a context.) It can be 'established' by a normal mathematical procedure. The problem is not the lack of any method of establishing a particular proposition but that too much can be 'established'. Both the proposition that the class of all classes which are not members of themselves is a member of itself and the proposition that the class of all classes which are not members of themselves is not a member of itself can be 'established'. Of course, the fact that both propositions can be proved can be used to show that there is something logically wrong with both propositions. For example, from the fact that both can be proved it might be deduced that no such class as the Russell class exists and thus that the question of whether it belongs to itself or not does not arise.[1] But this argument

1. See section 1 of this thesis.

is not open to Tucker because he believes that in formulating the paradox an informal language rule has been broken, and any such conclusion is blocked.

Granted that the totality of all classes that are not members of themselves forms a class, Tucker has to show how his rule is broken by asking if the class is not a member of itself. This he asserts but does not show. But this is the most important question. It is one which most proposed solutions of the Russell paradox have attempted to answer. Russell's own solution in terms of the theory of type, for example, is designed to show that it is meaningless to assert of any class that it is not a member of itself and a fortiori it is meaningless to assert of the Russell class that it is not a member of itself. The many different solutions may be said to be just so many different ways to answer the question of why it is meaningless to suppose that the class of all classes that are not members of themselves is either a member of itself or not. If any of these solutions were to be accepted then there would be no temptation to break the rule. The position is analogous to that created by Ryle's solution of the Epimenides. If it is recognised that 'this is false' does not express a proposition when taken by itself, then there is no temptation to ask whether it is true or false. If it is recognized that it is meaningless to assert that the Russell class is not a member of itself, then there is no temptation to ask whether it is a member of itself or not. The informal rule in each case need not be invoked for it would not be broken. Tucker seems unaware that he himself has not given a solution.

Furthermore, he has the added difficulty of explaining why, in the case of some classes it does not break an informal rule to ask whether they are members of themselves, and in others, the Russell class for example, why it does. To ask of the class of all classes which are not members of themselves whether or not it is a member of itself seems to me to be logically similar to asking of any class whether it is a member of itself or not. If we call the predicate from which a class is obtained by abstraction the classifying predicate, the idea behind Tucker's solution seems to be that one must not ask whether the class so obtained satisfies its classifying predicate. Thus, the predicate 'does not belong to itself' collects into a class certain classes, but since this is the classifying predicate of that class it must not be asked of that class if it satisfies this predicate. But if this is the idea behind his solution then the possibility of asking of any class whether it is a member of itself is ruled out. For in asking this one is asking a question which is equivalent to asking whether the class satisfies its classifying predicate.

For to ask of the class of all unit classes, say, whether it is a member of itself is equivalent to asking whether the class is itself a unit class. To ask of the class of all finite classes whether it is a member of itself is to ask whether it is a finite class. I do not see how Tucker can draw the line and say of one class that it is meaningful to ask if it is a member of itself and of another class that it is meaningless. Russell's solution of course was to banish 'member of itself' into the realm of meaningless expressions regardless of which class it is applied to. Tucker does not intend to do this, for he is willing to allow that it is meaningful to ask of some dlasses whether they are members of themselves, for otherwise there would be no problem of applying 'member of itself' to the Russell class because there would be no such class.

In addition, Tucker has the problem of giving the rule and showing how it is broken for those paradoxes which closely resemble the Russell paradox. I refer to these classes of the form:

$$x \epsilon y \text{ if and only if } (z_1)(z_2) \ldots (z_n) \sim (x \epsilon z_1 . z_1 \epsilon z_2 . \ldots z_n \epsilon x)$$

and

$$x \epsilon y \text{ if and only if } (z_1)(z_2) \ldots (z_n) \left[(x = z_1 . z_1 = z_2 \ldots z_{n-1} = z_n) \supset x \notin z_n \right]$$

Each of these classes give rise to paradoxes and it is Tucker's job to show why these paradoxes break some informal rule. If he is to maintain his thesis he must be able to do just this. It is made more difficult by the fact that the class of all empty classes, referred to above, bears a striking resemblance to the paradoxical class y such that $x \epsilon y$ if and only if there is no z such that $x \epsilon z$ and $z \epsilon x$. The only difference between this definition and the definition of the class of all empty classes is the addition of '$x \epsilon z$'. What informal rule could be invoked to stop the application of 'member of itself' to one whilst allowing it for the other?

I do not deny that Tucker is able to deal with these paradoxes (and the paradox generated by the class of all grounded classes) along the lines that they break informal language rules. I assert only that it is difficult to see how they can be so explained.

I turn to the next set-theoretic paradox with which Tucker deals - the Burali-Forti paradox. He outlines it as a consequence of two theorems. Firstly, there is a theorem which states that 'the series of all ordinals up to and including any given ordinal exceeds the given ordinal by one. It follows from this theorem that there is no greatest ordinal. The other states that the series of all ordinals has an ordinal number. It follows that there is a greatest ordinal, namely, the ordinal number of the series of all ordinals. The two theorems contradict each other.'[1.]

Tucker then gives his solution of the paradox, tracing its formation to the breaking of informal language rules. He points out that the first theorem shows that the class of ordinals is a self-generating class, by which he understands a class such that whatever group of members is considered, it follows by the property of self-generation that there is yet another member. He argues that we cannot go on generating from such a class since self-generation can only be applied to it by the use of 'more than all' which is a clear case of breaking a rule of informal language.

The difficulty arises when the condition that the class of all ordinals has an ordinal is brought in, for it would seem that self-generation must apply to this ordinal also. But, Tucker contends, it is clear that self-generation cannot apply to this ordinal without breaking an informal rule governing the use of the word 'all', for "when we say all we really mean the whole lot, we mean there are no more to come. So when we speak of the class of all ordinals we really mean all of them. We cannot apply self-generation to this class because to do so would break the rule for this use of 'all'. We cannot speak of 'more than all'."[2]

---

1. ibid. p.511
2. ibid. p.512

Tucker's argument as expressed in his paper is confused. He does not seem sure whether to apply his concept of self-generation to the class of all ordinals or to the ordinal number of the class of all ordinals. Further, there seems some confusion over the theory of ordinal numbers. He says that 'the series of all ordinals up to and including any given ordinal exceeds the given ordinal by one.' Yet it is not the series that exceeds the given ordinal by one but the ordinal number of that series. Again, this property of the series of all ordinal numbers up to and including a given ordinal that its ordinal number exceeds the given ordinal by one is not sufficient to prove that the class of all ordinal numbers is a self-generating class in the sense given by Tucker. For it is the property of the classes of rationals, integers, prime numbers and many other classes that given any class of them up to and including any member then there exist yet other members not belonging to the particular sub-class. These latter classes, however, are not self-generating, for it is not the case that whatever class of them is chosen there are other members not included in that class. The class of all rationals, for example, does not yield another rational nor included in the class. What is needed in addition to ensure that the class of ordinals is self-generating is the theorem that the ordered sum of a class of ordinals among which there is no greatest member is greater than any member of that class.

But these objections may only be muddles which Tucker can clarify by a little more precision. There remain much greater objections. Firstly, his contention that self-generation cannot be applied to the totality of a self-generating class because it breaks an informal rule for the use of 'all' succeeds only in blurring the distinction between self-generating classes and other classes. For, if self-generation cannot apply to the totality of the class, it must be applicable to proper sub-classes only. But the definition of a self-generating class then degenerates into a tautology applicable to any class, self-generating or otherwise. For any class, there exists members not included in any proper sub-class. What, then, is the force of the distinction drawn by the definition of a self-generating class? All classes must become self-generating under Tucker's restriction of the applicability of that term to only proper sub-classes. Secondly, if, when we say of any class that <u>whatever</u> group of its members be considered there exist other members of the class not included in the group, it would seem to break the informal rule for the use of the word 'whatever' if it does not cover the class of all members of the class, since 'whatever' means 'whatever' and not 'whatever, except'. Yet this is a conclusion Tucker is forced to draw by preventing the concept of self-generation applying to the totality of a class.

If, for the moment, we accept Tucker's argument, it still leaves many questions which it is imperative to answer before the paradox is cleared up satisfactorily. In particular, if he is willing to accept that the ordinal number of the class of all ordinals, exists, which it appears he is, then if this ordinal is denoted by Ω, what is to be said of $\Omega + 1$? Since Ω is the greatest ordinal number then there are only two alternatives open to him. Firstly, that $\Omega + 1$ is equal to or less than Ω in which case there will be further contradictions arising because no class can be similar to any section of itself which this would imply (see below). Secondly, that $\Omega + 1$ does not exist. This would seem to be the most likely alternative for Tucker as he says that self-generation cannot apply to Ω. But what does it signify to deny existence to this 'ordinal number', for it is the ordered sum of two ordinal numbers, Ω and 1. It is even possible to find set representatives of it, namely, for Ω, the set of all ordinal numbers arranged in order of magnitude and for 1 the set consisting of - 1 alone. $\Omega + 1$ will then be the order type of the ordered sum of these two classes. Since the ordered sum of these two classes will be well-ordered, $\Omega + 1$ will be the ordinal of this class. How can existence be denied in this case? Tucker must answer this question before he can be said to have 'solved' the paradox.

The above argument rests on Tucker's assumption that there is a greatest ordinal, namely the ordinal number of the class of all ordinals, which he states as the second theorem of set-theory needed for the construction of the Burali-Forti. But this statement is not the only way in which the paradox may be expressed. A much more precise statement of the paradox would show less grounds for supporting Tucker's thesis that the paradox is grounded in the misuse of 'all'. For example, consider the set of all ordinals arranged in order of magnitude and let Ω be the ordinal number of this set. Consider the set of all ordinals up to and including this ordinal arranged in order of magnitude, then the ordinal of this class will equal $\Omega + 1$. Providing it is not assumed that Ω is the greatest ordinal number, there is no danger of misusing 'all' in the sense of Tucker. For Ω is only one ordinal amongst many and lies somewhere within the series of all ordinals and not necessarily at the end. (Analogously, if the continuum hypothesis is assumed, then then the set of all cardinal numbers up to and including the cardinal number of the continuum has a cardinal number which lies within the given set of numbers.) But this latter set is a section of the set of all ordinal numbers and, by a theorem of set-theory its ordinal number will be less than or equal to the ordinal number of the whole set. I.e. $\Omega + 1 \leq \Omega$. But this contradicts the theorem which says that for any ordinal number w, w is less than w + 1. In this statement of the paradox, there is no misuse of the word 'all' in the sense that we have tried to use it to mean 'more than all'. The fact that the class of ordinals is self-generating has not been used. Nowhere in this demonstration have we relied on the fact that the class of all ordinals gives rise to an ordinal not in the class. It was assumed that Ω lay somewhere <u>within</u> the class. Nowhere did we use as part of the demonostration that there was an ordinal lying outside of the class of all ordinals.

Of course, Tucker may still object that although the property of self-generation has not been used explicitly in this proof it has been included implicitly by the use of the ordinal number $\Omega + 1$. Tucker would no doubt say that this 'ordinal number' can only be produced by self-generation from the class of all ordinals which has the ordinal Ω Again, he will be confronted with the difficulty of the status of $\Omega + 1$. In the version of the paradox that I have given above this would be the place most vulnerable to an attack along his lines. It would seem that $\Omega + 1$ has been generated from the class of all ordinals up to and including Ω. Tucker, claiming uniqueness for this ordinal, would deny that this is a legitimate move since Ω is the only ordinal to which self-generation does not apply. This in turn would imply that $\Omega + 1$ does not exist (since if it existed the proof of the paradox would proceed unharmed) but, as I have argued previously, there is no clear meaning to this assertion. It would appear that Tucker objects to the self-generation of the class of all ordinals up to and including Ω, but there are ways in which the ordinal $\Omega + 1$ may be generated, other than by this process of self-generation. One has only to consider the definition of an ordered sum of two ordinal numbers to see that $\Omega + 1$ can be generated from other sets than the set of all ordinal numbers. By definition $\Omega + 1$ is the ordinal number of the ordered sum of the two representative sets, the set of all ordinal numbers (it must be remembered that this assumes that such a set exists, an assumption Tucker is willing to allow) and a set consisting of one member alone, - 1 say. A set-representative of $\Omega + 1$ can thus be found which is different from the set of all ordinals arranged in order of magnitude up to and including Ω. Provided that the set of all ordinals exists, there seems little reason to deny existence to this new set and by definition $\Omega + 1$ will be its ordinal number. To assert, in the face of this set, that there is no such ordinal as $\Omega + 1$, as Tucker's view implies, is to misunderstand the notion of an ordered sum of two ordered sets. His arguments could only apply if there were just one way of generating the ordinal $\Omega + 1$ and that way was by self-generation. As I have shown, there are other ways and for these more explanation is needed than he has given.

Even if Tucker could show that there was no such ordinal number as $\Omega + 1$, there are other ways of stating the paradox which do not depend on this number in any way and do not, as far as I can see, depend upon self-generation from the class of all ordinals.

Let Ω be the ordinal number of the class of all ordinal numbers. Then, by a theorem from the theory of ordinal numbers, the class of all ordinal numbers, arranged in order of magnitude, less than Ω has the ordinal Ω. But this latter class is a section of the class of all ordinal numbers which has the same ordinal number, Ω. Now, two classes have the same ordinal number if and only if they are similar. Therefore, the class of all ordinal numbers is similar to a section of itself, the section consisting of all ordinal numbers less than Ω. This contradicts the theorem of set-theory which states that no set can be similar to a section of itself.

Stating the paradox in this way, there is no use made of the self-generative property of the class of all ordinals explicitly or implicitly. It is not based on the class of all ordinals up to and including Ω but only with the class of all ordinals less than Ω. Tucker has to show, if his thesis is to be maintained, that some informal language rule has been broken in such a demonstration of the paradox. So far as I can see he will be unable to appeal to either the property of self-generation or to any suspicious use of 'all'.

Finally, there is the additional drawback that if Tucker's reasoning is accepted many proofs in mathematics become more doubtful. Some mathematical proofs depend on a <u>reductio ad absurdum</u> which involve classes that generate members not included in the totality of members of that class. This is perhaps easier to see in examples; I shall give two.

a) If every class of integers has a least member then the principle of mathematical induction holds, i.e. if p(1) and if p(n) implies p(n+1) then for all n p(n). Suppose the statement a) to be false. Then, every class of integers has a least member, p(1), p(n) implies p(n+1), and there is an n such that $\sim$ p(n). Let s be the class of all n such that $\sim$ p(n), then there is a least member of s, m say. $m \neq 1$ since p(1). Therefore m is greater than 1. Now, if p(m-1) we should be able to deduce p(m), therefore, $\sim p(m-1)$. I.e. m-1 is a member of s, but m is the least member of s which is a contradiction. Hence a) holds.

b) A similar mapping of a well-ordered set W onto a subset never relates a member w of W to an image which preceded w in W. Suppose b) false. Then, there is a mapping f such that at least one member of W is mapped onto an image which precedes it in W. Let t be the set of all members of W which, by the mapping f, are related to images preceding them in W. Since t is a subset of W t will be well-ordered and have a first member x. Let f(x)=y. y is less than x. Because f is a similar mapping f(y) is less than f(x), i.e. y. Therefore y belongs to t, but y is less than x and x is the least member of t. Since this is a contradiction b) holds.

In both of the above examples use has been made of classes which generate members different from any member of the totality of that class. In a) s was such a class and m-1 the member generated from it which was different from the totality of members. In b) t was such a class and y such a member. In both cases the resulting contradictions were used to negate the hypothesis from which they were proved. How does such a use of these classes differ from the use made of the class of all ordinals to generate $\Omega + 1$? If such a use really does break an informal rule for the use of 'all', then not only does Tucker prevent self-generation being applied to the class of all ordinals, but also self-generation as applied to these classes. The theorems which a) and b) state can no longer rely on the proofs presented above, since these proofs depend upon a misuse of 'all' (according to Tucker). One consequence of Tucker's argument is that accepted mathematical proofs like the above can no longer stand. Such proofs are often used in analysis as well as in set-theory (e.g. the proof of the theorem that a continuous function in a closed interval is bounded and takes every value between the values of the function at its end points). Perhaps Tucker believes that these proofs are invalid, but it is important to realise these implications of his arguments.

I shall conclude my objections to Tucker's method of solving the Burali-Forti paradox by observing that it is misleading to talk of <u>the</u> Burali-Forti paradox and it is this way of talking which leads (or misleads) paradox-solvers into thinking that there is but one paradox to solve. In fact, 'the Burali-Forti paradox' covers a class of different, though connected, contradictions.[1] I have given above two different statements of the paradox, one of which has the contradictory conclusion

---

~~1. ibid, p.153~~

that $\Omega + 1$ is less than or equal to Ω , the other that there is a set which is similar to a section of itself. There is a third, which Tucker seems to concentrate on, that, stemming from the theorem that for any set of ordinals there is an ordinal greater than any in the set, states that there is an ordinal greater than any ordinal. It is only in the last that the notion of self-generation is applied to the set of all ordinals and which can be attacked along the lines Tucker suggests. Concentrating only upon this expression of the paradox, it would appear that he neglects the first two which do not involve the notion of self-generation applied to the set of all ordinals. A solution of the Burali-Forti paradox must 'solve' all of the contradictions which are referred to by that name.

If, as Tucker claims,[1] it is necessary to pin-point the exact place at which the contradiction occurs and then see what language rule has been broken, then it is precisely this which he has failed to do. He has chosen only one of many contradictions which fall under the same head and is thus led to pin-pointing self-generation as the cause of the contradiction. A wider view of the paradox which would include all three expressions of it might have prevented this. For, if we can specify the

1. ibid, p.153

one concept which is essential to allow the working of the paradox and which is common to all three forms then that concept is not self-generation but the concept of <u>the set of all ordinals</u>. It is this which leads us to talking about <u>the</u> Burali-Forti paradox. By 'the Burali-Forti paradox' we really mean any paradox which involves this concept and as long as it is understood in this way no harm ensues. The mistake of confusing just one paradox with a class of paradoxes leads to such unsatisfactory solutions as Tucker's. It is, perhaps, the fact that the set of all ordinals enables contradictions in a variety of contexts to be deduced that arouses suspicion of the set in question rather than the deductions which are made from it. Only if Tucker can show that each paradox belonging to this class relies on a faulty application of a language rule can he be said to have given a solution of the Burali-Forti paradox. It is for a very good reason that logicians have been unhappy about the set of all ordinals.

The next paradox with which Tucker deals illustrates all the faults of his discussion of the Burali-Forti. This time he manages to confuse not two paradoxes belonging to the same class but two altogether different paradoxes. I give his discussion in full.

'There is a theorem to the effect that the cardinal of the set of all sub-sets of a given set is greater than the cardinal one concept which is essential to allow the working of the paradox and which is common to all three forms then that concept is not self-generation but the concept of <u>the set of all ordinals</u>. It is this which leads us to talking about <u>the</u> Burali-Forti paradox. By 'the Burali-Forti paradox' we really mean any paradox which involves this concept and as long as it is understood in this way no harm ensues. The mistake of confusing just one paradox with a class of paradoxes leads to such unsatisfactory solutions as Tucker's. It is, perhaps, the fact that the set of all ordinals enables contradictions in a variety of contexts to be deduced that arouses suspicion of the set in question rather than the deductions which are made from it. Only if Tucker can show that each paradox belonging to this class relies on a faulty application of a language rule can he be said to have given a solution of the Burali-Forti paradox. It is for a very good reason that logicians have been unhappy about the set of all ordinals.

The next paradox with which Tucker deals illustrates all the faults of his discussion of the Burali-Forti. This time he manages to confuse not two paradoxes belonging to the same class but two altogether different paradoxes. I give his discussion in full.

'There is a theorem to the effect that the cardinal of the set of all sub-sets of a given set is greater than the cardinal of the given set. So for any given cardinal there is a greater cardinal. The notion of a cardinal is self-generating. By applying this result to the set of all sets Cantor had already obtained a contradiction which is isomorphic with that obtained later by Burali-Forti. It is solved in the same way.'[1]

From this paragraph it is necessary to extricate the two paradoxes and see how well the solution in terms of mis-applying the notion of self-generation to a totality fits each of the two cases.

Firstly, as was the case with Tucker's treatment of sets of ordinals, it should be noticed that Cantor's theorem - that the cardinal number of a given set is less than the cardinal number of the set of all the sub-sets of that set - does not ensure that the notion of cardinal number as self-generating in the sense Tucker uses it.[2] To deduce that the notion of cardinal number is self-generating it is necessary to use the theorem of set-theory which asserts that for any set of cardinals amongst which there is no greatest member there is a cardinal greater than any cardinal in the set, namely the cardinal number given by the sum of all cardinals in the set. It is only by combining this theorem with Cantor's theorem that we are able to

---

1. ibid. p.512

deduce that for any set of cardinals there is a cardinal which is greater than any in the set.

Secondly, only one of the paradoxes involved in his discussion can be said to rely on self-generation. The other seems independent of it. It is with the latter that I deal first.

Cantor's theorem leads directly to a paradox which does not need the notion of cardinal number at all; it needs only the concept of one-one correspondence. It states that the set of all sub-sets of a given set cannot be put in one-one correspondence with any sub-set of that set. Hence, if S denotes the set of all sets and US the set of all its subsets, US cannot be put in one-one correspondence with any sub-set of S. On the other hand, since US <u>is</u> a subject of S, US is in one-one correspondence with a sub-set of S. This is a contradiction. It is one of the paradoxes with which Tucker should be dealing. The question that presents itself is where does the notion of self-generation enter? We are no longer dealing with cardinal numbers but with sets and one-one correspondences between them. Nor has the paradox been rephrased in terms of sets in such a way that self-generation is still necessary for a deduction of the paradox. Certainly the concept of set is self-generating as a paraphrasing of the second paradox involved would show. But the fact that the notion of set is self-generating is not applied here. The set of all subsets of S is not another set different from any member of S, nor does the paradox 'prove' this. It depends on the equivalence and the non-equivalence of this set with any <u>sub-set</u> of S, not with any member of S. Hence the notion of self-generation is irrelevant for the deduction of this paradox.

The second paradox does involve the notion of self-generation and here Tucker's solution is relevant. The paradox involves the set of all cardinal numbers. Since to any set of cardinal numbers there is a greater cardinal - if, among the set there is a greatest cardinal, c, then the cardinal of theset of all sub-sets of a representative set of that cardinal c is greaterthan any cardinal in the given set; if there is no greatest cardinal number in the set then the sum set of these cardinals is greater than any member of the set - there will be a cardinal greater than any cardinal in the set of all cardinals. That is, there will be a cardinal greater than any cardinal, which is absurd. Tucker's solution to this paradox, since he says that it is solved in the same way as the Burali-Forti would be to maintain that an informal language rule has been broken:- the rule governing the use of 'all'. The same objections apply to this argument as were made against the solution of the Burali-Forti.

It should be clear that this solution leaves open many important questions which have to be answered before the paradox can be said to be solved. For example, if it is correct to talk of the set of all sets and the set of all cardinals and also to talk of these sets having cardinals as Tucker seems to imply, what is interesting and raises grave problems for his solution is the relationship between these cardinals and the cardinal of the set of all sub-sets of S and the cardinal of the sum of all cardinals. If both S and US have cardinals, which is the greater or are they equal? Whatever answer to this question is given, it will conflict with at least one theorem from set theory. If the cardinal of US is less than or equal to the cardinal of S then Cantor's theorem is contradicted; if the cardinal of US is greater than S then the very definition of 'greater' in cardinal number theory must come under revision. Tucker is silent on these implications of his solution. Again, if it is admitted that for any one cardinal there is a greater, what is to be said of the sum of all cardinals? For it is not the case that we simply use the theorem that for any set of cardinals there is a greater cardinal and, by applying this theorem to the set of all cardinals, arrive at a contradiction. We can be more explicit than this. The sum of all cardinals will be a cardinal greater than any cardinal, providing that there is no greatest cardinal. What relationship holds between this cardinal (equal to the sum of all cardinals) and any cardinal? Tucker's solution implies that it is less than or equal to some cardinal in the set of all cardinals. But this implies in turn that the theorem of set-theory which says that the sum of any set of cardinals amongst which there is no greatest member is greater than any in the set. He has not removed the paradox but shifted it so that other theorems of set-theory become paradoxical. If there were a greatest cardinal then of course there would not be a paradox involved in the notion of the set of all cardinals but there would be a paradox produced by Cantor's theorem. It is from Cantor's theorem that both the paradoxes under discussion spring. If there were some set - the set of all sets, say - which had the highest cardinal number then the cardinal number equal to the sum of all cardinal numbers would again be the greatest cardinal. But can it be deduced that there is a greatest cardinal? Cantor's theorem says there is no greatest cardinal. So it should be with this theorem that Tucker should be concerned since it is fundamental to the construction of the two paradoxes.

As I have made clear above the notion of self-generation and its limitations is of no help in this context.

There are axiomatic set-theories, and in particular Quine's 'New Foundation', in which Cantor's theorem is not forthcoming.[1] In this axiomatisation there is a set - the universal set - which is equivalent to the set of all its sub-sets. Working in this set-theory, there would be a set of highest cardinality and consequently no problem over the set of all cardinal numbers. But the fact that there is a set-theory which provides a set of greatest cardinality in no way supports Tucker's contention that there is a greatest cardinal number. There is no primarily philosophic motive behind Quine's system other than to effect a simplification and clarification of the theory of types in terms of stratified and unstratified formulae. It is a suggested axiom system among others. It has not been constructed from any conviction that the universal set must be a set of highest cardinality. Quine's axiom system does not then support Tucker's thesis that there is a set the cardinal of which is greater than any other cardinal. The difference between the two positions is between what a theory says to be so

---

1. W. Quine, 'New Foundations for Mathematical Logic', in <u>From a Logical Point of View</u>, Cambridge, Mass, 1953.

and what is claimed to be so.

The difficulty would be more apparent if Tucker tackled the task of axiomatising his set-theory so that one could see from exactly what assumptions he deduces such theorems that state, for example, that the set of all ordinals arranged in order of magnitude has the greatest ordinal number. One could then see which theorems of classical settheory remained and measure how adequate the theory was for the tasks asked of it and whether there was some way of reconciling, for example, Cantor's theorem with the theorem that there is a greatest cardinal number. Until such an axiomatised system is constructed it will be impossible to judge the success of the solutions proposed by Tucker. It is possible, however, (and this I have attempted to do) to show where the main difficulties lie and why the solution appears unsatisfactory and in some cases irrelevant.

I shall now consider Tucker's arguments concerning the diagonal argument and impredicative definitions. According to Tucker impredicative definitions do not cause the paradoxes for 'the sole causes of the contradictions are those already mentioned'.[1] However, certain uses of impredicative definitions occur in invalid arguments, not because they are impredicative

---

1. Tucker, op.cit. p.514

but because the arguments involving them are invalid. As an example of such an invalid argument Tucker takes the diagonal argument contending that the use of impredicative definitions to establish that the set of all sets of natural numbers cannot be put in one-one correspondence with the set of all natural numbers involves an invalid argument. Furthermore, he maintains, the diagonal argument can be reformulated in such a way that the use of impredicative definitions is unnecessary.

The proof of the non-denumerability of the set of all natural numbers that Tucker wishes to show contains an invalid argument is taken from Wang.[1]

'... suppose that the set of all sets of positive integers is denumerable. Then each positive integer has its corresponding set and each positive integer either is or is not a member of its corresponding set. Consider the set N of all those positive integers which are not members of their corresponding sets. Is n the positive integer whose correlate is N a member of N or not? If it is a member of N then by the condition of membership of N it is not a member of N. If it is not a member of N, then by the same condition it is a member of N.

---

1. Hao Wang, 'Formalisation of Mathematics', Journal of Symbolic Logic, 1954, p.246.

Wang thinks that these contradictions prove that the premiss is false and concludes that there is no such one-one correlation.'[1]

Tucker continues his argument:

'Now this argument is isomorphic with the heterological family of paradoxes. For N is the set of all those positive integers which are not members of their correlating sets, and in order to be a member of N a positive integer must already be correlated with a set. N is a second-order set which is parasitic for its members on other first-order sets. .... "is n a member of N?" could only be answered in the affirmative or the negative if n were already assigned to some set other than N. But this would be contrary to the condition that it is assigned to N and only to N. The argument turns on the breaking of this rule and is therefore invalid.'[2]

There are two points to be made with respect to this argument. Firstly, the use of 'already' which occurs twice in the above quotation, and, secondly, the use of 'first-order' and 'second-order'. What is the force of the word 'already'? Presumably, that the set N is not one of the sets of positive integers in the enumeration. But since, by hypothesis, the

---

1. Tucker, op.cit. p.515
2. ibid, pp.515-516

enumeration is of all sets of positive integers and N is a set of positive integers, N will occur somewhere in the enumeration. So it is not the case that the question 'Is n a member of N?' can only be answered if n is assigned to some set other than N. The enumeration is of all sets of positive integers and N will be one such. Tucker's argument would be more to the point if he were supporting a radical constructivist view which demands that a definition describes a construction involving the creation of some new entity which cannot be assumed to exist independently of the construction. This is the constructivist argument against impredicative definitions which Wang was discussing in the paper cited.[1] But no such scruples activate Tucker, for he says: 'If impredicative definitions are needed in mathematics, mathematicians can have as many of them as they like'.[2] He is prepared to accept that an impredicative definition is a method of picking out one entity from a pre-existing totality of entities. Thus he must be prepared to accept that N will be one of the sets by means of which N was defined. The consequence of this is that he may not use the phrase 'already be correlated with a set' to mean 'correlated with some set other than N'.

---

1. Wang, op.cit. p.246 et sec.
2. Tucker, op.cit. p.514

Secondly, the use of the expressions 'first-order' and 'second-order' to describe sets seems an invalid one, or at least constitutes a retrograde step. It was pointed out by Ramsey in a discussion of Russell's axiom of reducibility that the property of being elementary or non-elementary belongs not to a proposition, as Russell claimed, but more properly to instances of a proposition.[1] A proposition, according to Ramsey, may occur in two instances: one instance elementary, the other non-elementary. As an example he gives the proposition instances 'ϕa' and '$\phi a.(Ex)\phi x$' which are two instances of the same proposition, yet the first is elementary and the second non-elementary (in the sense of Russell). Such an argument rests on the assumption that two proposition symbols are instances of the same proposition if and only if they express agreement or disagreement with the same set of truth possibilities. Thus the whole hierarchy of orders stratifies proposition symbols rather than propositions (as the axiom of reducibility itself seems to suggest). Turning from Ramsey's theory of propositions, the same point may be made in connection with Tucker's division of sets into first-order and second-order. The notion of first-order and second-order properly belong to <u>the manner of</u> definition of a set rather than the set itself.

---

1. F. Ramsey, The Foundations of Mathematics, p.34

A set of positive numbers remains a set of positive numbers no matter how defined, providing the definition itself is unobjectionable. It is even clearer in the case of sets than in the case of propositions since there is a well recognised criterion for the identity of two sets: two sets are identical if and only if they have the same members (the axiom of extensionality). Thus, the set defined may have yet another definition of first-order. In order to clarify these two objections I shall give an illustration from set-theory which will show why it is absurd to say that we cannot properly ask of a set whether its co-relate belongs to that set and also why it is dangerous to talk in terms of the order of a set.

It will be instructive to consider not the set of all sets of natural numbers but the set of all <u>finite</u>[1] sets of natural numbers. I do so because this set is equivalent to the set of natural numbers and a correspondence can be established between each natural number and each finite set of natural numbers, thus simplifying the construction of 'second-order' sets. The correspondence can be established by means of arranging the sets in a sequence: A precedes B

---

1. In the case I am discussing, it is the set of all <u>non-empty</u> finite sub-sets of the set of all natural numbers.

in the sequence if the sum of the members of A is less than the sum of the members of B; if their sums are equal A precedes B if the least member of A is less than the least member of B or, if these are equal, if the next to least member of A is less than the next to least member of B, and so on. It is clear that before any set in this sequence there will only be a finite number of sets at most and that all finite sets of natural numbers can be reached in this way after only a finite number of sets. Hence, a correspondence has been set up, the first few terms of which are:

1	$\leftrightarrow$	(1)
2	$\leftrightarrow$	(2)
3	$\leftrightarrow$	(1, 2)
4	$\leftrightarrow$	(3)
5	$\leftrightarrow$	(1, 3)
6	$\leftrightarrow$	(4)
7	$\leftrightarrow$	(1, 4)
8	$\leftrightarrow$	(2, 3)
9	$\leftrightarrow$	(5)
10	$\leftrightarrow$	(1, 2, 3)
11	$\leftrightarrow$	(1, 5)
12	$\leftrightarrow$	(2, 4)

13	$\leftrightarrow$	(6)
14	$\leftrightarrow$	(1, 2, 4)
15	$\leftrightarrow$	(2, 5)
16	$\leftrightarrow$	(3, 4)
17	$\leftrightarrow$	(7)

In this correspondence it is clear that the set of all natural numbers which do not belong to their corresponding sets (a second-order set in Tucker's usage) is the set of all numbers greater than 2 (a first-order set). So that although the set has been defined by means of a second-order expression it does not entail that the set is of 'second-order'. The difference between 'second-order' and 'first-order' as applied directly to sets is seen to be unreal.

The correspondence also illustrates a much greater objection to Tucker's argument. On his own account we are debarred from asking of the 'second-order' set N whether n is a member of N or not. On exactly similar grounds he would have to admit that the same reasoning applied to the set of all positive integers that were members of their corresponding sub-sets. For this set too is a 'second-order' set parasitic for its members on 'first-order' sets. In the above correspondence between the set of positive integers and the set of all finite sub-sets of this set the set of all positive integers which are not members of their corresponding sub-sets is the set of all positive integers greater than 2 and the set of all positive integers which are members of their corresponding sub-set is the set (1, 2). Since there is no positive integer corresponding to the set of all integers greater than 2 (the set being infinite) the question of whether the n which corresponds to it is a member of the set does not arise. But the question may be aksed of the other 'second-order' set, the set of all positive integers which do belong to their corresponding sub-sets and, in this case, answered negatively, for 3 (which corresponds to the set (1, 2)) does not belong to (1, 2). In this case it is clearly absurd to maintain that all questions of the form 'Is n a member of N?' where N is a second-order set can only be answered in the affirmative or in the negative if n is assigned to some set other than N. For although we are discussing a different set from the set N it is still of the same 'order' as N and the question has been answered even though 3 is not assigned to some set other than (1, 2).

I am not asserting that in all cases a second-order definition can be replaced by a first-order one. Such an assertion would be tantamount to asserting the axiom of reducibility and dismissing the difficulties involved in the notion of impredicative definition. It is sufficient to point out that where a second-order can be replaced by a first-order definition Tucker's differentiation between first and second-order sets disappears and his arguments about what questions cannot be answered is seen to be invalid. Nor is it open to him to say that in other cases where the definition cannot be so replaced his argument still stands. For, since he expresses no worries about impredicative definitions, he must accept that a set of positive integers, no matter how defined, is still a set of positive integers and thus belongs to an enumeration of all sets of positive integers, should such an enumeration exist. If impredicative definitions of this type are legitimate then he must accept the reasoning involved in the proof of the theorem that the set of all positive integers is not equivalent to the set of all sub-sets of that set. To admit that the definition of N is impredicative is to admit that N is one of the sets by means of which N is defined. This is the meaning of 'impredicative'. By allowing impredicative definitions, it is invalid to argue that 'N is a second-order set which is parasitic for its members on other first-order sets'.[1] For N is one of the sets among these 'first-order' sets. His argument expresses a contradiction: on the one hand, he states that impredicative definitions are legitimate, that is, it is legitimate to define such a set N by means of a

1. Tucker, op.cit. p.515

totality of sets of which N is a member, and, on the other hand, that the definition of N ensures that N is somehow different from each of the members of that totality.

After arguing that those forms of Cantor's diagonal argument which involve impredicative definitions are invalid, Tucker proceeds to show that there are forms of the diagonal argument which do not make use of impredicative definitions. As an example he takes the proof that the set of all unending decimals is not denumerable. The proof, he says, consists in giving a rule whereby a decimal is constructed which differs from each decimal in a denumerable set of unending decimals. If the unending decimals are arranged in a sequence then corresponding to each positive integer there will be an unending decimal. To construct the required decimal all one needs is the rule that in its nth. place is an integer different from the integer in the nth place of the decimal corresponding to the positive integer n in the enumeration. 'It (this decimal) cannot appear (in the enumeration) because the rule for writing it down consists in making it differ from each unending decimal in the denumerable series. .... It follows that there are more unending decimals than there are integers.'[1]

1. ibid. p.516

Apart from the fact that the rule as Tucker gives it need not succeed in giving an unending decimal different from each of the decimals in the denumerable set (since the possibility of replacing the integers with 0 has been overlooked, thus producing a terminating decimal), there is a more important objection. It does not follow, from the fact that there is an unending decimal not included in the sequence that the set of unending decimals cannot be put in one-one correspondence with the set of all positive integers. It only follows from the theorem that there is a decimal not included in any enumeration of unending decimals. The addition of one object alone would not alter the cardinality of any infinite set. This mistake would not, perhaps, be very important if it were not the case that Tucker uses this incorrect account in his argument and which a correct account of this form of the diagonal argument would invalidate. He says:

'Now this argument differs entirely from that given by Wang! For the decimal which differs from each decimal in the denumerable list is not defined in terms of a totality of which it is a member. It is not defined in terms of a totality at all ..... It is not laid down that the totality of unending decimals is required prior to its construction. No totality is mentioned.'[1]

1. ibid. p.516

A more careful statement of the proof would have shown that reference to the totality of unending decimals is inevitable. There are at least two ways of proving that the set of all unending decimals is not equivalent to the set of all positive integers, though the two proofs are not essentially different. The first is to assume that an enumeration of the unending decimals is possible and then proceeds to construct a decimal which is different from all decimals in the enumeration which is a contradiction and, hence, by reductio ad absurdum the unending decimals cannot be put in one-one correspondence with the positive integers. Such a proof does involve the totality of all unending decimals since it is assumed that the set of all unending decimals can be enumerated. (This type of proof is barred to Tucker because from all of a collection we have generated one of that collection which does not belong to that collection.[1]) The totality of all unending decimals is required prior to its construction.

The second method of proof is based on a lemma. The lemma states that given any denumerable sub-set of the set of all unending decimals there exist members of that set which are not in the sub-set. This is proved by the usual construction of a decimal following the rule given above. This certainly does not

---

1. see above, p.

require the totality of all unending decimals. But the lemma does not prove that the set of all unending decimals is not equivalent to the set of all positive integers. The proof of the latter from the lemma is easy and proceeds by reductio ad absurdum. It is assumed that there is a one-one correspondence between the two sets and from this it follows that the set of all unending decimals is a proper sub-set of itself, which is a contradiction. It can be seen that the totality of unending decimals is again required in the assumption that they can be put in one-one correspondence with the positive integers. Nor is it the case that the totality of undnding decimals is not required for the construction of the decimal. For, although the lemma was proved before the latter theorem, the theorem is only a disguised version of the first method of proof. The lemma states that all denumerable sub-sets of the set of unending decimals are proper sub-sets of that set. The theorem is only one particular case of this lemma where the set in question is the set of all unending decimals.[1] It is equivalent to the first proof in this respect: that it is based on the assumption

1. This is perhaps clearer if we write the lemma $(D_0)\ (P{\sim}D_0{\subseteq}D) \supset D_0 \subset D$ where P is the set of all positive integers and D the set of all unending decimals. The theorem follows by the substitution of D for the bound variable D_0 Thus $(P{\sim}D) \supset (D{\subset}D)$ The theorem is proved because it is assumed that the lemma is true when D_0 is D and this is so only because the construction of the decimal is assumed to be possible even when D_0 is the totality of all unending decimals.

that the totality of all unending decimals are laid out in an enumeration. Tucker is therefore wrong in assuming that the construction of such a decimal does not require the totality of unending decimals.

If he were right that it involves no mention of the totality and that this decimal only differs from each of the decimals in the enumeration then it would be equally possible to maintain that the set N of all positive integers not belonging to their corresponding sub-sets differ from each of those sub-sets. Indeed such a proof is often given by means of an analogous lemma followed by a theorem.[1] It is curious that Tucker did not consider this to be a way out of referring to a totality of which N is a member.

Tucker's view also stops Cantor's theorem from being proved because it relies on the impredicative definition of such sets as N, whether we prove it directly or by means of a lemma. Cantor's theorem is about any set and the set of all sub-sets of that set. Thus the particular form of diagonal argument which meets with Tucker's approval is not available for the theory of abstract sets. It may be that he does not

1. S. Kleene, '<u>Introduction to Metamathematics</u>.' Amsterdam 1952, pp.14-15

object to the rejection of Cantor's theorem but this implication should at least be realised.

Lastly, Tucker deals with the Richard paradox. Let E be the class of all finitely definable decimals. Then E has $\aleph_0$ members. Let N be a decimal defined by means of the diagonal rule applied to an enumeration of all finitely definable decimals. Then N differs from each member of E. But since N is thus a finitely definable decimal N belongs to E. This is a contradiction. The paradox is swiftly dealt with:

'But this is not a paradox at all. The diagonal method does not of itself generate contradictions. For from the fact that N is finitely defined by the diagonal rule, it follows that E has more than $\aleph_0$ members. So the assertion that E has $\aleph_0$ members is thereby proved to be false, that is all. The form of the argument is simply that A is asserted and not-A is shown to be the case.'[1]

Once again a paradox is not solved but only moved to another place. It is not the case that it is simply asserted that the set of all finitely definable decimals is equivalent to the set of all positive integers. It can be proved on one highly plausible assumption: that the number of letters and punctuation marks employed in the English language,

1. Tucker, op.cit. p.516

say, is at most denumerable. Now the set of all finitely definable decimals will be a sub-set of the set of all sequences of letters and punctuation marks and this in turn will be equivalent to the set of all finite sequences of positive integers. This last set can be proved to be equivalent to the set of all positive integers. Hence, the set of all finitely definable decimals has the cardinal $\aleph_0$. If Tucker really means to pursue his argument to its logical conclusion, what is proved by reductio ad absurdum is that the number of different letters and punctuation marks available in the English language is greater than the number of positive integers, an assertion which is only a little less repugnant than the paradox with which we started. It is clearer still if we restrict the symbols which we are to use in a language to a finite number, to all the symbols on this page, say. Again the number of decimals finitely definable by means of the symbols on this page is $\aleph_0$. and the paradox follows through once N is defined as the decimal constructed by the diagonal rule for an enumeration of these decimals. Since all symbols occurring in the definition of N are on this page it must belong to the set of all decimals finitely definable by means of symbols on this page. The definition of N however ensures that it does not belong to that set. By reductio ad absurdum the set of all symbols on this page is non-denumerable. This is the conclusion we have to accept if we are to allow the diagonal argument as a valid argument and the notion of 'finitely definable decimal' as a legitimate concept. It is clear that the number of symbols on the previous page is not non-denumerable but finite. A solution of the paradox which involves admitting that a set of symbols which has thirty members (the number of symbols used on the previous page) has a non-denumerable number of members is unlikely to convince one of its plausibility.

III

I have devoted the whole of the last chapter to a detailed discussion of Tucker's solution of the paradoxes in terms of informal language rules because it is in direct contradiction with the position outlined in section I of this thesis. Also, the attack on formalism contained in his paper seems to be a backward step in foundational studies. The doctrine of formalism that is being attacked is not what might be called the 'strict' formalism of Hilbert (although this position would of course be open to the same attack); the net is spread wider to catch such different views as those of Quine, Curry and Church. The attack is to deny the need to formalise, to deny that formalisation succeeds in any clarification of mathematical concepts.[1]

There are two separate theses involved. The first is that there are indispensable concepts without which we should be unable to operate formal systems. The second is that any formalisation of these concepts must retain all the inprecision of their informal counterparts. The two concepts Tucker picks out for examination are the concepts of rule and the concept of substitution. These are both needed in the operation of

---

1. ibid. pp. 501-2

the formal system and any attempted formalisation of them will carry over any vagueness that the informal concepts have.

The first thesis is correct in as much as there must be one meta-level where the rules of a formal system are stated in informal discourse. For, in order to operate a formal system it is necessary to understand the rules of formation and transformation of that formal system. If these are formalised in turn, then clearly the rules of the meta-metalanguage have to be understood. The formalisation of the various meta-levels will still need a meta-language in which the rules are of an unformalised nature - framed in terms of informal discourse. But this thesis does not refute formalism, since formalists, including Hilbert have not claimed that one could operate a formal system without having some pre-formal concepts. Kleene, for example, in a book devoted to meta-mathematics (in the strict Hilbertian sense of finitary methods as opposed to what Kleene calls set-theoretic methods) writes:

'The meta-theory belongs to intuitive and informal mathematics. ..... The assertions of the meta-theory must be understood. The deductions must carry conviction. They must proceed by intuitive inferences, and not, as the deductions in the formal theory, by applications of stated rules. Rules have been stated to formalise the object theory, but now we must understand without rules how those rules work. An intuitive mathematics is necessary even to define the formal mathematics.'[1]

If Tucker were only arguing this thesis theme would be little or no disagreement between him and those he attacks.[2] To understand and operate a formal language it is necessary to understand and employ those concepts both formal and informal used in the meta-theory. But he is not content to stop there. He maintains that formal concepts, dependent as they are on informalconcepts retain all the unclarity of those informal concepts and therefore a formalist rejecting informal language because it is unclear must reject the formal language also. But even for those indispensible concepts that Tucker lists it is by no means certain that because they are informal concepts any unclarity attached to them carries over to a formal system which needs them as a basis.

For example, Tucker's argument that the informal notion of a rule is necessary in setting up a formal language

---

1. S.C. Kleene, Introduction to Metamathematics, Amsterdam, 1952, p.62

2. Even Church states that "In order to set up a formalised language we must of course make use of a language already known to us, ..., stating in that language the vocabulary and rules of the formalised language." A. Church, Introduction to Mathematical Logic, I. p.47

does not imply that because the informal notion is unclear the formal language is unclear also. His argument is based on the fact that any formal language depends on giving a set of rules: rules of formation which give conditions for a formula to be well-formed and rules of transformation giving conditions for a sentence to be an immediate consequence of another and for a string of sentences to be a proof of another. Certainly, the informal notion of a rule is employed in constructing a formal language. It must be understood that the rules of formation and transformation are rules; it does not follow, however, that the formal language is unclear because the informal concept of rule is unclear. For there is a perfectly harmless way in which informal notions can be said to be unclear. It is unfortunate that Tucker does not say in what way the informal notions are unclear. Körner, on the other hand, has given a definition of a type of unclarity especially useful in discussions of the problems of pure and applied mathematics.[1] Concepts are divided into two categories: exact and inexact. It is his contention that mathematicsl concepts such as 'line', 'group', '3' etc. are exact whereas the corresponding empirical concepts are

---

1. S. Körner, The Philosophy of Mathematics, London, 1960, p. 159ff.

inexact. A concept is exact if it does not admit of borderline cases; a concept is inexact if it does. If a concept is unclear because it admits of borderline cases, i.e. because it is inexact, then this type of unclarity is not necessarily transferred to a formal language which employs such a concept. In the case considered here, if 'rule' is unclear in the sense that it is inexact then it does not follow that the formal system must share this inexactness. For although the construction of any formal system is dependent on the informal notion of a rule, what is required in connection with the rules of formation and transformation is the recognition that they are rules. The fact that the concept of 'rule' admits of borderline cases does not imply that the rules of formation and transformation are borderline cases. The fact that there are some cases which are neutral instances of the concept 'rule' does not imply that all cases are neutral instances of the concept 'rule'. The inexactness of the concept 'rule' does not carry over to a formal system given as a body of rules.

The same argument applies to Tucker's second thesis that the formalisation of the concept of a rule carries with it all the unclarity of the informal notion it is supposed to formalise. To formalise the concept of 'rule' a rule must be given to show that the formalised concept of a rule is intended to replace the informal notion. A rule is employed in laying down what is to be counted as a rule in the new, formalised sense. Again the concept of rule is informal and Tucker again argues that any unclarity pertaining to this concept of 'rule' is carried over to the formalised concept. But if the notion of 'rule' is unclear because it is inexact then the formalised notion of rule need not be unclear in that sense. Providing the rule laying down that the formalised notion of a rule is to replace the unformalised notion of a rule for some specified meta-language can be recognized as a rule, then there is no inexactness carried over to the formalised concept. Similar arguments apply to the other indispensible concepts listed by Tucker.

Another, though related, sense in which the concept of rule may be said to be unclear is in the difficulty of giving a precise definition of 'rule'. In a discussion of rules Waismann claims that it is indeed difficult to give a clear definition. This, he says, is because '... the word "rule" like the term "ostensive definition" is one which stands for all sorts of different things which merely have a certain similarity in their use, and which the use of language groups loosely together.'[1] In other words, the things that we call rules have a 'family likeness' in Wittgenstein's phrase. If rules are like this then clearly any proposed formalisation of the rule concept is going to differ considerably from the informal concept. We may look at such a formalisation as an explication of the concept.[2] If this is the sense in which 'rule' is thought to be unclear then the formalisation will escape this kind of unclarity.

It may be that Tucker understands 'unclear' in some other sense than 'inexact' or 'lacks a clear definition' in which case it may be that the unclarity is transferred from the indispensable informal concepts to the formal language itself. But it is necessary to explain exactly what sense of the word 'unclear' is to be understood when he affirms that 'if the formalist premiss that all the usage which occurs in informal language is unclear is correct, then we are condemned by the premiss to perpetual unclarity'.[3] The formalist may still argue that informal language is unclear in the sense that it is

---

1. F. Waismann, The Principles of Linguistic Philosophy, ed. R. Harre, London, 1965. p.140

2. The concept of explication is examined in section 3 of this thesis.

3. Tucker, op.cit. p.503

inexact and yet maintain that his formal language does not contain this unclarity.

In the same way, the fact that in the string of meta-languages that a formalist may set up to construct and study a formal system there must be one which is stated in informal language does not imply that the formal language is as unclear as the informal language used in that meta-theory. Any notion used in the informal language of the meta-theory may be inexact. There may be contexts where the concepts are unclear but this does not imply that they are unclear in the context of the meta-theory. For example, in a meta-theory the class of provable formulae may be defined as 'the smallest class of formulae which contains the axioms and is closed with respect to the relation "immediate consequence of"'.[1] The concept of 'relation' employed in this definition may be inexact in the sense that it is not always clear whether any particular term is a relation or not. It is sufficient, however, for the formal language not to carry with it this inexactness, that the concept of 'immediate consequence of' be recognised as an instance of the concept 'relation'.

In summary, then, formal languages depend upon informal languages and informal concepts, but any unclarity in the

---

1. K. Gödel, <u>On Formally Undecidable Propositions of Principia Mathematica</u> trans. B. Meltzer, London, 1962, p.45

sense of inexactness of those informal concepts does not imply that a formal language is doomed to the same inexactness. At one point Tucker seems to be saying that no clarification can ever be made by using mathematical logic since mathematical logicians frequently fall into conceptual confusions. He cites Skolem's view that the concepts of set-theory are relative. Yet surely Skolem's proposal to replace Zermelo's 'definite proposition' by the notion of an expression which contains as atoms only expressions of the form 'a∈b' or 'a=b' is a clarification of that notion.[1] The fact that Skolem may have fallen victim to a conceptual confusion elsewhere does not mean that no clarification has been achieved at this point.

Before concluding this chapter I should like to discuss two other theses Tucker holds. Firstly, that contradictions occurring in a formal language are understood if they are understood in informal language. Secondly, that the contradictions can be solved only by looking to see which informal language rule has been broken. The syntactic contradictions of set-theory are just as much paradoxes when stated in informal languages as they are when stated in formal systems. One may agree to this without rejecting as Tucker does the

1. E. Zermelo, 'untersuchenden über die Grundlagen der Mengenlehre I' Mathematische Annalen 59, p.262. T. Skolem, 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', Matematikerkongressen i Helsinfors den 4-7 Juli 1922, Den femte Skandinaviska matimatikerkongressen, Redogorelse, Akademiska Bokhandeln, Helsinki 1923. p.218

distinction drawn by Russell between the mathematical and philosophical aspects of foundational problems. They remain distinct even though the paradoxes of set-theory may be solved in the same manner for both the formal and informal statements of them. What Russell means in the passage[1.] cited by Tucker is that the mathematics of the theory of types, i.e. the deduction of theorems from assumptions embodying type-theory, is separable from the philosophic justification of the theory of types. Even if the theory of types turns out to be without such justification and philosophically unsound, the mathematics of type-theory may still be developed. The mathematics of a certain set of assumptions is independent of the justification of those assumptions although interest in the mathematical development of them may not be.

Syntactic paradoxes that are contained in any <u>formalised</u> system of set-theory are statable in informal language. Formalists have not denied this and often introduce the problems involved in the construction of a formal set-theory by the paradoxes with an informal discussion of the paradoxes.[2] This should be qualified, however. A paradox occurring in a <u>formal</u> system is a purely formal characteristic. It occurs when,

---

1. B. Russell, <u>Logic and Knowledge</u>, London 1956, p.102
2. e.g. R.L. Goodstein, <u>Mathematical Logic</u>
   Curry, <u>Introduction to Mathematical Logic</u>

say, two formulae are provable one of which is identical in form to the other except that it is prefaced by '$\sim$'. A formal language consisting of a certain vocabulary and certain syntactic rules may be given such a definition of inconsistency. Any two formulae which have the above the form would constitute a contradiction in that formal language. But, what does it signify to say that such a contradiction if understood can be understood in informal language? As yet, the language is considered to be formal and not as a formalised language. In the role of formal language it is a game played with certain pieces according to a certain set of rules. A contradiction occurring in a formal language will act (if the propositional calculus is an interpretation of a sub-system of that language) as a license to give theorem status to any well-formed formula of that language.

In order to 'understand' the contradictions in informal language it is necessary to give semantic rules in addition to the syntactic rules. There have to be some rules of translation for the formal system to receive an interpretation. Formal languages as opposed to formalised languages may have no obvious translation or interpretation. If a contradiction occurred in such a formal language, it would be puzzling to say that one could 'understand' why the contradiction occurred. The most one could hope for would be some alteration to the axioms of the system which would avoid the proof of that contradiction being available. Nor, in this case, would it be possible to use the notion of interpretation, for an inconsistent formal system has no interpretation. For a purely formal language it makes no sense to say that we can 'understand' the contradictions that occur. At most the axioms may be altered so that no contradictions can be proved, but the alterations that may be made will not be made as a result of examining the informal translation of the formal system. (It would in any case be wrong to talk of <u>the</u> translation or <u>the</u> interpretation of a formal system, because for formal systems which have an interpretation in an infinite domain there will be two non-isomorphic interpretations[1]. Rather, the corrections will be made as a result of technical expertise resulting from working with formal languages. Of course, any solution of an informal paradox which gains general acceptance will be incorporated in the <u>formalised</u> language. An informal solution will be reflected in the formalised theory. There are, as Tucker says, no formal contradictions to solve as well, once the informal contradictions are ironed out. For formal languages, however, there is no possible way of 'understanding' the cause of any contradictions that occur, so that there will be a fear that the formal system is inconsistent. Tucker is wrong in saying

---

1. see section 4 of this thesis for further discussion.

that in formal systems '.... there is no such thing as a contradiction which has no cause, no such thing as a contradiction that cannot be tracked to its source.'.[1] For a formal system the only thing that could be regarded as the cause of a contradiction would be an axiom of the system such that its removal would result in a consistent system. But this is not to give an explanation of why its contradicts the other axioms; it is to say only that it does contradict them.

---

1. J. Tucker, Formalisation of Set-Theory, p.513. More explicitly he maintains that the following are false: 'That in a formal system for which there is no guarantee of consistency a contradiction may turn up unexpectedly anywhere. That there may be latent contradictions spread through such a system.' ibid. p.510

IV

Having disposed of formalism to his own satisfaction and given solutions to the paradoxes in terms of informal language rules, Tucker turns to the problem of constructivity which he considers to be logically prior to the problem of consistency[1] Once it has been shown that a language is constructive there will be no worry over the problem of consistency.

To make sense of this thesis it will be necessary to see how Tucker uses the word 'constructive' as it is not clear how it is to be used in all the contexts in which it occurs. As employed by Tucker it is an adjective that can qualify 'conditions'[2], 'procedures'[3] and 'language'[4]. He defines it, however, only as it applies to procedures.

'Constructive procedures are defined as procedures which can, in some sense, be carried out whereas non-constructive procedures are those which, while they can (in some sense) be specified cannot be carried out.'[5]

It is not at all clear how such a definition can be extended to cover the case of a natural language being constructive.

---

1. J. Tucker, 'Constructivity, Consistency and Natural Languages', Proceedings of the Aristotelian Society, 1967 pp.145-168.
2. ibid. p.164 3. ibid. p.152 4. ibid. p.145
5. ibid. p.152

Tucker admits that his definition of 'constructive' differs from any given in the past. For there will be some procedures that have been regarded in the past as constructive that are not constructive according to the new definition and vice versa.[1] He maintains, in addition, that the limitations imposed on the methods of proof by constructivists are merely arbitrary chosen restrictions having no significance for foundational studies.[2]

There are two objections I wish to make at this point.

Firstly, it does not follow from the definition that there will be procedures which turn out to be constructive (non-constructive) under the definition, but have in the past been considered non-constructive (constructive). It is doubtful whether any constructivist would wish to disagree with the definition. What is doubtful is whether there would be agreement over what counts as 'a procedure which can be carried out'. It is not over the definition that there would be dispute but over what procedures can be carried out. The reason that there appear to be many different standards of constructivity is due to there being many different views as to what constitutes a procedure which can be carried out. For varied reasons one may reject impredicative definitions, pure existential theorems,

---

1. ibid p.153
2. ibid p.153

any proof involving the notion of the totality of all real numbers, any proof involving the notion of an arbitrary set, etc. as non-constructive. Constructivists who reject some or all of these as being examples of non-constructive procedures may do so because the procedures cannot in some sense be carried out. Tucker's definition can be seen to be virtually useless as a clarification of constructivity and for differentiating between his notion and those of other constructivists. He claims that there is a ' ... a single basic notion of constructivity that is essential to foundations' and, further, that non-constructive procedures are non-constructive ' .... in the precise sense that they are impossible of execution'[1] (my italics). It is clear from the above argument that the definition has not given any precision to the notion of constructivity nor does it help us to classify procedures that are constructive and those that are not.

Secondly, Tucker's assertion that the limitations on methods of proof have in the past been arbitrary restrictions imposed by constructivists is a gross misrepresentation of the facts. A general account of constructivity is beyond the scope of this thesis but in order to see that Tucker's account of the 'arbitrary limitations' is incorrect it will be necessary

---

1. ibid. p.153

to examine some of the constructivists' views and why they reject certain proof procedures.

The intuitionists, for example, believe that mathematical assertions are reports of successful mental constructions. The exact nature of these mental constructions is difficult to specify and its dependence on an intuition or acts of intuition unappealing to empiricist or analytic philosophers. But if we grant, for the moment, that it makes sense to talk of mental constructions it should be clear that the logical connectives which the intuitionists themselves use in their reports of mental constructions will receive very different interpretations from the usual truth-table interpretation.

The proposition '$\sim p$', since even negative propositions are reports of a mental construction, is not just a report of the absence of a construction but is the report of a construction which deduces a contradiction from the supposition that the construction reported by 'p' were brought to an end.[1]

Existential propositions of the form $(Ex)A(x)$ have no other meaning than: 'A mathematical object x satisfying the condition A(x) has been constructed'.[2]

For the intuitionist mathematical objects, whether they are sets or natural numbers or real numbers are essentially

---

1. A. Heyting, Intuitionism - an Introduction', Amsterdam, 1956
2. A. Heyting, 'Some Remarks on Intuitionism' in Constructivity in Mathematics, Amsterdam 1959, p.70

<u>constructible</u> objects.[1] The apparent peculiarities of intuitionist mathematics spring from this conception of mathematical objects. If mathematical objects do not exist prior to their construction - Heyting claims that he is unable to make sense of the assertion that they do - then the rejection of pure existential proofs follows as a consequence. Similarly some instances of the law of excluded middle must be rejected since, both 'p' and '~p' being reports of constructions, there will be cases - in particular, cases involving quantification - where we are in possession of neither construction. Other logical laws to which the intuitionists object can be considered in the same way. The justification for their rejection is the nature of mathematical objects.

There is nothing arbitrary about the restrictions and limitations on methods of proof, for the limitations are laid down by the nature of the mathematical objects and, Heyting says, there is nothing arbitrary in the notion of a constructible object.[2] The notion of a constructible object must itself be a primitive undefined notion since any attempt to define those operations that are constructive would need existential quantification[3]. Nevertheless, what is meant by a construction

---

1. ibid. p.70 2. ibid. p.70
3. R. Peter, 'Rekursivitat und Konstruktivitat' in <u>Constrmctivity in Mathematics</u>, p.228.

can be made clear by examples.[1] Since the notion of a constructible object is not arbitrary neither are the methods of proof which the intuitionists allow.

Another constructivist, Wang, gives good reasons for rejecting impredicative definitions. There is nothing arbitrary in this rejection. It is not a ban on impredicative definitions imposed simply because impredicative definitions sometimes lead to paradoxes. Nor is it necessarily a ban on all kinds of impredicative definitions. One may, for example, accept impredicative definitions of natural numbers but not of sets of natural numbers.[2] But where one allows impredicative definitions and where one disallows them is not purely arbitrary; it will depend on what one considers to be the nature of the objects over which the quantified variable in the impredicative definition ranges.

Wang, speaking of the vicious-circle principle, says that the principle is directed against the <u>introduction of new objects</u>.

'Impredicative characterisations are objected to not just as such but only as a means for initially introducing an object.'[3]

If a set can be said to exist only after it has been defined

---

1. A. Heyting, '<u>Intuition - an Introduction</u>' .
2. H. Wang, 'Ordinal Numbers and Predicative Set-Theory', in <u>A Survey of Mathematical Logic</u>, Peking, 1963 p.642
3. H. Wang, ibid. p.640

then clearly it is circular for this definition to contain a quantifier ranging over this set. Only if sets exist prior to their definition, in which case the 'definition' would be a specification of one object from an existing totality of sets, would predicative definitions be legitimate.[1] The ban imposed by constructivists on impredicative definitions is an outcome of how the mathematical objects - in this case, sets - are conceived. Constructivists would reject Tucker's contention that since impredicative definitions do not give rise to contradictions there is no difficulty over their legitimacy.[2]

Similar constructivist arguments may be given for rejecting proofs involving the notions of 'all real numbers', 'all sets of positive integers', 'arbitrary set', 'arbitrary law', etc.

In all the cases so far considered there have been no purely arbitrary decisions on what is to count as a constructive proof. There are differences between constructivists as to what constitutes a constructive proof but the differences can be traced to the different ways that the mathematical objects are seen by them. But the fact that differences exist does not imply

---

1. It is odd that Quine can treat the problem so lightly. There is no harm in impredicative specification, he maintains, for 'we are not to view classes as literally created through being specified. .... The doctrine of classes is rather that they are there from the start. This being so, there is no evident fallacy in impredicative specification.' (W. Quine, <u>Set-Theory and its Logic</u>, Cambridge, Mass., 1963, p.243) The question here is, surely, whose doctrine of classes, Brower's? Wang's?

2. J. Tucker, 'Formalisation of Set-Theory' p.514

that they are arbitrary.

Tucker's notion of constructivity is not based on any previous view about the nature of mathematical objects. It is put forward to us as 'what can be carried out'. There is no attempt to expand this definition although he does give an example of a procedure which though it appears to be non-constructive turns out to be constructive and an example which though normally taken to be constructive (even by some intuitionists) turns out to be non-constructive under his definition.

(i)

I shall deal with the former example first. It is an attempt to show that diagonal procedures are constructive in his sense. The attempt depends on his analysis of the term 'indenumerable set'. 'Taken in the referential sense, the expression "indenumerable set" means an actual infinity which is greater than an actual denumerable infinity'[1] But there is another interpretation open to us, namely, 'a non-referential interpretation in which it means a set which contains an indenumerable element; where by an indenumerable element is meant an element which differs systematically from each element in an unending series whose generative recipe is given.'[2]

1. J. Tucker, 'Constructivity, Consistency and Natural Languages', p.156

2. ibid. p.156

It is difficult to make sense of this interpretation. What is needed is a clarification of the terms 'indenumerable element' and the term on which it depends, 'generative recipe'. From a later exposition that Tucker gives[1] it appears that D is an indenumerable element of S if D belongs to S and is different from each element of S given by some law determining an initial element and the successor of any element. Under this definition it will turn out that many sets thought of as denumerable will be indenumerable. (Perhaps both denumerable and indenumerable, but Tucker does not define denumerable.) Even the set of natural numbers would become indenumerable. For the generative recipe: - initial element 3, successor of an element x, x+1 - will give two indenumerable elements, 1 and 2. 1 and 2 belong to the set of natural numbers and yet differ from each of the elements given by the generative recipe. That the set of natural numbers is non-denumerable is an absurd consequence and makes nonsense of the distinction initially brought in by Cantor.

It might be said that my example ignores the fact that there is a generative recipe for the natural numbers and if I had taken this recipe then I should not have succeeded in obtaining

---

1. ibid. p.159

an indenumerable element. In general this would mean that there is only one proper generative recipe and the notion of a proper recipe would need to be defined. For non-denumerable sets (in the usual sense) there can be no proper generative recipe; in fact, there can be no generative recipes at all, for that is what the proofs of non-denumerability show.

Talk of an indenumerable element of a set prompts the question: which element is an indenumerable element? But the production of an element D, purported to be the indenumerable element, would result in our being able to give a generative recipe in which D would occur. (By taking D as the initial element and tacking on the other elements given by the original generative recipe which left out D.) A denumerable totality does not become indenumerable by adding one element.

Rather than talk of an indenumerable element, we could talk of an indenumerable element relative to a given generative recipe. Perhaps Tucker would then say that an indenumerable set would be one in which for any given generative recipe there remained an element of the set not included in the generative recipe. But does this mean that we have a 'non-referential' interpretation of 'indenumerable set'? The non-referential interpretation is the interpretation in which the sense of 'indenumerable set' is a set that contains an indenumerable an indenumerable element. In general this would mean that there is only one proper generative recipe and the notion of a proper recipe would need to be defined. For non-denumerable sets (in the usual sense) there can be no proper generative recipe; in fact, there can be no generative recipes at all, for that is what the proofs of non-denumerability show.

Talk of an indenumerable element of a set prompts the question: which element is an indenumerable element? But the production of an element D, purported to be the indenumerable element, would result in our being able to give a generative recipe in which D would occur. (By taking D as the initial element and tacking on the other elements given by the original generative recipe which left out D.) A denumerable totality does not become indenumerable by adding one element.

Rather than talk of an indenumerable element, we could talk of an indenumerable element relative to a given generative recipe. Perhaps Tucker would then say that an indenumerable set would be one in which for any given generative recipe there remained an element of the set not included in the generative recipe. But does this mean that we have a 'non-referential' interpretation of 'indenumerable set'? The non-referential interpretation is the interpretation in which the sense of 'indenumerable set' is a set that contains an indenumerable element. The only sense that can be given to this is an element which escapes every generative recipe. This is far stronger than Cantor's original definition and, if the Zermelo-König paradox is to be avoided, demands a precise definition of 'generative recipe'.

The non-referential interpretation that Tucker gives is unsuccessful. Again it shows a misunderstanding of constructivist objections to infinite sets. According to Tucker, the referential sense of 'indenumerable set' is 'an actual infinity which is greater than an actual denumerable infinity'. But the meaning of 'indenumerable set' in most set-theories, is given by some such definition as: a set which cannot be put in one-one correspondence with the set of natural numbers and which contains a subset which can. Its meaning is fixed by this definition. The definition does not mention 'greater than' or 'actual infinity'. It might be objected that even though this definition does not mention actual infinities it nevertheless presupposes them. In fact, the definition says nothing of whether the sets involved are infinite in the sense that they lie spread out before us in their entirity or in the sense that given any finite number of elements of the set there are yet others of the set.

The diagonal procedure which, Tucker says, has been regarded as non-constructive is, in fact, regarded as constructive by some intuitionists.[1] It is the conclusions drawn from the diagonal procedures that are regarded as non-constructive. Take, for example, the proof that the set of all sequences of positive integers is non-denumerable. First, suppose a correlation has been set up between the natural numbers and a set of sequences of positive integers. The usual diagonal procedure then gives a sequence which is not correlated to any natural number. It follows that the set of all such sequences cannot be correlated to the set of all natural numbers. Now, the intuitionist does not object to the diagonal procedure employed here. For, given a law which correlates the natural numbers with a set of sequences of positive integers, it is possible to construct a sequence of positive integers which is not correlated by the law. The construction needed is, of course, provided by the diagonal rule. It is the conclusion drawn from this to which the intuitionist objects. We cannot concude that the set of all sequences of positive integers cannot be correlated with the set of natural numbers since, he would say, it does not make sense to speak of all such sequences. The diagonal procedure is not rejected because it appeals to an actual infinity, as Tucker maintains.[2] Nor

---

1. see A. Fraenkel, Abstract Set Theory, Amsterdam 1961. p.55
2. J. Tucker, 'Constructivity, Consistency and Natural Languages' p. 159

is it regarded as non-constructive. The reason for rejecting such sets as the set of all real numbers, the set of all sets of positive integers, etc. is not just that the sets involved are 'actually' infinite. As with predicative definitions constructivists argue that a set must be defined by a rule or a law. A real number must be defined by a law - for intuitionists, spread laws.[1] Wang argues that the totality of laws is ill defined. We can have no 'clear and distinct idea of the totality of all sets or laws defining enumerations'.[2] In other words, the set of all sets of positive integers is non-constructive because each set of positive integers would have to be given by a law and we are never in a position to contemplate all laws, having knowledge of only a finite number at any time.[3] Similar reasoning applies to the set of all real numbers etc.

To say _anything_ of the set of all real numbers is non-constructive, so to say of that set that it cannot be correlated with the set of natural numbers is non-constructive. The fact that a proof of Cantor's theorem involves the diagonal procedure does not mean that the procedure is non-constructive. Cantor's theorem would still be non-constructive even if it involved only the intuitionist propositional calculus, for the very statement of the theorem is non-constructive in that it refers to a non-

1. A. Heyting, _Intuitionism - An Introduction_ p.34
2. H. Wang, 'The Formalisation of Mathematics' in '_A Survey of Mathematical Logic_', p.580
3. ibid, p.580.

constructive set. Tucker's plan to rehabilitate the diagonal procedure as a constructive procedure is unnecessary and if he is to rehabilitate Cantor's proof as constructive, he will need to show that such phrases as 'the set of all real numbers' are constructive. Since, however, he has defined only constructive procedures it is difficult to see how he can cope with what have been regarded as non-constructive entities.

These are hints of how he would deal with such entities in his discussion of the interpretations of '$\aleph_0$'. Again he refers to the referential and non-referential interpretations of symbols. On the referential interpretation, according to Tucker, '$\aleph_0$' stands for an actual denumerable infinity of elements. But there is a non-referential interpretation, he says, in which 'the function of a class symbol is to express the notion "the elements of the class such that ...." where the elements mentioned fall under the recipe for the generation of an unending series of elements.'[1]

Exactly what '$\aleph_0$' stands for can only be determined within some specified set-theory. There are theories in which '$\aleph_0$' is the set of all sets which can be put in one-one correspondence with the set of natural numbers.[2] Here '$\aleph_0$' does not denote

---

1. J. Tucker, 'Constructivity, Consistency and Natural Languages', p.158
2. e.g. Cantor, Russell.

a denumerable set at all, since it is the set of all denumerable sets, i.e. an indenumerable set. There are other theories in which '$\aleph_0$' will be a denumerable set[1] but because of such variety of usage it is unsafe to claim that '$\aleph_0$' has some one particular denotation.

As all that can be said with the use of transfinite numerals can be said without their use, there are no extra difficulties brought in for the constructivist by their introduction. It is not to the introduction of transfinite numerals that the constructivists object but to the sets which have transfinitecardinals. Tucker should then deal with the interpretation of 'N' (the set of natural numbers) rather than with '$\aleph_0$'. The remarks he makes about the interpretation of '$\aleph_0$' must be considered as if they are about 'N'.

Interpreted in this way, the function of class symbols, instead of referring to an actually infinite number of elements, expresses on Tucker's non-referential interpretation the notion 'the elements such that ....' where the elements fall under the recipe for the generation of an unending series of elements. The distinction here, if indeed there is one, is very fine. 'N' does not refer to the set of all natural numbers but expresses the notion 'the elements such that ...'.

---

1. e.g. von Neumann

It would seem that Tucker wishes to escape by taking 'N' as a way of talking about the elements of a set rather than the set. Now there are things which we wish to say about the set 'N' and other things which we wish to say of the elements of N. To say that each natural number has a unique representation as a product of primes is not to say that N has a unique representation. To say that each non-empty sub-set of N has a least member is not to say something of each element.

Perhaps Tucker means something other than the reading above by 'expresses the notion of' but I find it difficult to regard such a phrase as 'the elements such that ....' as a notion at all. I can understand 'the set of all elements such that' or 'being an element such that ....' as notions. But the first seems to be the referential interpretation of 'N' and the second not what would be meant by 'N' in any theory.

However, Tucker does talk of '$\aleph_0$' as being the class of entities generated from 0 by the successor operation despite his analysis of transfinite class symbols in terms of elements rather than classes.[1] In this case he says that it commits a category a mistake to ask for this class to be constructed. The reason given is that it is a class of classes of classes

1. ibid. p.158

whereas the elements of the class are classes of classes. (Again this is only true in certain theories) But just because '$\aleph_0$' (or 'N' to be safer) is of a different category from its elements it does not follow that all question of its constructivity does not arise. A strict constructivist could say that the class of all natural numbers is not constructible because, even though each natural number may be constructed (in principle), there will never be a time when we have constructed all of them. To a constructivist classes have to be constructed. The fact that the elements of a class belong to a different category from the class is irrelevent.

But the non-referential interpretation of class symbols that Tucker gives will not work for classes which are non-denumerable. There is no way of filling out the expression 'the elements such that ...' by any generative recipe giving an unending series of elements. What could be the generative recipe for the real numbers? That there is no such recipe is just what Cantor's theorem proves. Class symbols for non-denumerable sets cannot receive such an interpretation. If there could be an interpretation in terms of generative recipes then the constructivist who accepts as constructible sets given by a generative recipe would have no worries about the sets denoted by these symbols. One may look at the constructivist demand as a demand for generative recipes. It is the notion of a set not given by any generative recipe, the idea of an arbitrary set, that worries the constructivist.

It is not clear how infinite sets of greater cardinality are to be interpreted. Even if indenumerable sets could receive a non-referential interpretation in terms of containing an indenumerable element, there is no guide given by Tucker for finding a non-referential interpretation for such sets as the set of all real functions of a real variable. To show that this set is of greater cardinality than the set of real numbers it will not be possible to replace the usual 'diagonal' procedure by a procedure showing that there is a function different from each function in an unending series of functions given by some generative recipe. The most that the latter would show would be that the set of all real functions was indenumerable. Generative recipes are out of place here since the set of all real numbers itself is not given by a generative recipe.

Similar problems arise when sets of the same non-denumerable cardinality are considered. What is Tucker's constructive interpretation of the proof that the set of all real numbers has the same cardinality as the set of all continuous functions? We could perhaps show that both are indenumerable in Tucker's sense of containing an indenumerable element but, since neither is given under a generative recipe and Tucker's non-referential interpretation always mention the existence of a generative recipe, there would seem little chance of showing 'constructively' that they have the same cardinal.

Lastly, the axiom of choice is dealt with swiftly by Tucker. According to him the axiom is non-referential in character.[1] But it is not at all clear what Tucker means by non-referential in this context as he has only discussed 'non-referential' for the case of '$\aleph_0$' and 'indenumerable'. If he wishes to say that in addition to being non-referential it is also constructive - as it would seem from his allegation that it is the referential interpretation which makes the constructivist regard certain procedures as non-constructive[2] - he must introduce some other notion than that of generative recipes, for the axiom is needed precisely when there is no generative recipe. If there were a generative recipe for a set with the property stated by the axiom of choice then there would be no need of the axiom of choice. The axiom of choice is a purely existential axiom of the form $(\exists x)\, F(x)$. If there were a generative recipe giving a set with the property F it would follow from the predicate calculus alone that $(\exists x)F(x)$. The axiom of choice is needed only if there is no way of obtaining the set from the other axioms of set-theory.

1. ibid. p.157
2. ibid. p.156-157

In the case of a set of indenumerably many sets there could not be a generative recipe (in the sense of a recipe giving an unending series of elements) which gives a set containing just one element from each set since there are indenumerably many of them. But even in the case of a set of denumerably many sets the axiom of choice could not be dealt with in Tucker's terms for the existence of a generative recipe would imply that there is no need to invoke the axiom.

In conclusion, it would seem that each of Tucker's attempts to rehabilitate the non-constructive as constructive fails. Also, his approach ignores what seem to me the main problems that the constructivists bring to the fore. Since he sees their problems as arising from the doctrine that class symbols refer to actual infinities he misses the most interesting and clearest of their objections - their objections to impredicative definition, the notion of an arbitrary law and the notion of arbitrary set.

(ii)

The example of a constructive proof which Tucker says is in fact non-constructive is proof by reductio ad absurdum.

'All arguments to contradiction are non-constructive since the emergence of a contradiction shows that what has been tried in the given argument cannot in fact be carried out.'[1]

There are different arguments which may be labelled as arguments to contradiction, some of which even intuitionists accept. Consider an argument of the form: $\{p \supset (q . \sim q)\} \supset \sim p$. This form of reductio ad absurdum some intuitionists accept. (Indeed, as stated earlier in the chapter, $\sim p$ may only be asserted after having derived a contradiction from the supposition that the construction denoted by p has been carried out. Some intuitionists would say that this is what '$\sim p$' means.[2]) It would appear that this form of argument has certainly been accepted by constructivists. But there is another form which has been rejected by constructivists since it relies on the law of excluded middle. Consider an argument of the form: $\{\sim p \supset (q . \sim q)\} \supset p$. Clearly this is unacceptable on constructivist grounds, for the fact that a contradiction has been derived from the supposition that $\sim p$ entitles us to say only that $\sim p$ is absurd, i.e. $\sim\sim p$. We could move to p from $\sim\sim p$ only if we assumed some such logical rule as $\sim\sim p \supset p$ which is tantamount to assuming the law of excluded middle.

1. ibid. p.152

2. It cannot be quite as simple as this since $\sim q$ would have to be explained first, and so on. To break this infinite regress some intuitionists have two interpretations of negation. Kolmogorov speaks of a primary interpretation in terms of the incompatibility of a subject with a predicate. Brower's notion of absurdity could then be defined in terms of this primary interpretation. See Kolmogorov, 'On the Principle of Excluded Middle' (first) published in 1925) included in From Frege to Gödel, ed. J. van Heijenoort, Cambridge, Mass., 1967. pp.420-421

Constructivists have certainly objected to the second of the two schemata mentioned above[1], but in general they have accepted the first.[2]

It remains to investigate why Tucker regards the first as non-constructive. His explanation rests on the idea that the appearance of a contradiction shows that 'what has been tried cannot be carried out'. In geometry one might, I suppose, talk in a rather imprecise fashion of 'trying to construct two tangents at the same point on a circle' and, from the contradiction that results from supposing this to be done, say that what we tried to do cannot in fact be carried out. Elementary geometry text books may be written in such language. To do so is to treat geometry as a description of the physical world and reductio ad absurdum proofs look as though they report that certain lines cannot be drawn etc. Such a view of geometry has long been abandoned.

Talk of 'construction' in arithmetic, analysis or set theory remains metaphorical unless backed up by some definition or explanation. Tucker speaks of 'constructive procedures'

1. e.g. R. Goodstein, 'Proof by Reductio ad Absurdum', Mathematical Gazette, vol xxxii, 1948

2. Some intuitionists reject the whole idea of negation in mathematics, so that reductio ad absurdum as a legitimate proof procedure would be rejected. But their arguments are not directed against the reductio ad absurdum procedure in particular See the discussion of Griss's and van Dantzig's attitude in Fraenkel and Bar-Hillel, Foundations of Set-Theory, pp.239-244.

and 'non-constructive procedures', of 'procedures which can be carried out' and 'procedures which cannot be carried out'. But what 'procedures' are there in mathematics? Mathematics consists of proofs. Are procedures supposed to be different from proofs?

In a reductio ad absurdum proof what is it that I 'try' and that I find cannot be 'carried out'? Both phrases suggest that it is some kind of action. What happens in a reductio ad absurdum proof is that I suppose something to be the case and find that what I supposed cannot be the case. There is no mention here of something that I try to do and find that I cannot do. It is true that I indulge in the activities of proof-making and supposing. But it is neither of these activities that I try and find that I cannot carry out. For I have successfully carried out the proof and, although what I supposed turns out to be impossible, it does not follow that I cannot suppose what I did suppose. If there is something else in the reductio ad absurdum proof which I tried and found I could not carry out Tucker has given no hint of what it might be.

Apart from this difficulty, there remains the problem of finding out when a proof is of the reductio ad absurdum form. A discussion of reductio ad absurdum proofs by Goodstein[1.] will illustrate this problem. Goodstein, disliking reductio ad

1. R. Goodstein, 'Proof by Reductio ad Absurdum'

absurdum arguments because of the lack of information that they give, tries to give direct proofs of theorems normally proved by this method. As an example he considers a direct proof of the theorem that the square root of 2 is irrational.

Starting from the fact that for all positive integers p and q $|p^2 - 2q^2| \geq 1$, it follows that $|p^2/q^2 - 2| \geq 1/q^2$. This, he says, is a direct proof that 2 is not the square of a rational number. But this latter statement is surely an inference made from the above inequality. It may follow almost immediately but an inference does have to be made nevertheless. The inference, it seems to me, that has to be made here will be made in the following way. Suppose that 2 is the square of a rational number p/q. Then $p^2/q^2 - 2 = 0$. Therefore $|p^2/q^2 - 2| < 1/q^2$ which contradicts the above inequality. In other words, the proof that the square root of 2 is irrational still needs a reductio ad absurdum proof. Goodstein has not shown conclusively that the use of reductio ad absurdum in this example is unnecessary.

When proving theorems in an informal way, without reference to any axiom systems, it is difficult to say when reductio ad absurdum has been used. In the above example there is no indication of what we are allowed to assume. An axiomatisation of arithmetic would settle this. If, among the axioms, there occurred the schema '$a > b \supset a \neq b$' then there would be no need to employ reductio ad absurdum. Suppose, instead of this schema, one of the axiom schemata was

'$\sim(a > b . a = b)$'. Then the proof would continue as indicated in the previous paragraph. Whether we have to use reductio ad absurdum can be decided only after we have laid down an initial set of assumptions.

Tucker, as can be seen from his talk of 'trying' and 'carrying out', takes reductio ad absurdum in its rule form rather than in its schematic propositional form. That is, in the form:- if a certain hypothesis leads to a contradiction then the negation of that hypothesis holds. Without entering too deeply into the technical details of the propositional calculus, it may be pointed out that this rule corresponds to the rule of the propositional calculus:- if there is a hypothetical proof '$s \vdash t . \sim t$' then there is a categorical proof of '$\sim s$'. In most systems of the propositional calculus this will be derived as a subsidiary rule from the axioms and rules of the system. But the import of the rule is that a categoric proof of '$\sim s$' can be found whenever we have found a hypothetical proof of the form '$s \vdash t . \sim t$'. In other words we can prove categorically from the axioms alone the formula '$\sim s$' without the use of any hypothesis. Suppose some mathematical theory is formalised within the first order predicate calculus. It is a short cut to use the derived rule of reductio ad absurdum in order to prove a proposition $\sim P$ of that theory. Nevertheless, there will be a categoric proof of $\sim P$ which will not involve using P as a hypothesis in a deduction. If we now consider the categoric proof of $\sim P$, what is it that has been tried and cannot be carried out in this proof? Can it make any sense at all to talk of the categoric proof in this way?

Tucker may still object that, even though categorical proofs as opposed to hypothetical proofs do not involve suppositions or hypotheses from which a contradiction can be derived, reductio ad absurdum has been used implicitly in the sense that the proposition $\{P \supset (Q.\sim Q)\} \supset \sim P$ or axioms from which this proposition can be derived have been used in the categoric proof. This may be the case. Tucker then has to show that these axioms themselves are non-constructive. We may, for example, prove $\{P \supset (Q.\sim Q)\} \supset \sim P$ from the two axiom schemata $\sim(P.\sim P)$ and $(P \supset Q) \supset (\sim Q \supset \sim P)$ added to suitable axioms for the logical connectives '$\sim$' and '.'. Which of these axioms is non-constructive? Which of these does it make sense to talk in terms of 'trying' to do something and finding that it 'cannot be carried out'?

Non-constructive procedures in general Tucker regards as impossible of execution because their execution would require the contravention of already accepted constructive conditions.[1] As I have argued above I am unhappy about Tucker's use of 'procedure' in the context of mathematical proofs. It is clear from his use

1. J. Tucker, 'Constructivity, Consistency and Natural Languages' p. 153

of the word that it is not a synonym of 'proof'. Yet it is difficult to see what he can mean if it is not 'proof'.

Reductio ad absurdum proofs, since they are non-constructive, cannot be admitted as proofs proper, he says, but should be regarded as arguments. They may provide us with the only information that we have. But they should be regarded only as 'temporary scaffolding' from which we may later construct a proofs proper.[1]

The point that arises here is how these arguments manage to provide any information when they can only be made by breaking rules. 'The impossibility of the procedures is of a rule-breaking character'.[2] It seems odd, if not inconsistent, to maintain that certain rules have been broken and yet that information is provided by breaking those rules. If information and correct information at that, for Tucker nowhere suggests that the information so given is wrong, can be gained by breaking the rules, what possible purpose do the rules serve? One would expect to get misleading information, in some cases, from rule-breaking, just as fallacious reasoning would produce, in some cases, incorrect consequences.

From the arguments presented in (i) and (ii) it can be seen

1. ibid. p.155
2. ibid. p.153

that Tucker has failed to make his new distinction between constructive and non-constructive procedures clear. It is not at all clear how the term 'procedure' itself is to be understood. It is true that we sometimes speak of Cantor's theorem as involving the diagonal procedure. But this way of speaking is harmless. It means only that a certain way of defining a particular object has been used. The intuitionists can make their notion of constructive proof clear in the examples they give. To say, for example, that an existence theorem in arithemtic has a non-constructive proof is to say that the proof does not tell us how to compute a number with the required property. Tucker's notion of constructivity is made no clearer by the examples he gives. It is essential that he makes this notion comprehensible if he is to go on and maintain that natural languages are constructive in tendency.

What is meant by a natural language being constructive in tendency is even more obscure than of procedures being constructive. The only evidence he gives for this conclusion about natural languages is that the paradoxes of set-theory are produced by breaking constructive conditions. He gives the Russell paradox as an example. The condition '$\sim a \epsilon a$', he maintains, is a constructive condition, as is '$a \epsilon k$ if and only if $\sim a \epsilon a$'. What is meant by a constructive condition is never enlarged upon. The solution of the paradox is then given as described in chapter II of this Section.

Since the paradoxes are generated by breaking constructive conditions, natural languages, in which the paradoxes can be expressed, must be constructive.

Because of the undefined notion of a language being constructive, it is difficult to see what indeed has been established by this argument. Even if the phrase 'constructive condition' were defined it would still be difficult to see what Tucker means by 'a natural language being constructive in tendency'. It seems to Tucker that this is 'a significant discovery and one which is contrary to Tarski's thesis about natural languages'.[1] But it is hard to see what the discovery is or how the discovery contradicts Tarski. Certainly Tarski says that natural languages seem to preclude a consistent use of the expression 'true sentence'[2] and, further, that natural languages must be inconsistent.[3] But there is no indication in Tarski's paper of what it would mean to say that a natural language is constructive (or non-constructive). Consequently, reference to Tarski's paper fails to clarify Tucker's contention.

---

1. ibid. p.145
2. A. Tarski, 'The Concept of Truth in Formalised Languages' in <u>Logic, Semantics, and Metamathematics</u>, Oxford, 1956, p.165
3. ibid. p.164-165

V

Throughout Tucker's papers there is constant reference to his view that the contradictions have to be explained.

'The paradoxes have to be explained, they have to be fully understood, and the manoeuvres of formalisation cannot provide any such information or insight'[1]

'.... in any satisfactory account of the paradoxes of set theory a foundational account must be an explanatory account. For this reason the usual devices for avoiding the paradoxes of set theory are unsatisfactory since they do not satisfy the explanatory requirements of the foundational level.'[2]

His insistence that the paradoxes have to be explained shows that he believes there is an explanation. What sort of explanation is made clear by his purported explanations criticised in chapter II of this section. The explanations will be in terms of the rules of language. Now these rules of language must already be embeddied in the language before the appearance of the contradictions. The rules which have been laid down by philosophers and logicians to prevent the occurrence of contradictions he regards as evasions and not explanations. They are rules designed solely for the purpose of avoiding the contradictions. Since this is the

---

1. ibid. p.165
2. ibid. p.150

only reason for the rules' existence they lack any explanatory force. Instead of such ad hoc rules Tucker says that we must find rules which can be seen to hold before the contradictions arise.

'There is the view that the appearance of a paradox is quite unpredictable, that nothing can be done beforehand, that we just have to wait for them to turn up and then avoid them.' This, Tucker claims, is not the case. Instead, we can investigate the constructive working conditions of these words prior to, and independently of, the appearance of contradictions. The view that the paradoxes are unpredictable is irrational since we could in each case have avoided the contradictions by giving due attention to the constructive working conditions of the words involved.[1]

The paradoxes can always be explained by drawing attention to the linguistic rules of language. Underlying this thesis is the thesis that natural languages are consistent; that the rules of language never give rise to contradictions. He offers solutions to the paradoxes by locating a linguistic rule which has in some way been broken in the 'proof' of the supposed paradox. If natural languages did produce contradictions, and, in particular these contradictions as has been maintained in the past (for example by Tarski[2]), then Tucker's search for an explanation would be totally

1. ibid. pp.164-165
2. A. Tarski 'Concept of Truth in Formalised Languages'

misplaced. I have tried to show in a previous chapter that Tucker's attempts to solve the paradoxes are each unsuccessful. He has not, so I maintain, located a linguistic rule which has been broken. How does the thesis that natural language is consistent stand up? If it could be shown that such a language is inconsistent without breaking any linguistic rule, then we should have less reason to continue looking for 'explanations' of the paradoxes.

In another paper[1], Tucker claims that 'formalisers reject informal language because it gives rise to contradictions. Yet there is no evidence whatever for their view.' He goes on to demand

'... an example of an intralinguistic contradiction which is obtained by conforming to the working conditions of a natural language. Formalisers do not back up their faith with mere examples. They are committed. They do not look at the facts.'[2]

If the phrases 'working conditions' and 'linguistic rules' are interchangeable, this is the underlying assumption that Tucker has been making throughout his other papers.

The inclusion of the word 'intra-linguistic' in the above quotation succeeds only in confusing the issue. According to Tucker 'extra-linguistic' contradictions can occur in natural languages without a breakdown of working conditions, though it is

1. J. Tucker, 'Philosophical Argument' Supplementary Volume XXXIX 1965, The Aristotelian Society.
2. ibid. p.57

by no means clear how this comes about. The only example of an extra-linguistic contradiction that he gives is of a man who says that it is raining and it is not raining. Both the assertion that it is raining and that it is not raining have empiric content. They assert contradictory things about an extra-linguistic state of affairs. 'Each has content. Each is well-used. They simply contradict each other head on.'[1] In this example I cannot see how, at the same time, this can be a contradiction and for both expressions to be well-used, for either it is meant as a report of light-drizzle, in which case it is not a contradiction or one of the expressions is not well-used, at least in any sense of 'well-used' with which I am familiar.

If 'extra-linguistic' and 'intra-linguistic' are to be distinct mutually exclusive categories into which we can divide propositions and in particular contradictory propositions, then we need more of a guide than is given by one example. Into the intra-linguistic category Tucker wishes to put the set-theoretic paradoxes and into the other every contradiction which has not been labelled a paradox. If we did not know that this is the division he wants we would not be able to see which propositions belonged to which category. It may, however, be the case that the distinction is between a priori and empirical propositions, but

---

1. ibid. p.58

this would be unlikely as this would render the two new terms superfluous. In order to see the inadequacy of the purported distinction, has the man who says that two and two are four and two and two are not four made an extra-linguistic or an intra-linguistic contradiction? Certainly both are well-formed and each has content. So it would appear from Tucker's example that it is extra-linguistic, but so too would Russell's paradox appear as an extra-linguistic contradiction. Clearly, he would like to separate out the arithmetic contradiction from the paradoxical ones of set-theory, but he has not given any criterion to enable us to do so.

Leaving aside the question of the precise meaning of 'intra-linguistic', we can return to the thesis contained in a previous quotation:- that there is no evidence that contradictions occur in ordinary discourse when conforming to the working conditions of that language and that formalisers overlook this fact.

For any one committed to the belief in the consistency of natural languages, as Tucker is, there is no way of refuting him. Each time an apparent contradiction turns up which does not seem to violate the working conditions of that language, it is always open to him to say that although it does not seem to violate any of the working conditions that we have found, nevertheless it does violate some condition, but it just happens that we have not found it yet. The thesis is irrefutable. Unlike formal languages where the 'working conditions' are laid down in advance in the form of rules of inference, informal languages have to be inspected after their use for their working conditions. We have no guarantee at any time that we have found all these conditions and that since a given contradiction does not break any of these conditions it must be a contradiction which does not violate any working condition of that language.

The reason why some philosophers have rejected such a thesis in the past is the existence of several contradictions which do not seem to break any linguistic rule. The existence of these contradictions is the fact which formalisers look at. They are the evidence which formalisers produce. To accuse them of not producing evidence and of not backing up their faith with 'mere examples' is to ignore the amount of research into the paradoxes in the last sixty years. If some philosophers maintain that informal language is inconsistent then it is because the 'explanatory' solutions offered in the last sixty years fail to satisfy them. The justification such a philosopher would give then for saying that informal language is inconsistent, though not of course conclusive, is reasonable and not the result of an irrational belief.

In passing, it may be noted that many of those Tucker refers to as formalisers have themselves offered solutions of the paradoxes in informal terms and maintained that they do arise from the violation of some implicit linguistic rule. Even the formalists themselves have given an explanation of the paradoxes in terms of an unjustified extension of the usual rules of logic from finite domains to infinite domains.

'Does material logical deduction somehow deceive us or leave us in the lurch when we apply it to real things and events? No Material logical deduction is indispensible. It deceives us only when we form arbitrary abstract definitions, especially those which involve infinitely many objects. In such cases we have illegitimately used material logical deduction; i.e. we have not paid sufficient attention to the preconditions necessary for its valid use.'[1] This quotation shows that even Hilbert, the foremost formalist, believed that contradictions occurred only when the rules implicit in the language were forgotten. Indeed, if the word 'preconditions' in the above quotation were to be changed to 'working conditions' then its last sentence would not look out of place in Tucker's paper.

Russell also is included in Tucker's list of formalisers and the theory of types which Russell devised to deal with the paradoxes Tucker regards as an evasion and not an explanation. Now, although Russell does say that the main recommendation for the theory of types is that it solves the paradoxes he also believes it to conform with common sense.[2] But behind the theory of types

1. D.Hilbert, 'On The Infinite', in Philosophy of Mathematics, ed. P. Benacerraf and H.Putnam, Oxford, 1964.
2. B. Russell, Principia Mathematica, Cambridge, 1913 p.37

there is the vicious circle principle which is the justification for the theory. The vicious circle principle is derived in turn from the principle that in a definition the definiendum must not appear in the defininiens, which has been considered a sound principle from the time of Pascal at least, and may be found in elementary logic textbooks. It is true that belief in the principle does not by itself lead to a theory of types. Hintikka's more recent work on applying the vicious-circle principle in its simplest form does not lead to the theory of types.[1.] The theory of types cannot be derived from the vicious circle principle alone: it needs Russell's analysis of classes in terms of propositional functions, for example.

The reasons for accepting the theory of types as put forward by Russell are philosophical. The theory of types was not just an evasion but an outcome of the vicious-circle principle and Russell's philosophical doctrine of propositional functions. The subsequent rejection of Russell's theory of types by sympathetic logicians was caused not by the lack of any philosophical justification for the theory but by the unsatisfactory nature of the axiom of reducibility and the doctrine of propositional functions.

---

1. J. Hintikka, 'Identity, Variables and Impredicative Definition', Journal of Symbolic Logic, 21, pp.225-245

The so-called formalisers have looked at the facts and have presented examples. It would seem that the implicit rules of informal language do give rise to contradictions. The set-theoretic paradoxes provide good evidence for this. Their existence cannot of course show conclusively that informal language is inconsistent, i.e. that the implicit rules of language allow a situation to occur where two sentences one of which is the negation of the other both appear to have the same truth value. For the rules of informal language are not open to our inspection as the rules of formal languages are. We may inspect a formal language and show conclusively that it leads to a contradiction, the rules of inference and any axioms that it may have are precise and explicit. For natural languages the rules have to be found and even then precision cannot be expected. One may draw the analogy between extracting the rules of natural languages and extracting the rules of a game from the observation of the game itself. If the only guide to the rules of that game was our observation of that game then we could never be sure that the rules we had extracted were the complete rules of the game, nor that any of the extracted rules corresponded with precision to any actual rule of the game. For there may always be the possibility that some rule has not been employed while the observer was watching and that the rules have the disjunctive form; do A or do B or .... for some finite number of possibilities A, B, ...

Since the observer may only watch for a finite time the disjunctive rule may have been employed for doing only a finite number of these possibilities and at no time could the observer be sure he has listed them all.

The analogy breaks down when we consider that, unlike the hypothetical observer, we are not only extracting the rules but at the same time playing the game, Nevertheless, the point brought out by the analogy is that it cannot be proved that informal language is inconsistent. The most that we can say is that there is evidence for this conclusion.

It may be that Tucker believes that not only is it the case that natural languages are consistent but also that they must be consistent. But if it is correct to talk of language rules, as Tucker does, then it does seem possible that these rules could conflict. The rules we use have been made by us, and, as we are unable to see all the consequences of these rules straight away, it may turn out that they conflict. In designing a game or a system of laws we may find that the rules or the laws are such that direct us to do contradictory things. In chess, for example, there are two rules, one of which says that the king must be moved out of check and another which says that the king must not move into check. On certain occasions these rules conflict; they conflict when a position of checkmate is reached. Suppose that winning the game of chess consisted, not in forcing checkmate, but in removing all the opponent's pieces from the board. Then the rules would create an impasse when the checkmate position was reached. If language is thought of in this way, as a system of rules which we have made, then it seems not unlikely that such conflicting rules should occur. Perhaps the classic example of this is Prior's introduction[1] of the propositional connective 'tonk' by means of two rules of inference. From a proposition A one can derive the proposition A-tonk-B and from the proposition A-tonk-B one can derive the proposition B. Consequently, from A one can derive the proposition $\sim$A. The rules for the connective 'tonk' are such that two contradictory propositions can be derived. It is, of course, easy to see that these two rules allow the derivation of contradictory propositions, but the example does show that in talking of language rules the possibility of rules which allow contradictory propositions to be derived may exist.

In the case of the set-theoretic paradoxes one could regard the axiom of comprehension (in its naive form) as a rule for the introduction of the phrase 'a-belongs-to-b'. What is shown by the appearance of paradoxes is that one cannot adopt such a rule (along with others) without falling into inconsistency, just as one cannot adopt the rules for 'tonk' without falling into

1. A. Prior, 'The Run-about Inference Ticket', Analysis, vol.21, 1960, pp.38-39

inconsistency.

As I have argued earlier in section 1 of this thesis, the appearance of contradictions in set-theory reveal only that we cannot operate consistently with the axiom of comprehension in its naive form. We are forced to change the rules for the use of the word 'set'.

To ask, as Tucker does, for some explanation of why certain purported sets lead to contradictions is as futile as asking for an explanation of why there is no greatest prime number.

Since each paradox of set-theory can be stated in a natural language there may be a temptation to think of solutions of paradoxes in terms of spotting a fallacy, rather as one spots that an arithmetic contradiction is produced by a fallacious move of dividing by 0. To spot such a fallacy it is necessary to recognise the rule that division by 0 is illegitimate. To spot some fallacy in set-theory we need to recognise the rule that has been illegitimately disregarded. In what sense there are such rules in abstract set-theory has been discussed in section 1. There I tried to show that the rules of abstract set-theory were our own creation, utilising certain analogies from a pre-formalised notion of collections of objects. A first attempt at set-theory included the axiom of comprehension. It produced contradictions in the field of abstract sets. Consequently it was necessary to abandon the axiom if we desired a consistent system. The rules introduced for the word 'set' would have to be revised. To look for explanations in terms of ordinary linguistic rules seems to me a mistake as they could at most guide us in choosing the axioms of set-theory. In abstract set-theory we may use the words 'set' and 'class' but it must be remembered that we have chosen to use them in a certain way, the way laid down by the axioms we have chosen.

Perhaps the simplest solution in terms of linguistic rules would be that the phrase 'abstract set' is itself illegitimate for whenever we speak of sets we must speak of sets <u>of</u> something, e.g. numbers, students, chairs etc. To talk of 'abstract sets' is to forget this rule, to think that there could be sets which are not sets <u>of</u> something.[1] But this demolishes not only the paradoxes but the whole edifice of set-theory. There would be no paradoxes of set-theory since there would be no set-theory.

1. I am not suggesting that this is a rule of language, although it seems to me just as acceptable a rule as those cited by Tucker in his solutions of the paradoxes.

Section 3

In this section I shall be concerned with Tarski's definition of truth for formalised languages. This definition is the subject of his two papers, 'The Concept of Truth in Formalised Languages'[1] and 'The Semantic Conception of Truth'[2]. In the former, Tarski constructs a definition of truth for a formalised language and explains how, and within what limits, this definition may be modified for other formalised languages. In the latter paper, the construction is only outlined but there are, in addition, replies to various criticisms made of Tarski's definition.

Whenever a definition is given for some stated purrpose, one method by which that definition can be judged is to see if it does achieve the purpose intended. Tarski's definition is in this catagory since he states the aim of his definition and the conditions that it must satisfy. In the examination that follows I shall show that he has not succeeded in constructing a definition which accomplishes the task that he has set for it. To show this, is not to show that the definition is either wrong or valueless. A definition may be regarded, for example, as a proposal to treat the definiendum as a synonym or an abbreviation of the definiens; the definition may then be accepted or rejected on other grounds than

1. A. Tarski, 'The Concept of Truth in Formalised Languages' included in Logic, Semantics and Metamathematics, trans. Woodger, 1956, pp.152-278

2. A. Tarski, 'The Semantic Conception of Truth', Philosophy and Phenomenological Research, vol. 4 (1944).

accomplishment of purpose. It may be held that acceptance of such a proposal might lead to a confusion or that the definition is fruitless because it allows few or no relevant consequencies to be drawn. Considerations like these may enable the definition's worth to be evaluated and, in general, they will be independent of those concerned with its satisfaction of the author's purpose or purposes. In the present section I shall leave aside all considerations that do not directly affect the question of whether Tarski's definition of truth fulfils, or fails to fulfil, his programme.

I

By means of quotations from his papers, I shall begin by isolating the purpose Tarski's definition of truth has to fulfil.

'The present article is almost wholly devoted to a single problem - the definition of truth. Its task is to construct - with reference to a given language - a materially adequate and formally correct definition of the term "true sentence". This problem, which belongs to the classical questions of philosophy, raises considerable difficulties.'[1.]

'The desired definition does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary, it aims to catch hold of the actual meaning of an old notion.'[2.]

Tarski elaborates further this 'old notion':

'... throughout this work I shall be concerned exclusively with grasping the intentions which are contained in the so-called classical conception of truth ("true - corresponding with reality") ...'[3]

'We should like our definition to do justice to the intuitions which adhere to the classical Aristotelian conception of truth - ...: To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.'[4]

---

1. A. Tarski, 'The Concept of Truth in Formalised Languages', Introduction. I shall refer to this work as CTF.
2. A.Tarski, 'The Semantic Conception of Truth', section 1. I shall refer to this work as SCT.
3. CTF, Introduction
4. SCT, section 3.

The definition of truth must be conformable to this classical conception of truth, if the definition is to fulfil the aims Tarski has set for it. According to Tarski the definition must have as consequences, such equivalences as the following:

"Snow is white" is true if and only if snow is white.

He maintains that in this equivalence "Snow is white" occurs as a name of a sentence and not as a sentence itself since the subject of "is true" can only be a noun or an expression functioning like a noun.

Tarski holds that the problem of constructing a definition conformable to the classical conception of truth becomes the problem of constructing a definition the consequences of which will be equivalences of the form:

X is true if and only if p.

In these equivalences, "p" will be replaced by a sentence and "X" by a name of that sentence.

For natural languages such as English, the construction of a definition which will imply consequences of the above type raises several problems. One of these is the difficulty which is produced by generalising such sentences as the following: "Snow is white" is true if and only if snow is white. The natural generalisation would be a sentence of the following form:

for all p, "p" is a true sentence if and only if p.

("p" is here a name of the sentence "p"). The difficulty lies in the function of names in such sentences, for if names like "Snow is white" are treated as syntactically simple expressions (like single words of a language) then parts of the name may not be replaced, just as parts of a word (the letters) may not be replaced in a natural language. Under these conditions, "p" denotes the letter of the alphabet, p. Consequently, the sentence 'For all p, "p" is a true sentence if and only if p' will have such implications as '"p" is true if and only if it is snowing.'. Clearly, this treatment of quotation-mark names leads to undesirable results. Similar considerations applied to other forms of names force Tarski to give up the attempt to construct a definition of truth for a natural language. Apart from the difficulties entailed by the function of names in such a definition of truth, there occurs in the application of the term "true" in a natural language a variant of the 'liar' antinomy. In view of these problems Tarski turns his attention away from natural languages to formalised languages.

For such formalised languages Tarski tries to construct a definition of truth, consequences of which he desires to be sentences of the following form:

X is true if and only if p.

(Here "X" is a name of the sentence "p".)

It is seen that this attempt is analogous to the previous attempt to construct a definition for a natural language. Tarski contends that for some formalised languages, a definition which would fulfil the above condition can be constructed and moreover that it is possible to give a precise condition, which these languages must fulfil if they are to allow of such a definition to be constructed for them.

Tarski then constructs a definition for one formalised language in some detail and shows how other definitions of truth may be constructed for other formalised languages.

The above is an outline of Tarski's paper 'The Concept of Truth in Formalised Languages', the details of which I shall consider later. At present it will be sufficient for my purpose to extract from this outline the several aims that he has met for his definition.

Firstly, the definition and the investigations relating to it should be concerned with concepts dealt with in classical philosophy i.e. be such that they have philosophical value and not only technical value.[1.]

Secondly, the definition should be conformable to the classical conception of truth, i.e. "true - agreeing with reality".[2.]

Thirdly, the definition should have as consequences sentences of the following type: X is true if and only if p, in which "p" is a sentence and "X" is a name of that sentence.[3.]

Fourthly, the definition should satisfy several formal conditions, e.g. the definiens should be in terms whose sense is precisely known, or in terms which are reducible to other known terms.[4.]

---

1. CTF, Introduction. SCT, paragraph 3.
2. CTF, Introduction. SCT, section 3.
3. CTF, section I; section 3, paragraph 4.
4. CTF, Introduction. SCT, section 1.

Now, if the investigations are to have any concern with the philosophical problems of truth then that part of the papers which is concerned with defining truth for formalised languages should not only be a concern of mathematicians but a concern of philosophers, since this part is the main subject of the paper. Implicit in the first aim is that the definition of truth for formalised languages should be of some philosophic value, since it is not to be supposed that classical philosophy has been concerned merely with a technical term related to formalised languages only, the study of formalised languages being of more recent origin than classical philosophy. Implicit in the second aim is that the 'correspondence' theory (or whatever Tarski understands by this term) should be applicable to formalised languages, for there would be little support gained for his contention, that he is dealing with a problem that has occupied philosophers, if he were constructing a definition of truth conformable to a theory which is inapplicable in the domain of formalised languages.

The third aim presupposes that consequences of the type X is true if and only if p, should be conformable to the Aristotelian conception of truth. But Tarski does not elaborate on what he understands by 'conformable'. Presumably, he wants these consequences to be consistent with the Aristotelian conception of truth, i.e. such that acceptance of the Aristotelian conception of truth implies acceptance of the schema: X is true if and only if p, where "X" and "p" are replaced according to the conventions mentioned above.

The fourth aim is of a formal nature and its fulfilment may be determined by inspection of the definition Tarski constructs.

In connection with the second and third aims listed above, the definition should have no consequences which are in conflict with the concept of truth with which Tarski is concerned. For, if the definition is to be conformable with (which is interpreted here as 'consistent with') such a conception of truth, then, besides giving consequences of the form X is true if and only if p, the definition should not imply sentences which would be unacceptable to anyone allowing those consequences.

In the past criticisms of the definition have been directed at the suitability of Tarski's procedure for natural languages and have ignored the prodedure for formalised languages. After a discussion of these criticisms I shall show that Tarski has not fulfilled all of the above aims for the more limited domain of formalised languages.

II

In this chapter I shall consider criticisms of the definition made by Black, Strawson and Kneale.

Black has argued[1] that Tarski's definition of truth for formalised languages would be inapplicable to any natural language and that consequently Tarski's investigations are without philosophic relevance.

Firstly, he maintains, Tarski's definition would necessitate a complete enumeration of the terms occurring in the language for which the definition is being constructed. But natural languages have an 'open' character, that is, they can have added to them new terms. There would have to be some rule that would stipulate that no new terms were to be introduced into those languages, if a definition based on Tarski's procedure were to be possible. The fact that there is no such rule would condemn any such attempt to failure.

Secondly, Tarski's definition of truth applies to only one language at a time. In other words, Tarski does not give a definition of truth in general, but gives a definition of truth for a language, L_i, say. If the definition of truth for L_i is known, then how is a definition of truth for another language L_j to be constructed? To extend the principle of the definition to another language, Black says, it is necessary to understand that principle.

---

1. Max Black, 'The Semantic Definition of Truth', Analysis, Vol. 9. No.4, 1948, pp.49-63

But to <u>state</u> that principle will only be a crude reformulation of the sentence: <u>For all p, "p" is a true sentence if and only if p</u>. This sentence was found to be unacceptable by Tarski because of the difficulty involved by the function of the name "p" . It is impossible to state a general criterion or definition of truth by Tarski's procedure. Black contends that the philosopher is searching for a <u>general</u> criterion for truth.

Thirdly, the philosophic problem of truth is left untouched by the definition that Tarski proposes, since adherents of the correspondence, coherence and pragmatist theories of truth would all agree (subject to certain qualifications) that "it is snowing" is true if and only if it is snowing.

Black's conclusion is that Tarski's procedure has no philosophic relevance, since it is inapplicable to natural languages and also neutral to conflicting theories of truth. This conclusion, however, does not follow from Black's arguments.

Assuming his arguments are valid and he is entitled to say that the definition cannot be applied to natural languages, that a general criterion for truth based on Tarski's lines cannot be stated consistently and that the definition of truth which Tarski gives is neutral to conflicting theories of truth, do these statements imply the conclusion that the investigations of Tarski are without any philosophical relevance? It very much depends on what Black understands by 'philosophical relevance'. If he regards any definition as without philosophical relevance which is neutral to conflicting theories of truth, then there does seem to be some support from his arguments for his conclusion. But to say that the definition is neutral to mutually inconsistent theories, because those theories would all accept sentences of the following form: "It is snowing" is true if and only if it is snowing., is incorrect. For there may be other consequences of the definition which are incompatible with these theories. For example, there is the conclusion that Tarski's definition implies: 'It turns out that for a discipline of this class (a very comprehensive class of mathematical disciplines) the notion of truth never coincides with that of provability."[1]. It may be the case that this consequence might be incompatible with some theory of truth; it is not sufficient for Black's conclusion to examine just those consequences represented by the schema, X is true if and only if p.

If Black understands by 'philosophical relevance' the relevance of the definition for natural languages, in which case consequences of the definition of the type quoted above, having reference to formalised languages only, would be ignored, then it still seems that the conclusion does not follow. For even though Tarski's procedure may not be applied to natural languages, there are consequences about them which can be infered

---

1. SCT. Section 12.

from his investigations. One such consequence is that no definition can be constructed, consistent in the language, which will imply all sentences of the form: X is true if and only if p. Again, whether this has 'philosophical relevance' depends on what Black means by that term, but Black's paper contains sentences of this same negative type of which he does not wish, presumably, to deny 'philosophical relevance'. Also, the claim that Tarski's investigations have only philosophical relevance if his procedure is applicable to natural languages seems to be an unwarranted restriction. If, as is the case, a term such as 'true' is used in connection with sentences of a formalised language as well as with sentences of a natural language, then there seems to be no reason why this term should not be of philosophical interest. In the same way, a philosopher may consider some term which is used mainly in connection with science, e.g., 'theory', 'hypothesis', 'model', without ceasing to be a philosopher. In addition, there is the philosophy of mathematics which, in general, is not concerned with applications to natural languages.

The arguments which Black proposes are not ones which Tarski would contradict. The first argument, that Tarski's procedure is inapplicable to natural languages, Tarski has made himself. Tarski contends that it is because of the indefinite character of a natural language which makes any definition of truth, materially adequate and formally correct in Tarski's sense, inapplicable in a natural language. '<u>The problem of the definition of truth obtains a precise meaning and can be solved in a rigorous way only for those languages whose structure has been exactly specified</u>.'[1]

Secondly, Tarski does not claim that his definition has anything to do with the philosophical problem of truth he writes: 'In general, I do not believe that there is such a thing as "the philosophical problem of truth". I do believe that there are various intelligible and interesting (but not necessarily philosophical) problems concerning the notion of truth, but I also believe that they can be exactly formulated and possibly solved only on the basis of a precise conception of this notion.'[2] Tarski might not object to any of Black's arguments; it is only with Black's conclusion that he might disagree. This conclusion, I have shown, rests, for its validity, on the extension of the term 'philosophical relevance', which Black appears to have restricted unduly.

Strawson has argued that the Semantic Theory of Truth is a misconception.[3] He maintains that the word 'true' is not normally used in the way the semantic theory describes, though it may be so used for some technical purposes. He maintains that

---

1. SCT. Section 6
2. SCT. Section 18
3. P.F. Strawson, 'Truth', <u>Analysis</u>, Vol.9, No.6, 1949 pp.83-97

the semantic theory has its probable genesis in a confusion: the confusion between the use of 'true' in a phrase employed metalinguistically and the use of the word 'true' when isolated from this phrase. Strawson considers the meta-statements:

(i) "The monarch is deceased" is true if and only if the king is dead.

(ii) "The monarch is deceased" is true in English if and only if the king is dead.

In these two statements "is true if and only if"is used synonymously by Strawson for the phrase "means that". (the case of a queen being disregarded by Strawson). He states that this use of the phrase "is true if and only if" is metalinguistic. He next considers the following sentence:

(iii) "The monarch is deceased" is true in English if and only if the monarch is deceased.

Sentences like (iii) he considers as degenerate cases of metalinguistic statements of the type of (i) and (ii). He then notices the similarity between the use of the phrase "if and only if" in this type of metastatement and its use in expressions of the following type:

(iv) The monarch is deceased if and only if the king is dead.

In sentence (iv) "if and only if" occurs, but the sentence is what Strawson calls a 'necessary or defining formula', whereas in (i), (ii) and (iii) "if and only if" occurs as part of "is true if and only if" in contingent metastatements. (They are contingent because it is a contingent matter that the sentences mean what they do mean.) The similarity of the use of the phrase "if and only if" in necessary formulas to the use of it as part of the phrase "is true if and only if" in contingent metastatements, Strawson contends, may have constituted a strong temptation to regard what follows the phrase "if and only if" in the degenerate cases of metastatements as the definiens of what precedes it.

Having analysed a probable source of the misconception involved in the Semantic Theory of Truth, Strawson argues that the normal uses of the word "true" are those in which the word might be replaced by some such phrase as "I confirm it".

These criticisms are not directed against Tarski in particular, but, as Tarski and Carnap are the only two writers mentioned by Strawson in his attack on the semantic conception of truth, I shall understand that Strawson does mean them to be included among those that he criticises. I shall now offer some objections to Strawson's arguments.

The premiss of Strawson's argument is that the semantic conception of truth rests on a mistaken idea of the actual or normal use of the word "true". This implies that those who put forward the Semantic Theory of Truth have either been unaware of the uses of the word "true", other than those proposed in the theory, or in some way have confused the uses. But in fact Carnap is well aware of the different uses of the word "true". Carnap writes in his 'Introduction to Semantics:'

'It is to be noticed that the concept of truth in the sense just explained - we may call it the semantical concept of truth - is fundamentally different from concepts like "believe", "verified", "highly confirmed", etc. The latter concepts belong to pragmatics and require a reference to a person.'[1.]

Strawson's contention that "true" may be adequately replaced by some such phrase as "I confirm it", "I concede that" etc. ensures that these uses belong to what Carnap calls pragmatics, that is, they require reference to a person. Consequently, the uses of the word "true" which Strawson takes to be the normal uses of it fall outside the Semantic Theory of Truth, but at least there is no confusion involved since the uses are clearly demarcated by Carnap. Tarski also accepts that there may be other uses of the term "true" and maintains that this will make no difference to his thesis:

'A time may come when we find ourselves confronted with several incompatible, but equally clear and precise, conceptions of truth. It will then become necessary to abandon the ambiguous usage of the word "true" and to introduce several terms instead, each to denote a different notion.'[2.]

---

1. Rudolf Carnap, Introduction to Semantics 1942 p.28
2. SCT. Section 14.

'We should reconcile ourselves with the fact that we are confronted, not with one concept, but with several different concepts which are denoted by one word; we should try to make these concepts as clear as possible (by means of definition, or of an axiomatic procedure, or in some other way).'[1]

It is clear from the above quotations from both Tarski and Carnap that the semantic theory does not claim to be the only theory of truth. The concept of truth with which the semantic theory is concerned may be different from Strawson's concept of truth, but, nevertheless, Strawson does not argue that the semantic theory offers merely a <u>different</u> concept of truth. Strawson is arguing that the semantic theory of truth is based on a <u>misconception</u>.

I think that I have made it sufficiently clear that Carnap and Tarski have taken considerable care to disentangle the semantic concept of truth from other concepts of truth and to guard themselves against the accusation of misconceiving the notion of truth. The claim that Strawson makes, that the semantic theory involves a misconception of the ordinary use of the word "true" can only be substantiated if those who put forward the theory contend that their definition of "true" results in a use of that word which is coincident with its ordinary use. This, however, is not the case, for Tarski writes:

1. SCT. Section 14.

'The problem of assigning to this word ("true") a fixed and exact meaning is relatively unspecified, and every solution of this problem implies necessarily a certain deviation from the practice of everyday language.'[1.]

This is not just a reiteration of the statement that more than one concept falls under the word "true"; it asserts that the replacement of a vague concept by a precise one necessitates a deviation from the ordinary use of the word that expresses that vague concept. Tennesson has argued[2] that verbal communication is dependent upon the use of linguistic locutions that are either a) suitable for some special purpose or b) clear or c) in accordance with ordinary language. If either of the first two conditions is considered most important, then, he concludes, the locution in question will no longer be in accordance with everyday usage. (He mentions in this context Strawson's use of "presupposition".[3]) Tarski's aims of clarity and precision ensure that his definition of truth will give a use of the word "true" that is not coincident with its everyday use.

More important still, the definition that Tarski has tried to construct for everyday language would not only imply a deviation from the standard use of the word "true", it is not even intended to make precise the normal use of the word, for Tarski writes:

---

1. SCT Section 17
2. H. Tennesson, 'Permissible and Impermissible Locutions' in Studies Dedicated to Professor Carnap on his Seventieth Birthday 1962
3. P.F. Strawson, Introduction to Logical Theory, 1952, p.175ff.

'A thorough analysis of the meaning current in everyday life of the term "true" is not intended here. ... I would only mention that throughout this work I shall be concerned exclusively with grasping the intentions which are contained in the so-called classical conception of truth.'[1]

It is true that Tarski believes that the semantic conception of truth does conform to some extent with common-sense usage, but he regards this as unimportant for his thesis. In the passage in which Tarski states this belief[2], he takes care to differentiate between the aims of the semantic definition of truth and his belief about the semantic definition. It would be wrong to suppose that this belief is part of the semantic theory's claims; it is only an opinion about the semantic definition of truth, an opinion which may be mistaken as Tarski admits. If Strawson had wanted to show that the semantic theory of truth involves a misconception, then he would have had to show that this misconception was of the Aristotelian conception of truth which is the only conception of truth with which Tarski was concerned. It may be that Strawson is correct in asserting that his own use of the word "true" is more prevalent than the metalinguistic use of the word, but this is no criticism of the semantic theory for the semantic theory does not aim at offering a definition that is in accordance with everyday usage.

---

1. CTF. Introduction
2. SCT. section 17

The foregoing arguments have shown that Strawson's criticisms are misguided for they are directed against claims that the semantic theory doesnot in fact make. These are my major arguments against Strawson but there is one point of detail in his paper that I should like to consider further.

When Strawson shows what he believes to be a probable genesis of the 'misconception' he emphasises the importance that sentences like '"It is snowing" is true if and only if it is snowing' play in the semantic theory of truth. He calls these sentences degenerate metalinguistic sentences because they are degenerate forms of sentences like '"The monarch is deceased" is true if and only if the king is dead' which are metalinguistic sentences. His main argument in this section of his paper is based upon the assumption that the semantic theory of truth is concerned primarily with such 'degenerate' sentences. Strawson writes:

'To read the degenerate cases, then, as specifications, or parts, of some ideal defining formula for the phrase "is true" is to separate the phrase from the context which alone confers this meta-linguistic use upon it, and to regard the result as a model for the general use of "is true".'[1]

And again:

'... the muddle of reading a degenerate case of contingent statements metalinguistically employing the phrase <u>is true if and only</u>

---

1. Strawson, 'Truth'.

if, as a pseudo-defining formula of which the definiendum consists of a quoted sentence follows by the phrase is true ... may have contributed to the plausibility of the theory.'[1]

The 'muddle' that Strawson attributes to the adherents of the semantic theory should not be ascribed to either Tarski or Carnap since they do not consider it necessary to insist on 'degenerate' cases like '"It is snowing" is true if and only if it is snowing'. In 'The Semantic Conception of Truth' Tarski does consider such sentences throughout his paper, but it should be remembered that this paper is only expositary in character and is limited to the non-technical aspects of his earlier investigations. Yet, even, here he writes:

'(This requirement that every sentence which occurs in the object-language must also occur in the metalanguage can be somewhat modified, for it suffices to assume that the object-language can be translated into the metalanguage)'.[2]

I.e. it is not necessary that "it is snowing" occur on the right of "if and only if" in the sentence '"It is snowing" is true if and only if it is snowing"; it is only necessary that there should be some translation of "it is snowing" on the right. The problem of constructing a definition which will have as consequences sentences represented by the Schema "X is true if and only if p",

1. Strawson, 'Truth', section II
2. SCT. section 9

where "X" is the name of a sentence, is not concerned solely with those cases in which "p" is that sentence; it is equally concerned with those sentences in which "p" is a 'translation' of it. Tarski has made a simplification of his arguments contained in 'The Concept of Truth in Formalised Languages' and, in doing so, leaves aside the extra complication which would arise from the consideration of such sentences as '"The monarch is deceased" is true if and only if the king is dead'; but this simplification should not be seen as any part of the muddle to which Strawson refers. Indeed, when Tarski comes to construct a definition for an actual language in 'The Concept of Truth in Formalised Languages', he no longer considers 'degenerate' cases even as part of his criterion of adequacy for a definition of truth.[1] I shall discuss this actual definition later; for the moment I should just like to show that Tarski, in the main body of his work, dispenses with these 'degenerate' sentences. It seems unlikely, rather than plausible as Strawson maintains, that such sentences which play so little part in Tarski's investigations should have been the basis of a muddle in the semantic theory.

Carnap also pays little attention to sentences of the 'degenerate' type for he writes:

'A predicate pr_i in M is an adequate predicate for the concept of <u>truth</u> with respect to an object language S $=_{df}$ from the

1. I shall not discuss this criterion further at this point as it involves a certain knowledge of Tarski's terminology.

definition of pr_i every sentence in M follows which is constructed out of the sentential function "x is F if and only if p" by substituting pr_i for "F", a translation of any sentence $\mathfrak{S}_k$ of S into M for "p", and any name of $\mathfrak{S}_k$ for "x".[1]

It is only as examples in the expository sections of Tarski's paper that these degenerate sentences occur, and, Carnap treats them as special cases of a more general type, not as sentences to which special importance is attached. It may be the case that some confusion of the kind Strawson points out may be the source of the semantic conception of truth, but it does not appear from the investigations of Tarski and Carnap. What does appear is that Strawson has exaggerated, if not mistaken, the role played by such sentences as '"It is snowing" is true if and only if it is snowing' in the semantic theory of truth.

Strawson has chosen to ignore the domains to which the semantic theory has been applied in detail by Tarski. Strawson is concerned only with empirical statements; the truth of "sentences in a formalised language does not concern him in this particular paper. He states, however, that "true" is certainly used metalinguistically for some technical purposes and presumably he considers that the definition of truth for formalised languages as given by Tarski is constructed for such a purpose. If this is so, then a

---

1. R. Carnap, Introduction to Semantics, pp.27-28

more thorough analysis of the term "technical purposes" is needed. It appears that there are at least two ways in which "technical purposes" may be interpreted. Firstly, a term may be used in a technical field e.g. mathematics or physics. Thus the definition of "force", "work", "energy and "mass" in physics and "group", "field" and "set" in mathematics are technical definitions, the definiens of which belong to symbols and terms of a technical subject. These words in a physical or mathematical context are certainly defined for a technical purpose; their definitions have little connection with their use in everyday language. It is not in this catagory that Tarski's definition of truth falls, for it is not constructed within any technical language. Secondly, the definition may be constructed for some purpose connected with a technical field but not as part of that technical field. Such terms as "model", "hypothesis" and "explanation" in connection with physics, "proof" and "implies" in connection with mathematics and "complete", "consistent" and "independent" in connection with formalised languages are used in this technical way. Tarski's definition of "true" should come in this catagory rather than the former. But it is still not clear for what technical _purposes_ the word "true" is defined or used by him. It is his intention to make more accurate the notion of truth that is contained in the classical conception of truth for actual languages or for artificial languages. There is no change of _procedure_ when he considers formalised languages; there is a change in the results of applying this procedure to formalised languages but it is the same conception of truth in both cases with which Tarski is concerned. The use of the term "true" that his definition would imply would be the same whether for formalised or actual languages. Tarski's use of "true" differs in this respect from the use of those terms that I have listed in the second catagory, for they are terms that are used in connection with their technical fields in a way that is not intended for everyday usage. Tarski intends to use the word "true" in the same way for both the technical field of formalised languages and for the non-technical field of everyday language. That is, his criterion for the adequacy of a definition of truth remains unchanged whether he is considering formalised or informal languages. If it is the case that Strawson is willing to allow the metalinguistic use of the word "true" in connection with formalised languages, then he should allow that it is so used in connection with everyday language. I shall not consider this point further as it may be that Strawson is referring to some other use of "true" when he talks of "technical purposes".

Kneale makes one objection against Tarski's definition in his discussion of truth.[1.] He contends that truth is applicable

---

1. W. Kneale and M. Kneale, The Development of Logic 1962, ch.X section I.

primarily to propositions and that Tarski holds that it is applicable primarily to sentences. He argues further that Tarski's assumption that truth is primarily concerned with sentences leads to difficulties in the case of those sentences that contain token-reflexive words (i.e. words which locate things or events by relation to the circumstance of their own utterance). A sentence may be uttered on one occasion to express a true proposition and on another to express a false proposition. For example, the sentence "I am hungry" may be used at the same time by two different people to express two different propositions, one of which may be true, the other false.

To defend himself from these difficulties, an adherent of Tarski's conception of truth, Kneale suggests may say that truth is ascribed in some primary sense to token utterances. (A token utterance in Kneale's sense is a passing event of speech, as, for example one might say that someone stuttered in his last sentence.) If it is to token utterances that those who subscribe to Tarski's theory attribute truth, then the difficulties of a sentence changing its truth value with varying circumstances are met. But then it becomes impossible to use Tarski's device of saying "if and only if" followed by the sentence under consideration as a condition of the truth of that sentence. For it is impossible to use quotation-mark names for <u>token</u> utterances or to use the structural-descriptive names, i.e. names given to the sentences by some such device as spelling. Also, it is impossible to use the <u>same token</u> utterance following "if and only if" since a token utterance is not a form of words but a passing event.

I shall attempt at this point to answer Kneale's objection.

Firstly, Kneale appears to be wrong on a question of fact. As indicated above, he maintains that Tarski assumes that truth or the adjective "true" is applicable primarily to sentences, whereas "true" is properly applicable to propositions. He writes:

'We hold that the adjective "true" is applicable primarily to propositions, whereas he (Tarski assumes that it is applicable primarily to sentences.'[1]

In the same paragraph he makes the stronger accusation:

'... the source of the trouble seems to be Tarski's unquestioned belief that truth is primarily a property of sentences.'[2] (the underlining is mine.)

In fact, Tarski makes no claim that "true" is applicable mainly to sentences, at least, in neither of the papers referred to in my discussion nor in those referred to by Kneale in his book. It does not appear that Tarski, although nowhere saying that truth is primarily a property of sentences, might still be assuming that it is such a property. On the contrary, Tarski is quite explicit on this matter:

---

1. Kneale, The Development of Logic, p.388
2. ibid, p.389.

'The predicate "true" is sometimes used to refer to psychological phenomena such as judgements or beliefs, sometimes to certain physical objects, namely, linguistic expressions and specifically sentences, and sometimes to certain ideal entities called "propositions". By "sentence" we understand here what is usually meant in grammar by "declarative sentence"; as regards the term "proposition", its meaning is notoriously a subject of lengthy disputations by various philosophers and logicians, and it seems never to have been made quite clear and unambiguous. For several reasons its appears most convenient to apply the term "true" to sentences, and we shall follow this course.'[1]
(The first underlining is mine.)

'Of course, the fact that we are interested here primarily in the notion of truth for sentences does not exclude the possibility of a subsequent extension of this notion to other kinds of objects.'[2]
(The underlining is mine.)
From these quotations it can easily be seen that Tarski does not hold that truth is primarily a property of sentences. What he says is that he is primarily interested in truth as applied to sentences, not that it is applied primarily to sentences. For the way in which Tarski considers "true", it is most convenient to apply the term to sentences; he does not say that it is the only application of the term "true" nor the primary application of it. No doubt the difficulties

1. SCT. section 2.

of extending his treatment of the word "true" from sentences to propositions involves considerable difficulties, both philosophical and technical, but Tarski does not consider it impossible or incorrect to extend his treatment to propositions. It appears that Kneale is mistaken when he accuses Tarski of assuming that truth is primarily a property of sentences. It is true that Tarski's papers deal only with sentences, but it is equally clear from the above quotations that he does not assume what Kneale suggests.[1]

Secondly, Kneale writes of the difficulties that treating sentencds as the objects to which the attribute "true" is applied. But he does not say expliitly what these difficulties are:

'When we are concerned with mathematical formulae ... or with other phrases that resemble his example in not containing token-reflexive words ...., Tarski's assumption leads to no serious difficulties. .... But these are special cases. A sentence of the commonest kind may be uttered at different times and in different circumstances to express different propositions, some true and some false. What Jones asserts by saying "I am hungry" is not the same proposition as that Smith asserts by uttering the words at the same time, nor yet the same as that Jones asserted by uttering the words

---

1. I should qualify this remark. Although it is clear from these quotations that Tarski has not assumed that truth is primarily a property of sentences, it may not be clear from the paper CTF, both quotations being taken from SCT. Indeed, the reader may get this impression from CTF, but section 2 of SCT does appear to deny explicitly the charges of such an assumption or "unquestioned belief".

yesterday! And when we say, as we sometimes do, that a sentence was true at the time of speaking or writing, we obviously mean that it was used then to express a true proposition though it could not be put to that effect now.'[1]

The next paragraph begins: 'In order to escape from these difficulties ...'. As no difficulties have been specified, it may be assumed that Kneale is referring to some difficulty entailed for Tarski's procedure by the changing truth-value of sentendes containing token-reflexive words. But in what way does the changing truth-value of a sentence affect Tarski's method? Tarski is not trying to establish a <u>criterion</u> of truth for sentences that will automatically decide whether that sentence is true. It is not the form of words that establishes the truth or falsity of the sentence by Tarski's definition of truth[2]; the truth of a sentence like "It is snowing" is decided eventually by making an observation. Tarski does not intend that the truth of "It is snowing" to be divined by looking at the form of the sentence. The fact that "I am hungry" has a changing truth-value in no way conflicts with Tarski's definition or procedure. He does not intend to fix the truth-value of all sentences for all time. The most that Tarski's definition allows as inferences are

---

1. <u>The Development of Logic</u> p.589.

2. Although I speak of "Tarski's definition", it must be remembered that he has not given a definition of "true" for ordinary language, but only "outlined" it.

such sentences as '"I am hungry" is true if and only if I am hungry". I do not see that any extra difficulty is entailed by the fact that "I am hungry" is now true and now false, for it is still the case that '"I am hungry" is true if and only if I am hungry' is true. The same 'difficulty' applies to all sentences of the English language, since all such sentences contain verbs and all these verbs are tensed. Tensed verbs are token-reflexive words according to Kneale[1], therefore the same argument applies to "It is snowing" as to "I am hungry". If he is to be consistent then he should place "It is snowing" in the same category as "I am hungry" rather than in the category of mathematical formulae. Similarly, however, mathematical formulae may also change their truth-value, i.e. they may be 'true' in one mathematical system and 'false' in another. (This will be the case when both mathematical systems have the same rules of sentence formation but differ in the rules of transformation, e.g. by taking different axioms for the two systems.) Kneale should conclude from these considerations that all sentences whether of ordinary language or of some formalised language are susceptible to the same or related difficulties; but this is perhaps wandering from the point. What is in question at the moment is whether the changing truth-value of a sentence is of any importance to Tarski's procedure.

---

1. <u>The Development of Logic</u>, pp. 51-2

I have shown above the reasons why I do not consider that it is relevant to Tarski's method, but I shall perhaps make my point more clear by giving an example. Carnap has given a simple semantic system which contains token-reflexive words in the form of tensed verbs:[1.]

'We construct a semantic system S in the following way. S (that is to say, the object language of S) contains seven signs: three individual constants, in_1, in_2, in_3, two predicates, pr_1, and pr_2, and the two parentheses "(" and ")". ... Sentences of S are expressions of the form pr(in). The truth-conditions are given separately for each sentence by the following rules:

1. $pr_1(in_1)$ is true if and only if Chicago is large.
2. $pr_1(in_2)$ is true if and only if New York is large.
3. $pr_1(in_3)$ is true if and only if Carmel is large.
4. $pr_2(in_1)$ is true if and only if Chicago is a harbour.
5. $pr_2(in_2)$ is true if and only if New York is a harbour.
6. $pr_2(in_3)$ is true if and only if Carmel is a harbour.

This is very similar to Tarski's procedure very much reduced in application. It is to be noticed that "Chicago is large" contains a token-reflexive word, namely, "is". If pr_1 denotes the word "large" and pr_2 denotes "a harbour" (or "is large" and "is a harbour" respectively) and in a similar fasion in_1 denotes the word "Chicago" etc., then the

---

1. Carnap, Introduction to Semantics, pp.23-4

pr_i are names, in Tarski's phraseology, of the words "is large", "is a harbour", "Chicago" etc. The system S is then Tarski's procedure exactly applied to the six sentences $pr_i(in_j)$. As can be clearly seen, there is no contradiction or difficulty involves in the semantic system S by the fluctuating truth-value of "New York is large" or, in Carnap's notation, $pr_1(in_2)$. It is not part of the semantic system S to fix the truth-value of the $pr_i(in_j)$, the semantic system fixes only the truth conditions of the $pr_i(in_j)$. The same applies to Tarski's procedure; it is only the truth conditions of sentences in which he is interested, it is not his intention to give a truth-value for each sentence that will remain unchanged for all time.

Thirdly, Kneale argues that to escape from these difficulties anyone who agrees with Tarski's prodedure might take refuge in token-sentences or utterances. That is, he might say that "true" is primarily an attribute of token-utterances. Apart from the difficulties involved in the use and mention of such an utterance which Kneale has indicated, it seems an unlikely hypothesis, bearing in mind what Tarski has written:

'Statements (sentences) are always treated here as a particular kind of expression, and thus as linguistic entities. Nevertheless, when the terms "expression", "statement", are interpreted as names of concrete series of printed signs, various formulations which occur in this work do not appear to be quite correct, and give the appearance of a widespread error which consists in identifying expressions of like shape. This applies especially to the sentence '"It is snowing" is a true sentence if and only if it is snowing.', since with the above interpretation quotation-mark names must be regarded as general (and not individual) names, which denote not only the series of signs in the quotation marks but also every series of signs of like shape. In order to avoid both objections of this kind and also the introduction of superfluous complications into the discussion, ... it is convenient to stipulate that terms like "word", "expression", "sentence", do not denote concrete series of signs but whole classes of such series which are of like shape with the series given; only in this sense shall we regard quotation-mark names as individual names of expressions.'[1]

From the above it may be deduced that Tarski would not apply his procedure to token-utterances. He is well aware of the difficulties that would arise if he were to do so. But (referring to the second argument) he is fortunately not obliged to use any such subterfuge. It must first be shown that the chaning truth-values of certain sentences do lead to real difficulties for his procedure. I do not believe that Kneale has shown satisfactorily that they do.

I have now dealt in some detail with the criticisms of Black, Strawson and Kneale. There is one point that may be noticed in their objections: they all consider the semantic conception of truth

---

1. CTF, Logic Semantics and Mathematics, p.156, footnote 1

truth in connection with ordinary language, but as I have said in the first chapter of this section, if Tarski's claims are to be discussed then it will be necessary to investigate the main part of his work, which is devoted to formalised languages. I shall leave the foregoing criticisms for the moment. In the next chapter I shall discuss in more detail Tarski's aims and purposes and then reconsider these criticisms in the light of that chapter.

III

Carnap asserts that the semantic conception of truth is intended as an explication of the concept of truth as used in everyday language and in all of traditional and modern logic.[1] By "explication", Carnap understands 'the task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept'.[2] In this task of explication, the earlier concept is called the explicandum and the new, replacing concept, the explicatum. Carnap enlarges further on the notion of explication:

'Generally speaking, it is not required that an explicatum have, as nearly as possible, the same meaning as the explicandum: it should, however, correspond to the explicandum in such a way that it can be used instead of the latter.'[3]

Although Carnap states that the semantic conception of truth is intended as an explication of the everyday concept of truth, it is not certain whether Tarski intended his definition of truth as an explication or if Carnap chooses to regard it as such. Since I have said that I shall consider Tarski's definition a success or a failure if it achieves or fails to achieve its intended purposes, it

---

1. Carnap, Meaning and Necessity, 1956, p.8
2. ibid p.8-9
3. ibid p.8

will be necessary to take some care over this point.

It is not necessarily the case that Tarski did intend his concept of truth to be an explication of any earlier concept or to be an explication of the everyday concept in particular, even though Carnap may regard his own approach to the semantic concept as explicatory. For example, Strawson's analysis of the actual usages of the word "true" could be regarded as an explication of the ordinary concept of truth, but, to judge from the number of categorical statements contained in his paper[1], it is extremely doubtful if it was intended as such.

It is, however, the case that Tarski is engaged on a task of clarification, thus fulfilling Carnap's definition of explication, and, moreover, from a reading of The Semantic Conception of Truth or from the quotations on pages 175, 176 and 177 of this thesis it is clear that Tarski, in maintaining that his conception of truth is not the only one possible, intends to give what is called by Carnap an explication. For it is one of the properties of an explication that it allows other explications of the same concept. (This distinguishes the type of analysis given by Strawson from that given by Carnap; the former analysises the actual uses of the word "true", the latter replaces the actual use by another.)

The further condition that Carnap gives in order that a clarification should rank as an explication is that the explicatum

---

1. Strawson, Truth.

'should ... correspond to the explicandum in such a way that it can be used instead of the latter'. As a condition, this is still very vague. Quine enlarges on and clarifies this notion.[1] He writes of explication:

'We do not claim synonymy. We do not claim to make clear and explicit what the users of the unclear expression had unconsciously in mind all along. We do not expose hidden meanings, as the words "analysis" and "explication" would suggest; we supply lacks. We fix on the particular functions of the unclear expression that make it worth troubling about, and then devise a substitute, clear and couched in terms to our liking, that fills those functions. Beyond those conditions of partial agreement, dictated by our interests and purposes, any traits of the explicans come under the head of "don't cares".'[2.]

He continues further:

'We have, to begin with, an expression or form of expression that is somehow troublesome. It behaves partly like a term but not enough so, or it is vague in ways that bother us, or it puts kinks in a theory or encourages one or another confusion. But also it serves certain purposes that are not to be abandoned. Then we find a way of accomplishing those same purposes through other channels, using other and less troublesome forms of expression.'[3]

---

1. Quine, Word and Object, 1960, p.257 ff.
2. ibid, p.258
3. ibid, p.260

As an example of an explication Quine uses the ordered pair. I shall use this example here, as it will serve as a model of explication with which to compare Tarski's definition of truth. According to Quine, the ordered pair, when first introduced by mathematicians, was subjected to the single postulate

(i) $(x,y) = (z,w)$ implies $x = z$ and $y = w$.[1]

There are many "explications" of the ordered pair. It is only necessary that they fulfil the condition (i). Thus, (x,y) may be taken as $2^x.3^y$ or as $x+(x+y)^2$ or as $\{\{x\}, \{x,y\}\}$. They are all adequate explications of ordered pair (with the reservation that the first two are only explications of an ordered pair of numbers) because all satisfy (i). Quine contends that the utility of "orderdd pair" depends on there being denoted objects for it. Any of the above explicantia will suffice. Not only do they fulfil condition (i), but they also define ordered pairs as numbers or classes which may be admitted as objects.

To return to Tarski's definition of truth, the situation is found to be similar to the definition of the ordered pair. The problem for Tarski is to construct a definition of truth for a language L which will be formally correct, that is, a definition the definiens of which is expressed in clear and unequivical terms or terms which are reducible to such. The difference between the

---

1. Quine may be right here but modern mathematicians would use logical equivalence rather than implication in (i).

explication of the concept of ordered pair and the concept of truth lies in the conditions that they have to satisfy. For the ordered pair the condition is clear and unambiguous as stated in (i) above; for truth the condition or conditions are not obvious nor unambiguous. Before any explication can be attempted it is necessary to specify the conditions that the explicatum must fulfil.

At this point Tarski fixes his attention upon sentences as opposed to propositions and upon the correspondence theory of truth as opposed to the coherence theory of truth, the pragmatic theory of truth etc. But the formulations of the correspondence theory of truth are insufficiently precise for Tarski's purpose. He finds it necessary to formulate a clearer condition than, for example, The truth of a sentence consists in its agreement with reality or A sentence is true if it designates an existing state of affairs. Finally, he formulates the condition thus:

(T) the sentence X is true if and only if p

where "p" is to be replaced by any sentence in the language and "X" by any name of that sentence. He modifies this condition to allow a translation of the sentence named by "X" to replace "p". The definition must have as consequences of type T. (T) functions as (i) in the case of the ordered pair. The explicatum of ordered pair must have as a consequence (i); the explicatum of truth must have as consequences sentences schematised by (T). It is to be noticed that (T) does not offer a <u>definition</u> of "true" but only furnishes what Tarski calls a condition for the <u>material adequacy</u> of any definition of "true". This explains why the criticisms of Strawson, Kneale, and Black appear to be misdirected.

Leaving aside the arguments Black proposes about the 'philosophical relevance' of Tarski's definition, with which I have already dealt, I shall now consider another of his arguments. Black argues that even if a complete enumeration of words in a natural language could be achieved and a definition of truth constructed for it, the definition would still be unsatisfactory because no extension of it to other languages would be legitimate. For suppose that a definition of "true in the English language as of January 1, 1940 "could be constructed, then the difficulty remains of extending the definition to cover, for example, "true in the English language as of January 1, 1941". According to Black it would be impossible to extend the definition of truth to this second language without involving the difficulties which were noted in the discussion of quotation-mark names. Black writes:

'Anybody who is offered a definition of "true in the English language as of January 1, 1940" must, therefore, resolutely abstain from supposing that he "understands" the principle of the definition, in the sense of being able to give an explicit definition of the concepts defined. If he tries to give such a formulation, he will succeed only in talking nonsense (uttering a sentence which breaks the syntactic rules of the language to which it belongs.)[1]

This argument seems to confuse the definition with the condition, a confusion that Black has been careful to avoid elsewhere. Black states the generalisation of the sentences, '"It is snowing" is true if and only if it is snowing', '"London is a city" is true if and only if London is a city' and so on in the following form:

> '(Θ) For all x, if x is a sentence, then "x" is true if and only if x.'[2]

(Black should have written "if "x" is a sentence" instead of "if x is a sentence" in the above.) The sentence Θ is unacceptable for reasons that have already been given in connection with quotation-mark names. Black agrees with Tarski that definitions of type (Θ) would fail to fulfil the condition which states that a materially adequate definition of truth must give as consequences '"It is snowing" is true if and only if it is snowing' etc. Black continues:

> 'In default of a simple defintion expressing the intent of the condition, the best we can do is to write a schema:
>
> (S) s is true if and only if x.
>
> We may say, informally and inexactly, that an acceptable definition of "true" must be such that every sentence obtained from (S) by replacing 'x' by an object-sentence and 's' by a name of definite

---

1. Black, 'The Semantic Definition of Truth' section 7
2. ibid, section 3

description of that object-sentence shall be true. But we must remember that to talk in this way is equivalent to paraphrasing the unacceptable formula θ. At all events, (S) is not a definition of truth, but at best a criterion to guide us in the search for a definition.'[1] Black's formula (S) is the same as the condition (T) given above. If Black's assertion is correct then Tarski is unable to formulate condition (T). But (T) (and likewise (S)) do not seem to be 'paraphrases' of (θ), for (θ) uses an instance of a name-function, in fact quotation-mark-names, and is only an attempted definition of truth which is found not to satisfy condition (T). It may be the case that a name-function cannot be found that could be used in a definition of truth and would satisfy condition (T), but at least condition (T) can be stated without inconsistency. If a definition were to be constructed in some language L_i then it may be the case that the definition could be extended to another language L_j without involving the formula (θ) either explicitly or implicitly. This is the case when Tarski extends his procedure from the calculus of classes to the calculus of relations and the calculi of many-termed relations. On the other hand, it may be the case that the definition is not extendible to other languages, as would be the case of extending Tarski's definition of truth for the calculus of classes to the general theory of classes. But it cannot be extended,

---

1. ibid, section 3.

because of the peculiarities involved in the language which expresses the general theory. It is not because it involves the formula (θ). In this respect it resembles the definition of the ordered pair, for, depending on the language in which "ordered pair" is to be defined, ordered pair may be defined as $\{\{x\}\ \{xy\}\}$ or $2^x.3^y$. If it is to be defined in the calculus of classes then $\{\{x\}\ \{xy\}\}$ will serve, if in the theory of numbers, $2^x.3^y$. The only requirement is that the definition should have the consequence:

$(x,y) = (z,w)$ if and only if $x=z$ and $y=w$

Similarly, for the definition of truth the only requirement is that it should give as consequences sentences schematised by (T). As can be seen from the case of the ordered pair, it is not necessary that the definition be capable of extension to another language, but only that the condition it satisfies should be.

Strawson too has mistaken the condition for a definition. He writes:

'... the muddle of reading a degenerate case of contingent statements meta-linguistically employing the phrase *is true if and only if*, as a pseudo-defining-formula of which the definiendum consists of a quoted sentence followed by the phrase *is true* ...'[1]

The objections of Strawson against the definition of truth, should be directed against the condition for the material adequacy of

---

1. Strawson, 'Truth' section II

such a definition, for, granted the condition, any objection against a proposed definition would be that it did not satisfy this condition. What arguments may be put forward against a condition? The argument of Strawson from actual usage may show that normal uses of the word "true" do not coincide with Tarski's use of the word "true", but it is clear that Tarski deals with none of these uses. Tarski is content to find a definition that will satisfy condition (T). It may certainly be objected that the condition is of no use for ordinary language because any definition that satisfies it will be inconsistent or that no definition can be found that will fulfil it. But such objections can only be discovered after the formulation of the condition and they do not stem from such considerations as Strawson's Condition (T) does not act as the conclusion of Tarski's investigations, it acts as the starting point for all later discussion. This in turn shows that Kneale is incorrect in his assertion that:

'... he (Tarski) even goes on to argue that the possibility of constructing the paradox of the Liar within ordinary language shows that for this, as distinct from a formalised language of science, there can be no satisfactory definition of truth.'[1]

To make this into a correct assertion 'that satisfies condition (T)' should be added at the end of the quotation. Tarski makes such assertions only about those definitions that would satisfy condition (T).

---

1. Kneale, The Development of Logic, p.589.

It is not his critics alone who exaggerate Tarski's claims. Russell has written:

'Tarski ... has shown that the words "true" and "false", as applied to the sentences of a given language, always require another language, of higher order, for their adequate definition.'[1] Unless Russell means by 'adequate definition' the same as Tarski's 'materially adequate definition' in the sense that the definition satisfies condition (T), then this too may be incorrect, for Strawson's use of the word "true" does not need a hierarchy of languages or even another language. If the role that explication plays is overlooked, it is possible that statements made by Tarski about definitions of truth <u>which satisfy condition (T)</u> may be confused with statements made about "true". Tarski has guarded himself against allegations that he is making categorical statements about the use of the word "true", by accepting the existence of uses other than his own. Although he writes in one paper:

'The concept of <u>truth</u> also is to be included here, (among semantic concepts) at least in its classical interpretation'[2].

It is not such a definite assertion as it appears, for by "semantics" Tarski in this context means:

'... the totality of considerations concerning those concepts which ... express certain connections between the expressions of a language and the objects and states of affairs referred to by these

---

1. B. Russell, <u>An Enquiry into Meaning and Truth</u>, 1940 ch.4
2. A. Tarski, 'The Establishment of Scientific Semantics' included in <u>Logic, Semantics, Metamathematics</u>.

expressions.'[1]

Thus, what appears to be a dogmatic assertion about the nature of truth turns out to be tautologous.

In the remainder of this section I shall consider Tarski's definition of truth as an explication and I shall judge it accordingly. In the next two chapters I shall be concerned with the definition of truth for formalised languages in which Tarski's investigations are conducted in more detail but which the above critics have neglected.

---

1. ibid.

IV

In this chapter I shall give an exposition of Tarski's procedure for the construction of a definition of truth for a formalised language.[1] It will be necessary to give this exposition in some detail since my criticisms of Tarski's definition will require it.

Firstly, Tarski characterises formalised languages as artificially constructed languages in which the sense of every expression is unambiguously determined by its form. The essential properties possessed by all formalised languages are the following:

a. for each language a list or description is given of all the signs with which the expressions of a language are formed;

b. by purely structural properties those expressions called sentences are distinguished from all other expressions of the language;

c. a list or description is given of the sentences called axioms;

d. in special rules, called rules of inference, certain structural operations are embodied which permit the transformation of sentences into other sentences; in particular, sentences which can be obtained by the application of this operation on the axioms are called provable sentences.

---

1. Tarski's procedure may be found in CTF, section 2.

Tarski adds that he is not concerned with formal languages the expressions of which have no material sense. The problem of defining truth for such languages is not even meaningful. He writes:

'We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the languages we shall consider. The expressions which we call sentences still remain sentences after the signs which occur in them have been translated into colloquial language. The sentences which are distinguished as axioms seem to us to be materially true, and in choosing rules of inference we are always guided by the principle that when such rules are applied to true sentences the sentences obtained by their use should also be true.'[1]

Before passing to a specific language, Tarski distinguishes an object language from its metalanguage. The metalanguage is the language in which we speak about the object language. Thus, the description of expressions of the object language and the names of expressions of the object language belong to the metalanguage.

For his object language Tarski chooses the calculus of classes which, he says, can be regarded as an interpretation of the algebra of logic.

I shall briefly summarise Tarski's description of the object language (which I shall call "O") and metalanguage (which I shall call "M").

---

1. CTF, section 2

Among the signs of O are "N", "A", "Π", "I" which comprise the constants of O and the variables "$x_{,}$", "$x_{,,}$", "$x_{,,,}$" and analogous signs consisting of "x" with a number of small strokes added below, which function as the variables of O. These are the primitive signs of O, all other constants being introduced by definition in terms of these.

In the meta-language M, there are 'translations' of the expressions of O and what Tarski calls 'structural-descriptive' names of those expressions. Thus:

"N" has the translation "not" in M and the name "ng",

"A" has the translation "or" in M and the name "sm",

"Π" has the translation "for all" in M and the name "un",

"I" has the translation "is included in" in M and the name "in".

"x" followed by k small strokes has the translation "v_k" in M and is translated in M by one of the class variables of M, "a", "b" etc.,

and "st" where "s" and "t" are expressions of O has the name "s⌢t".

It can be seen that every expression of the object language has both a translation and an individual name in M. For example, "$\Pi x_{,} I x_{,} x_{,}$" has the name "$(((un^\frown v_I)^\frown in)^\frown v_I)^\frown v_I$" and the translation "for all a, a is included in a".

In addition to these names and translations of the expressions of O, the metalanguage requires expressions of a general logical character, e.g. "if and only if", and expressions from the theory of equivalent classes and the arithmetic of cardinal numbers, e.g. "infinite cardinal number".

By means of definitions the following signs of the metalanguage are introduced.

1. $x = i_{k,l}$ if and only if $x = (in^\frown v_k)^\frown v_l$
2. $x = \bar{y}$ if and only if $x = ng^\frown y$
3. $x = y+z$ if and only if $x = (sm^\frown y)^\frown z$
4. $x = \Sigma_k^n t_k$ if and only if t is a finite n-termed sequence of expressions which satisfies one of the following conditions: a. $n=I$ and $x=t_I$, b. $n>I$ and $x = \Sigma_k^{n-I} t_k + t_n$.
5. $x = y.z$ if and only if $x = \bar{y}+\bar{z}$.
6. $x = \bigcap_k y$ if and only if $x = (un^\frown v_k)^\frown y$.
7. $x = \bigcup_k y$ if and only if $x = (((ng^\frown un)^\frown ng)^\frown v_k)^\frown y$.

Next, therefollows the definitions of sentential function and of sentence:

x is a sentential function if and only if x satisfies one of the following conditions: (a) there exist natural numbers k and l such that $x = i_{k,l}$; (b) there exists a sentential function y such that $x = \bar{y}$; (c) there exist sentential functions y and z such that $x = y+z$; (d) there exists a natural number k and a sentential function y such that $x = \bigcap_k y$.

x is a sentence if and only if x is a sentential function and no variable v_k is a free variable of the function x.

The latter definition depends upon the concept of 'free variable' which is defined as follows:

v_k is a free variable of the sentential function x if and only if k is a natural number $\neq$ o, and x is a sentential function which satisfies one of the following conditions: (a) there is a natural number l such that $x = i_{k,l}$ or $x = i_{l,k}$; (b) there is a sentential function y such that v_k is a free variable of y and $x = \bar{y}$; (c) there are sentential functions y and z such that v_k is a free variable of y and x = y+z or x = z+y; (d) there is a number l distinct from k and a sentential function y such that v_k is a free variable of y and $x = \bigcap_l y$.

The following are the axioms of M:

1. ng, sm, un and in are expressions, no two of which are identical.

2. v_k is an expression if and only if k $\neq$ o; v_k is distinct from ng, sm, un, in, and from each of the expressions v_k if k $\neq$ l.

3. $x^\frown y$ is an expression if and only if x and y are expressions; $x^\frown y$ is distinct from ng, sm, un, in and from each of the expressions v_k.

4. If x, y, z, and t are expressions, then $x^\frown y = z^\frown t$ if and only if one of the following conditions are satisfied: (a) x = z and y = t; (b) there is an expression u such that $x = z^\frown u$ and $t = u^\frown y$; (c) there is an expression u such that $z = x^\frown u$ and $y = u^\frown t$.

5. Let X be a class which satisfies the following conditions:

(a) $ng \in X$, $sm \in X$, $un \in X$, $in \in X$; (b) if k is a natural number distinct from o, then $v_k \in X$; (c) if $x \in X$ and $y \in X$, then $x^\frown y \in X$. Then every expression belongs to the class X.

The Axioms of O are:

(a) $\bigcap_I i_{I,I}$

(b) $\bigcap_I \bigcap_2 \bigcap_3 (\bar{i}_{I,2} + \bar{i}_{2,3} + i_{I,3})$

(c) $\bigcap_I \bigcap_2 \bigcup_3 (i_{I,3} \cdot i_{2,3} \cdot \bigcap_4 (\bar{i}_{I,4} + \bar{i}_{2,4} + i_{3,4}))$

(d) $\bigcap_I \bigcap_2 \bigcup_3 (i_{3,I} \cdot i_{3,2} \cdot \bigcap_4 (\bar{i}_{4,I} + \bar{i}_{4,2} + i_{4,3}))$

(e) $\bigcap_I \bigcup_2 (\bigcap_3 \bigcap_4 ((\bar{i}_{3,I} + \bar{i}_{3,2} + i_{3,4}) \cdot (\bar{i}_{I,3} + \bar{i}_{2,3} + i_{4,3})) \cdot \bigcap_5 (i_{5,2} +$

$\bigcup_6 (i_{6,I} \cdot \bar{i}_{6,2} \cdot i_{6,5})))$

together with the logical axioms schematised by the following:

(a) "ANAppp",

(b) "ANpApq",

(c) "ANApqAqp",

(d) "ANANpqANArpArq".

In these schemas the sentential variables p,q,r are replaced by sentential functions, the resulting expressions, if they are not already sentences, being converted into sentences by universal quantification over the free variables contained in them.

Tarski then defines the notion of consequence and of provable sentences. I shall not give the definitions here, as I shall not need them in the later sections of this thesis.

Having constructed the definition of provable sentence, Tarski turns his attention to the definition of truth for the formalised language O, the calculus of classes. Tarski rejects the identification of "true sentence" with "provable sentence". "Provable sentence" has been defined (in M), but in such a way that there remain sentences which are not provable and the negations of which are unprovable. For example, the sentence $\cap_I \cap_2 i_{I,2}$ is not provable, nor is its negation $\overline{\cap_I \cap_2 i_{I,2}}$. Such an identification would result in the contradiction of the law of excluded middle. Tarski is impelled to construct some other definition of "true sentence" in order to avoid this contradiction of the law. Reverting to the semantic conception of truth, he formulates the condition of material adequacy in the following convention:

'<u>CONVENTION T</u>. A formally correct definition of the symbol 'Tr' (denoting the class of all true sentences), formulated in the metalanguage, will be called an adequate definition of truth if it has the following consequences:

(a) all sentences which are obtained from the expression "$x \in Tr$ if and only if p" by substituting for the symbol "x" a structural-descriptive name of any sentence of the language in question and for the symbol "p" the expression which forms the translation of this sentence into the metalanguage;

(b) the sentence "for any x, if $x \in Tr$ then $x \in S$ (where "S" denotes the class of sentences).[1.]

The definition of "true sentence" depends on the concept of satisfaction which Tarski defines thus:

'The sequence f satisfies the sentential function x if and only if f is an infinite sequence of classes and x is a sentential function and these satisfy one of the following four conditions: (a) there exist natural numbers k and l such that $x = i_{k,l}$ and $f_k \subseteq f_l$; (b) there is a sentential function y such that $x = \bar{y}$ and f does not satisfy the function y; (c) there are sentential functions y and z such that x = y+z and f either satisfies y or satisfies z; (d) there is a natural number k and a sentential function y such that $x = \bigcap_k y$ and every infinite sequence of classes which differs from f in at most the k-th place satisfies the function y.[2.]
(In the above definition "f_k" and "f_l" denote the k-th and l-th members of the sequence f.)

From this definition it follows that a sentential function with no free variables (i.e. sentences) is satisfied either by all sequences of classes or by none. The definition of "true sentence" follows:

"x is a true sentence if and only if x ∈ S and every infinite sequence of classes satisfies x'[3.]

1. CTF, section 3
2. CTF, section 3
3. ibid.

This definition is materially adequate in the sense of convention T, but that it is so may only be shown in the meta-metatheory. It implies the following consequences:

(1) for all sentences x either $x \notin Tr$ or $\bar{x} \notin Tr$;

(2) for all sentences x, either $x \in Tr$ or $\bar{x} \in Tr$.

These last sentences (1) and (2) may be proved in the metatheory and they show that the class Tr is a consistent and complete deductive system. Thus, for every sentence of the language of the calculus of classes it will be either true or false and the law of excluded middle will not be violated.

Tarski proceeds to define a related concept, the concept of <u>correct or true sentence in an individual domain</u>:

'By this is meant (quite generally and roughly speaking) every sentence which is true in the usual sense if we restrict the extension of the individuals considered to a given class a, or - somewhat more precisely - when we agree to interpret the terms "individual", "class of individuals", etc., as "element of the class a", "subclass of the class a" etc., respectively.'[1]

With this restriction on the individuals considered, it is necessary to interpret expressions of the type '$\prod$xp' as 'for every subclass x of the class a we have p', and expressions of the type 'Ixy' as 'the subclass x of the subclass a is contained in the subclass y of the class a'.

1. CTF, section 3.

There follows the prcise definitions of the concepts of <u>correct sentence in an individual domain of k elements</u> and <u>correct sentence in an individual domain a</u>. These depend upon the definition of satisfaction, defined in accordance with the limitation of the individuals to a class a. The definition of the satisfaction of the sentential function x in the individual domain a by a sequence f is the preceding definition of satisfaction with the single modification that the sequence f must be an infinite sequence of subclasses of the class a. Then follows the definition of 'correct sentence' in two forms:

DEFINITION 25[1]. x is a correct (true) sentence in the individual domain a if and only if $x \in S$ and every infinite sequence of subclasses of the class a satisfies the sentence x in the individual domain a.

DEFINITION 26. x is a correct (trie) sentence in an individual domain with k elements - in symbols $x \in Ct_k$ - if and only if there exist a class a such that k is the cardinal number of the class a and x is a correct sentence in the individual domain a.

DEFINITION 27. x is a correct (true) sentence in every individual domain - in symbols $x \in Ct$ - if and only if for every class a x is a correct sentence in the individual domain a.

---

1. The numbers of the definitions given here are those in the text of CTF.

At this point in Tarski's paper there follows a number of important definitions and theorems, a knowledge of which I shall need in the next section.

DEFINITION 28. $x = \epsilon_k$ if and only if

$$x = \overline{\bigcap_{k+I} i_{k,k+I}} \cdot \bigcap_{k+I}(\bigcap_{k+2} i_{k+I,k+2} + \overline{i_{k+I,k}} + i_{k,k+I}).$$

(This states that the class denoted by the variable v_k consists of only one element.)

DEFINITION 29. $x = \alpha$ if and only if

$$x = \bigcap_I(\bigcap_2 i_{I,2} + \bigcup_2(i_{2,I} \cdot \epsilon_2)).$$

(This states that every non-null class includes a one-element class as a part.)

DEFINITION 30. $x = \beta_n$ if and only if either

$n = 0$ and $x = \overline{\bigcap_I \epsilon_I}$ or

$n \neq 0$ and $x = \bigcap_k^{k < n+I}(\Sigma_k^{n+I} \overline{\epsilon}_k + \Sigma_1^n \Sigma_k^1 (i_{k,1+I} \cdot i_{1+I,k})).$

(β_n states that there are at most n distinct one element classes.)

DEFINITION 31. $x = \gamma_n$ if and only if either

$n = 0$ and $x = \beta_o$ or

$n \neq 0$ and $x = \overline{\beta}_{n-I} \cdot \beta_n$.

(β_n states that there are exactly n distinct one-element classes.)

DEFINITION 32. x is a quantitative sentence (or a sentence about the number of individuals) if and only if there exist a finite sequence p of natural numbers such that either $x = \Sigma_k^n \gamma_{p_k}$

or $x = \overline{\sum_k^n \gamma_{p_k}}$.

THEOREM 8. If a is a class of individuals and k the cardinal number of this class, then in order that x should be a correct sentence in the individual domain a it is necessary and sufficient that $x \in Ct_k$.

THEOREM 9. For every cardinal number k the class Ct_k is a consistent and complete deductive system.

THEOREM 10. For every cardinal number k, $Pr \subsetneq Ct_k$ but $Ct_k \not\subseteq Pr$. ("Pr" is the symbol denoting the class of provable sentences).

THEOREM 11. If k is a natural number, and X the class consisting of all the axioms together with the sentences α and γ_k, then $Ct_k = Cn(X)$.

('Cn' denotes the class of consequences of the class denoted by the symbol in the brackets.)

THEOREM 12. If k is an infinite cardinal number, and X the class consisting of all the axioms together with the sentence α and all the sentences $\bar{\gamma}_l$ (where l is any natural number), then $Ct_k = Cn(X)$.

Theorems 11 and 12 depend on these three important lemmas:

LEMMA H. For every cardinal number k $\alpha \in Ct_k$.

LEMMA I. If k is a natural number and l a cardinal number distinct from k, then $\gamma_k \in Ct_k$ and $\gamma_k \notin Ct_l$, but $\bar{\gamma}_k \notin Ct_k$ and $\bar{\gamma}_k \in Ct_l$.[1]

1. See Appendix II. of my M.A. thesis Bristol 1963

LEMMA K. If $x \in S$ and X is the class consisting of all the axioms together with the sentence α, then there is a sentence y which is equivalent to the sentence x with respect to the class X and such that either y is a quantitative sentence, or $y \in Pr$ or $\bar{y} \in Pr$.

THEOREM 13. If k is an infinite cardinal number, then there is no class X which contains only a finite number of sentences which arenot axioms and also satisfies the formula

$$Ct_k = Cn\,(X).$$

THEOREM 14. If k is a natural number and l a cardinal number distinct from k, then $Ct_k \not\subseteq Ct_l$ and $Ct_l \not\subseteq Ct_k$.

THEOREM 15. If k and l are infinite cardinal numbers, then $Ct_k = Ct_l$.

THEOREM 16. If k is an infinite cardinal number and $x \in Ct_k$, then there is a natural number l such that $x \in Ct_l$ (in other words the class Ct_k is included in the sum of all the classes Ct_l).

THEOREM 17. If X is a consistent class of sentences which contain all the axioms together with the sentence α, then there is a cardinal number k such that $X \subseteq Ct_k$; if X is a complete deductive system, then $X = Ct_k$.

THEOREM 18. In order that $x \in Ct$ it is necessary and sufficient that for every cardinal number k, $x \in Ct_k$.

THEOREM 19. In order that $x \in Ct$ it is necessary and sufficient that for every natural number k, $x \in Ct_k$.

THEOREM 20. For every cardinal number k we have $Ct \subseteq Ct_k$, but $Ct_k \nsubseteq Ct$.

THEOREM 21. The class Ct is a consistent but not a complete deductive system.

THEOREM 22. $Pr \subseteq Ct$, but $Ct \nsubseteq Pr$.

LEMMA L. $\alpha \in Ct$, but $\alpha \notin Pr$.

THEOREM 23. If x is a quantitative sentence then $x \notin Ct$.

THEOREM 24. If X is the class consisting of all the axioms together with the sentence α, then $Ct = Cn(X)$.

THEOREM 25. If $x \in S$, $x \notin Ct$ and $\bar{x} \notin Ct$, then there is a quantitative sentence y, which is equivalent to the sentence x with respect to the class Ct.

THEOREM 26. If a is the class of all individuals then $x \in Tr$ if and only if x is a correct sentence in the domain a; thus if k is the cardinal number of the class a, then $Tr = Ct_k$.

THEOREM 27. $Ct \subseteq Tr$, but $Tr \nsubseteq Ct$.

THEOREM 28. In order that $x \in Tr$, it is necessary and sufficient that x is a consequence of the class which consists of all the axioms together with the sentence α and all the sentences $\bar{\gamma}_l$, where l is any natural number.

Tarski has been able to find a structural characterisation of true sentences, but, he says, this is purely accidental.

It is owing to the specific peculiarities of the calculus of classes and such a characterisation could not be carried over to other formalised languages.

In the next chapter I shall discuss in detail the preceding theorems and Tarski's arguments.

V

In a short discussion of the Tarski definition of truth Luschei has written:

'It is not <u>incorrect</u> to stipulate, for instance, that (an expression of the form) "It is snowing here now" is a true proposition in English if and only if it is snowing here now; indeed, any definition that yielded an incompatible consequence or failed to satisfy this criterion would be wrong or inadequate; but neither is it <u>illuminating</u>.'[1]

This would be so if Tarski's definition allowed no other consequences than those illustrated by the above example. Luschei shares with Black a total disregard for the important and interesting results of Tarski's investigations. There is one such result, which I have mentioned in the discussion of Black's criticisms, namely, that for certain mathematical disciplines:

'... <u>the notion of truth never coincides with the notion of provability;</u> for all provable sentences are true, but there are true sentences which are not provable.'[2]

That there are true but unprovable sentences follows from Lemma L and Theorem 28. Lemma L states that the sentence α is not provable and from Theorem 28, together with the definition of

---

1. Luschei, <u>The Logical Systems of Lesniewski</u>, 1962, p.314
2. SCT, section 12

consequence for the language O, it follows that α is true. Similarly, the quantitative sentences $\overline{\gamma}_i$, are true but not provable. The fact that these sentences have such properties is surprising. It is surprising when the attitude of many mathematicians is considered. Einstein has written:

'A proposition is then correct ("true") when it has been derived ... from the axioms,. The question of the "truth" of the individual geometrical propositions is thus reduced to one of the "truth" of the axioms. Now it has long been known that the question is not only unanwerable by the methods of geometry, but that it is in itself entirely without meaning. We cannot ask whether it is true that only one straight line goes through two points. .... The concept "true" does not tally with the assertions of pure geometry, because by the word "true" we are eventually in the habit of designating always the correspondence with a "real" object; geometry, however is not concerned with the relation of the ideas involved in it to the objects of experience, but only with the logical connection of these ideas among themselves.[1]

This attitude is found in a modern book on logic:

'... it has become more and more widely accepted during

---

1. A. Einstein, Relativity, translated R.W. Lawson, 1920, p.2.

the past hundred years, with the result that it is now the orthodox doctrine, that to say of a mathematical proposition p that it is true is merely to say that p is true in some mathematical system S, and that in turn is merely to say that p is a theorem in S. Thus, the semantic notion of truth of mathematical propositions is replaced by a syntactical one; instead of the ordinary meaning of truth, there is offered a criterion of "truth" solely in terms of logic - formal deducibility within a given postulational system.'[1]

Tarski has stressed the importance of his result (the second quotation of this section was italicised in the original) so that, in view of the above considerations, it would be useful to investigate how the result was deduced. I have already given the theorems from which the result that α is true but unprovable follows but I shall now investigate in more detail the assumptions necessary for such a deduction.

It follows from Lemma H that $\alpha \in Ct_k$ for every cardinal number k and thus from Theorem 19 that $\alpha \in Ct$, which, in combination with Theorem 27 yields the consequence that $\alpha \in Tr$ or, in other words, α is a true sentence of the calculus of classes. The origin of the theorem that α is a true sentence may therefore be

---

1. P.H. Nidditch, Elementary Logic of Science and Mathematics, 1960 pp. 286-287.

traced back to Lemma H, which, Tarski writes, is 'almost immediately evident'. The problem now is to prove Lemma H, for although Lemma H may be almost immediately evident it requires more than this for Lemma H to be asserted as a theorem. It is easy to see that Lemma H is "self-evident" if the procedure that Tarski illustrates in the third section of "The Concept of Truth in Formalised Languages" is followed. In this procedure, the sentence under consideration i.e. $\alpha \in Ct_k$ is submitted to a succession of transformation rules, which remain implicit in the unformalised meta-theory, until the following sentence is reached:

$\alpha \in Ct_k$ if and only if for all sub-classes a of a class g with cardinal number k either (for all classes b a is included in b) or (there is some class c such that {(c is included in a) and (not for all classes d is c included in d) and (for all classes e e is included in c implies [c is included in e or for all classes f e is included in f])}).

Having obtained this translation, the second part of the equivalence being inferred from the theorems of the calculus of classes, it can be deduced that α is a correct sentence in a domain with k elements. This procedure is exactly analogous to that given below:

'$\cap_I \cup_2 i_{I,2}$ $\in Tr$ if and only if for all classes a there is a class b such that $a \subseteq b$.

From this we infer without difficulty, by using the known theorems of the calculus of classes, that $\bigcap_{I} \bigcup_{2} i_{I,2}$ is a true sentence.'[1]

At this point it can be seen why " $\alpha \in Ct_k$" is immediately evident. It is because a translation has been effected into a more familiar language. But such a procedure is dangerous as it tends to slide over the question of how the translation is established as a theorem of the calculus of classes. This is not only the case of " $\alpha \in Ct_k$" but also of any other sentence of the formalised language. Such sentences as " $\bigcap_{I} \bigcup_{2} i_{I,2} \in$ Tr" can be established by reading off the translation of whatever procedes " $\in$ Tr" and checking to see if it is in fact the case.

Finally, then, the investigation of how it is deduced that α is a true sentence leads back to an examination of the initial assumptions of the meta-theory. It is quite clear what Tarski intended; the analogy with '"Snow is white" is true if and only if snow is white' is apparent. Although there is no paraphrase in the last mentioned sentence, there is a similarity of approach - a meta-linguistic sentence is asserted by appeal to an extra-linguistic fact. In the case of '"Snow is white" is true' (a meta-linguistic statement), it is asserted or denied

1. CTF, section 3.

after an empirical enquiry. In the case of "$\cap_{I} \cup_{2} {}_{2}i_{I,2} \in Tr$", it is asserted or denied after it is known whether, for all classes a, there is a class b such that a is included in b. It is here that the analogy breaks down, for how is it to be established that this is the case? Unlike "Snow is white" there is no empirical fact to which a statement about classes can correspond. There are only three ways in which such a statement could be established:- Firstly, by an appeal to intuition; secondly, by appeal to some model of the calculus of classes; thirdly, by an investigation to discover whether it is a proven sentence in some axiomatised system incorporating the calculus of classes. It is unlikely that Tarski intends that intuition should participate in the establishment of theorems, for intuition is a notoriously bad guide for the calculus of classes. If, as in the second case, it is assumed that the translations can be checked against some model (for example, the statements could be interpreted as being "about" the regions of a square) then generality is lost. (I do not use the word "model" in its technical mathematical sense as this would imply a postulate set for the translation statements; I use the term in the sense that each statement can be read as a statement about the regions of a square). The last case presents what appears to be a reasonable alternative, but even this is not without its difficulties. If the translation statement can be asserted if and only if it is a theorem in some axiomatised system, what system is it? Because the translation is written in terms of "classes", "is included in", "for all" and "there is" it would be natural to assume that the axiom system would be the calculus of classes. The difficulty of this approach is that the axiom system would then be the calculus of classes, differing only from the formalised axioms (given already as the axioms of O) in its notation.[1]

Unfortunately, this would not allow for true but unprovable sentences such as α or $\bar{\gamma}_1$. That this is so can easily be seen from the fact that the axiom system O yields exactly the same theorems as the axiom system from which the translation statements are deduced. So the new axioms system cannot be equivalent with the axiom system O if it is to give as a consequence the (translation of) the sentence α. These axioms, whatever they may be, will form part of the axiom set for the meta-language, some of which have been given in the last section. Tarski does not indicate what axioms they are in particular; he does say that they are general logical axioms which suffice for a sufficiently comprehensive system of mathematical logic.

The conclusion of the preceding paragraph is that the meta-

---

1. Of course, the axiom sets may differ but if both are to be called <u>the</u> calculus of classes it would be necessary that they be equivalent axiom sets.

theory has as its axioms those axioms of the meta-theory listed in section IV together with axioms of a general logic which allow as theorems α and $\bar{\gamma}_i$. Amongst the axioms there must be some from which the axioms of O, when translated, can be deduced since it is one of the theorems of the meta-theory that all provable sentences (of O) are true sentences. But the result of the meta-theory that there are true but unprovable sentences of O would then say little more than that the calculus of classes (given by the axioms of O) is incomplete, i.e. there are sentences of O which are unprovable and the negations of which are also unprovable, but can be completed by the addition of sentences as axioms (in this case the sentences α and $\bar{\gamma}_i$). This is the import of Theorem 28. That it is possible to complete the axiom set by additional axioms which are again provable from a more comprehensive axiom system is a peculiarity of the calculus of classes. It is so because the calculus of classes is part of a larger general system of logic. It would not be possible in the case of a geometry of "line_s" and "points", for then the principles of logic employed would be the same in the formalised geometry as in its unformalised counterpart. (I shall show this later). It is apparent that in the case of the calculus of classes, there are two different logics employed. In the formalised language O, the only axioms specified are the axioms for the calculus of classes and certain axiom schemata which are limited to the axiom schemata of the propositional calculus.[1] In the meta-language M, not only are the axiom schemata for the propositional calculus allowed as axioms, but also the whole set of primitive propositions that are included in Principia Mathematica.[2] This accounts to some extent for the difference between the theorems of O and the theorems of M. But this raises yet another question. What makes the axioms of O form an axiom system for the calculus of classes? Why are there just those axiom schemata belonging to the propositional calculus and no other logical axioms such as those used for the metatheory M? The question is now removed to more fundamental grounds and the relationship between the formalised calculus of classes (called by Tarski "the algebra of logic") and the unformalised language in which the formalised calculus finds an interpretation must be investigated further.

Besides the proof that α is a true but unprovable sentence of the calculus of classes, there are other proofs embeddied in the meta-theory that raise different though related problems.

---

1. cf. section IV of this thesis
2. I have assumed this as Tarski directs the reader to the Principia Mathematica for the general logical axioms; he does not, however, specify any one part of that work, cf. CTF, section 3.

Theorem 28 is proved from Theorems 12 and 26, but only because:

'We can show, on the basis of the system of assumptions here adopted, that the class of all individuals is infinite.'[1]

These assumptions must be derived from the general logical axioms of the metatheory since the specific axioms of the metatheory (see section IV) do not include any axioms about 'individuals'. These auxiliary axioms do not occur as axioms of the object language O. Again the legitimacy of the approach depends on the relationship between the formalised axiom system O and the axiom system for the unformalised interpretation of O.

At this point it will be necessary to look more closely at the nature of the terms 'formalised language', 'interpretation', 'the calculus of classes', and 'the algebra of logic', all of which play a fundamental role in Tarski's investigations. Unfortunately, Tarski is not very explicit about his use of these terms. At the beginning of Section IV I have given a list of what Tarski considers to be the essential characteristics of a formalised language. The one characteristic that is

---

1. CTF, section 3.

important for the present enquiry and serves to distinguish formalised languages from formal languages is the existence of an interpretation of the symbols of the formal language.

'We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the language we shall consider. The expressions which we call sentences still remain sentences after the signs which occur in them have been translated into colloquial language. The sentences which are distinguished as axioms seem to us to be materially true, and in choosing rules of inference we are always guided by the principle that when such rules are applied to true sentences the sentences obtained by their use should also be true.'[1]

The above quotation still leaves a certain vagueness for what does 'materially true' mean? Surely it cannot mean 'intuitively true' nor 'true for some model' since these terms are also surrounded by difficulties, as I have already indicated. The 'materially true' must refer to provability in some axiom system. It is the same for the term "concrete and ... intelligible meanings' - how much more concrete and intelligible is 'for all classes x, x is included in x' than '$\Pi x_i I x_i x_i$'? It is true that 'for all classes x, x is

---

1. CTF, section 2

included in x' has the appearance of being more understandable, but for the reason that it is written in everyday English. It must not be forgotten, however, that the terms in this sentence differ in their use in a mathematical context from their use in colloquial language. They are subject to exact rules in a mathematical context; their use in colloquial language is not exactly defined.[1] It is a false impression of 'intelligibility' that is gained. There is a psychological impression of 'intelligibility' because of the paraphrase into colloquial language; it is a false impression because it ignores the essential difference between the colloquial and mathematical uses of the terms. The rules which govern 'for all classes X', 'is included in' etc. in the mathematical sense would turn out to be those rules already formalised in the axiom system from which the paraphrases were made.

It may be assumed then that 'the concrete meanings' ascribed to the signs of the formalised language are elements of another axiom system. It remains to discover between which two axiom systems the semantic rules establish this 'meaning' relationship. There is no doubt that the formalised language

---

1. Ryle, Dilemmas, 1954, chapter VIII; F. Waismann, 'Verifiability' in Essays on Logic and Language, edited by Antony Flew, 1951.

O is one of the axiom systems and, in fact, the one to which meanings are ascribed. The other is the axiom system for a general logic, for example, the set of primitive propositions taken from the Principia Mathematica, suitably translated into the language of the meta-theory. Thus, if the calculus of classes is interpreted as a part of this larger general logic, certain sentences of the calculus of classes follow from the general logical axioms alone. In this way, the sentences α and $\bar{\gamma}_i$ can be proved as theorems in the calculus of classes in the metatheory. That this is possible in the meta-theory and not in the object language is because of the restricted number of logical axioms allowed in O. It must be remembered that O contains only a limited number of logical axioms schemata, namely, an axiom schemata set sufficient for the deduction of all true sentences from the sentential calculus.

This explains the difference in extension of "true" and "provable": a different set of logical axioms is taken in the meta-theory. The additional assumptions account for the fact that α appears as a theorem in the meta-language.[1] The validity of this approach now rests on there being good reason for the restriction on the logical assumptions employed

---

1. I have written "α" and "$\bar{\gamma}_i$"; when strictly I should have written "translation of α" and "translation of $\bar{\gamma}_i$"; but which one is intended should be apparent from the context. Sometimes both are understood by "α", but here again it should be apparent from the context.

in the object language.

Ultimately, it must be decided which axioms are necessary and sufficient for the calculus of classes. This in turn demands that '<u>the</u> calculus of classes' be adequately defined. It is Tarski's responsibility to provide a precise and rigorous definition of the calculus of classes so that there may be no doubt as to which assumptions are made for it.

Unfortunately, Tarski nowhere defines what he means by the calculus of classes. The only reference that is made which relates to this point is as follows:

'The calculus of classes is a fragment of mathematical logic and can be regarded as one of the interpretations of the algebra of logic.'[1.]

By "the algebra of logic" Tarski means the following:

A class K of elements with $\cup$ and $\cap$ elements of combination subject to the set of postulates:

Ia) $a \cup b$ is an element of K when a and b are elements of K.

Ib) $a \cap b$ is an element of K when a and b are elements of K.

---

1. CTF, section 2.

IIa) There is an element $\wedge$ of K such that $a \cup \wedge = a$ for all elements a of K.

IIb) There is an element $\vee$ of K such that $a \cap \vee = a$ for all elements a of K.

IIIa) $a \cup b = b \cup a$ whenever a, b, and $a \cup b$ are elements of K.

IIIb) $a \cap b = b \cap a$ whenever a, b, and $a \cap b$ are elements of K.

IVa) $a \cup (b \cap c) = (a \cup b) \cap (a \cup c)$ whenever a, b, c, and $a \cup (b \cap c)$ are elements of K.

IVb) $a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$ whenever a, b, c, and $a \cap (b \cup c)$ are elements of K.

V) If $\wedge$ and $\vee$ exist and are unique then for all all elements a of K, there exists an element -a of K such that $a \cup -a = \vee$ and $a \cap -a = \wedge$.

VI) There are at least two elements x and y of the class K such that $x \neq y$.[1.]

At least the algebra of logic is rigorously defined. It cannot be said that the same is true of the calculus of classes. Tarski's readers know only that the calculus of classes is an

---

1. Tarski indicates this by referring to Whitehead and Russell, Principia Mathematica 2nd. edition, Vol.i, pp.205-12 from which I have taken the postulates in the above form. They are due to Huntington, 'Sets of Independent Postulates for the Algebra of Logic', Translations of the American Mathematical Society, V, (1904), pp.288-309.

interpretation of the algebra of logic defined above. It is not clear what is meant by an "interpretation". Presumably, "interpretation" is used hdre in the sense of "model", i.e. A is a model of B if and only if there is a correlation between the elements of A and the elements of B and between the operations of A and the operations of B such that the elements of A satisfy the axioms of B under these operations. But this does not necessarily imply that the algebra of logic exhausts the calculus of classes. For example, the postulates of group theory have a model in the domain of integers but not all properties of integers are provable from these postulates alone, e.g. the commutative law of addition. In the latter case the postulates of the group theory would not be called a formalisation of the theory of integers. Analogously, unless all the theorems of the calculus of classes were given by the postulate set for the algebra of logic, the algebra of logic should not be regarded as a formalised theory of the calculus of classes. For it is not with the truth of sentences belonging to the algebra of logic that Tarski is concerned but the truth of the sentences of the calculus of classes.

Russell also writes[1.] that the calculus of classes is

---

1. Whitehead and Russell Principia Mathematica, vol.i, p.205.

an interpretation of the algebra of logic and with the introduction of definition of '$\subset$', '$\cup$', '$\cap$', '$-$' proves that the calculus of classes satisfies the postulates of that algebra.[1] But this is no help in the takk of clarification. It does not say what the calculus of classes is taken to be. Does the calculus of classes include the axiom of infinity for example? (The axiom of infinity is crucial for the proof of $\bar{\gamma}_i$ in the meta-theory M).

There are two alternative explanations that now account for the difference between the extensions of "true" and "provable". Either there is some axiom set in the calculus of classes which has as consequences the sentences α and $\bar{\gamma}_i$ or there is some extraneous logical axiom set, which, together with the axioms of the calculus of classes, allow the deduction of the sentences α and $\bar{\gamma}_i$. Neither alternative is satisfactory. For, if the first alternative is the case then the axioms should appear in the formal axioms of O and if, on the other hand, the second alternative is the case then the logical axioms of O would be unduly restricted.

The resulting confusion between the axioms of O and the axioms of the meta-theory M is due to one peculiarity of the

---

1. ibid. Definitions *22.01 - *22.05

formal mathematical discipline considered by Tarski. The calculus of classes is part of a more general logic. (The other disciplines that Tarski considers in this paper possess the same characteristic. The calculus of two-term three-term, and n-term relations and the generalised theory of classes all fall within the province of mathematical logic). To place the problem in a clearer light it will be sufficient to consider some formalised language, the content of which does not form part of logic. There seems no reason why this should not be done as Tarski does not indicate that only formalised languages belonging to mathematical logic can be treated in this manner. I shall consider some axiom set from the axioms of plane projective geometry.

I wish to consider the following propositions:
a) for any two lines there is a point that lies in both, b) there is at most one point belonging to two distinct lines and c) for any two points there is a line which passes through them. If these three propositions are treated as axioms of some projective geometry, they will form a set of axioms of incidence for that geometry. In order to bring these considerations into closer analogy with the preceding, I shall call propositions a), b) and c) "axioms of incidence theory" in the same way that the axioms of O are called "axioms of the caldulus of classes".

I may now formalise these axioms thus:

<u>Axiom I</u> $(L_i)(L_j)(E\,x_k)(Ix_kL_i.Ix_kL_j)$.

<u>Axiom II</u> $(L_i)(L_j)\ (L_i \neq L_j)\ ((x_k)(x_l)((Ix_kL_i.Ix_kL_j.Ix_lL_i.Ix_lL_j)$

$(x_k=x_l)))$

<u>Axiom III</u> $(x_i)(x_j)(EL_k)(Ix_iL_k.Ix_jL_k)$.

(The difference of type indicates difference of semantic category[1]).

In a similar way, other propositions from projective geometry may be formalised. An approach analogous to that adopted by Tarski for the calculus of classes may be made to this calculus of incidence theory, treating the latter as an object language. A new metalanguage may be constructed in which the object language may be talked about. Following the procedure outlined in section IV of this paper it is possible to construct in the metalanguage the definitions of sentential function and sentence. The definition of "axiom" (in the metalanguage) will be such that Axioms I, II, and III would be axioms under the definition. Also included in the definition of "axiom" would be axioms from the sentential calculus (as in

1. For a further discussion of semantic categories, see CTF, section 4.

the previous definition for the calculus of classes) and axioms governing "=".

The axioms for the metatheory would be similar to the axioms for the metatheory M of the calculus of classes, that is, axioms relating to expressions of O and the formation of these expressions. The notion of consequence and provable sentence may then be defined in the metalanguage; the definitions would not be different, in principle, from those of the same notions for the class calculus. (The logic of quantification theory would be incorporated in the definition of consequence)

After this meta-language has been constructed it becomes possible to formulate a convention similar in outline to Convention T (Section IV of this paper). In fact, Convention T may be transferred to the new metatheory exactly as it stands[1], if it is remembered that the language referred to in the Convention is, what I have called, "incidence theory".

It still remains, however, to construct some notation in which names of the various expressions in the new object language could be formulated. This offers no new difficulties; a procedure analogous to that exemplified in 1 - 7 on page

---

1. CTF, section 3 or section IV of this thesis.

of this paper could be adopted.

The concept of satisfactioh is more complicated because of the difference of semantic category between the variables of the object language. This difficulty may be overcome by employing the method of two-rowed sequences which Tarski explains.[1.] The definition of satisfaction would then be formulated as follows:

The sequence f of points and the sequence F of lines together satisfy the sentential function x if and only if these satisfy one of the four following conditions: (a) there exist natural numbers k and l and f_k lies on F_l and $x = i_{k,l}$; (b), (c) and (d) as in the previous definition of satisfaction[2] but with relevant changes made as in (a).

(It should be noted that "$i_{k,l}$" in condition (a) is the structural-descriptive name of "Ix_kL_l", and not "Ix_kx_l" as it was in the last section of this paper. Also, condition (d) will now be divided into two divisions corresponding to the two distinct operations of universal quantification over points and universal quantification over lines. The symbolism employed in Axioms I-III may be transformed into a symbolism closer to that employed by Tarski in his paper by using the Lukasiewicz-Tarski notation rather than the Peano-Russell, reducing the sentential connectives to "A" and "N", changing

---

1. CTF, section 4
2. Section IV, p.54

the notation of the quantifier "( )" to "Π" and defining all other logical terms by means of these. Axiom I would then read:

$$\text{"}\Pi L_1 \Pi L_{11} N \Pi x_{111} NNANIx_{111} L_1 NIx_{111} L_{11}\text{"}.$$

Strictly, these changes would be necessary for the conditions (c) and (d) above to be transferred to this theory simply and directly. I shall, however, continue to use the Peano-Russell notation as I think it is easier to understand in longer sentences).

Results follow which are analogous to the consequences of the definition of satisfaction for the calculus of classes. "True sentence" may now be defined: x is a true sentence if and only if x is a sentence and every infinite sequence of points and every infinite sequence of lines together satisfy x.

It follows that the class of true sentences is consistent and complete. The class of provable sentences, on the other hand, although consistent is not complete. There remain sentences which are not provable, the negations of which are also unprovable. Such an example is provided by Desargues' Theorem or Pascal's Theorem. Desargues' Theorem may be written in the notation employed for Axioms I-III as:

$$(x_1)(x_2)(x_3)(x_4)(x_5)(x_6)(((x_1 \neq x_2 \neq x_3 \neq x_4 \neq x_5 \neq x_6) . EL_7(Ix_1L_7 . Ix_4L_7)$$

$$\cdot EL_8(Ix_2L_8 \cdot Ix_5L_8) \cdot EL_9(Ix_3L_9Ix_6L_9) \cdot EL_{10}(Ix_1L_{10} \cdot Ix_2L_{10})$$

$$\cdot EL_{11}(Ix_4L_{11} \cdot Ix_5L_{11}) \cdot EL_{12}(Ix_2L_{12} \cdot Ix_3L_{12}) \cdot EL_{13}(Ix_5L_{13} \cdot Ix_6L_{13})$$

$$\cdot EL_{14}(Ix_1L_{14} \cdot Ix_3L_{14}) \cdot EL_{15}(Ix_4L_{15} \cdot Ix_6L_{15}) \cdot Ex_{16}(Ix_{16}L_7 \cdot Ix_{16}L_8 \cdot Ix_{16}L_9)$$

$$\cdot Ex_{17}(Ix_{17}L_{14} \cdot Ix_{17}L_{15}) \cdot Ex_{18}(Ix_{18}L_{12} \cdot Ix_{18}L_{13}) \cdot Ex_{19}(Ix_{19}L_{10} \cdot Ix_{19}L_{11}))$$

$$(EL_{20}(Ix_{17}L_{20} \cdot Ix_{18}L_{20} \cdot Ix_{19}L_{20}))).$$

The translation of this sentence in the metalanguage would be: if two coplaner triangles are in perspective then the intersections of their corresponding sides are collinear. It is known that this sentence may not be deduced from Axioms I-III, nor its negation be so deduced. It is undecidable relative to these axioms. It is possible to construct theories for which Desargues' theorem does not hold. In the meta-theory, therefore, any additional geometric assumptions must be arbitrary to some extent. The adoption of any such assumption as Pascal's Theorem from which Desargues' Theorem may be deduced is decided on grounds outside the bounds of the constructed meta-theory.

For example, it may be proved that Desargues' Theorem is a consequence of the axioms of incidence for three-dimensional projective geometry. This will, of course, necessitate the introduction of more axioms dealing with the incidence of planes with planes, lines and points. If it is wished that the geometry under consideration be a special case of the geometry of three-dimensions subject to the axioms of incidence, then the inclusion of Desargues' Theorem or some logically equivalent or stronger theorem is no longer an arbitrary decision. But, considerations such as these were not taken into account in the construction of the meta-theory. The semantic definition of truth for this language does not give a truth-value to all the sentences that may be constructed in it. It does not give a truth-value for the sentence stating Desargues' Theorem, nor does it give a truth-value for the sentence "$Ex_1 Ex_2 (x_1 \neq x_2)$."

Unlike the case of the calculus of classes extra-logical axioms in the meta-theory are unable to determine the truth or falsity of the sentences mentioned in the last paragraph, for the translations of them, namely, "if two triangles are in perspective then the intersections of their corresponding sides are collinear" and "there exist two different points" contain the words "points" and "collinear" which are not part of the logical vocabulary of the meta-theory. (I do not maintain, I should like to point out, that there is any definable distinction between 'logical vocabulary' and some 'extra-logical vocabulary' of which "points" and "lines" form part, but only that those branches of mathematical logic assumed in the metatheory would be insufficient for the deduction of such sentences).[1]

From the semantic definition of truth constructed for "incidence theory" it may be deduced that '"$Ex_1Ex_2(x_1 \neq x_2)$" is true if and only if there exist two distinct points' but neither '"$Ex_1Ex_2(x_1 \neq x_2)$" is true' nor '"$Ex_1Ex_2(x_1 \neq x_2)$" is false' may be deduced. If these conclusions were all that the semantic definition of truth involved then there would be little to argue against, for there would be no sentence that could be produced which would be both true and unprovable at the same time. This was not the case with the calculus of classes; both α and the sentences $\bar{\gamma}_i$ were true but unprovable.

I have shown that in order to establish the existence of true but unprovable sentences for the calculus of classes, it was necessary to make in the mata-theory additional assumptions, that is, to have in the axiom set of the meta-theory sentences which may be translatable into the object language O but which occur there neither as axioms nor theorems. For the calculus of classes it was necessary to introduce some axiom from which the translations of the sentences $\bar{\gamma}_i$; could be deduced in the meta-theory. In the case of

1. For arguments against the theory that there is a division between logical vocabulary and scientific vocabulary, vide Quine, 'Carnap and Logical Truth', in Logic and Language, 1962, p.53 et seq.

the geometry sketched above, additional geometric assumptions may be brought into the meta-theory only with a subsequent loss of generality for the application of the geometric theory embodied in Axioms I-III. Since "$Ex_1Ex_2(x_1 \neq x_2)$" is independent of these axioms, it would be possible to have the translation of it, namely, "there exist two distinct points" as an axiom of the meta-theory. On the other hand, for exactly the same reason it is equally possible to have the negation of it as an axiom. To do either would result in "$Ex_1Ex_2(x_1 \neq x_2)$" having a truth-value. The point to notice is that the truth-values would be different depending on the axiom chosen. It would thus be possible to construct a semantic definition of truth resulting in a contradiction without altering either the definition or the semantic rules of translation from the object language into the meta-language. Of course, it is possible to say that a 'sensible' geometry demands that there exist two distinct points, but this is to go beyond the object language and the meta-theory designed for the construction of the definition of truth.

It may still be objected that this is merely playing with words for Tarski has said that the language to be investigated has a vocabulary which has 'quite concrete, and, for us, intelligible meanings'[1] ascribed to its constituents, whereas in considering

---

1. CTF, section 2

the geometrical theory I have merely offered a translation of the expressions occurring in the formalised language in terms of a vocabulary which is part of another formalised language.

In answer to this, it is possible to ask for some concrete and intelligible meanings to the terms of the object language to be produced. I have given an outline of how the semantic rules of translation give a 'meaning' to the terms of the object language, and in this case I have given points and lines as the 'meanings' of the signs occurring in it. Now I do not know what Tarski had in mind by the term 'quite concrete and intelligible' but it seems to me that expressing the meanings of the sentences of the object language in terms of points and lines is as far as one may safely go; to demand more would be dangerous. As I have said earlier in this section, the translation of the object language in terms of points and lines at once gives the impression of 'concrete and intelligible meanings' attached to the signs of that language. But the impression is a false one, for although "points" and "lines" are words in common currency and thereby gain in intelligibility if not in 'concreteness', it must not be forgotten that for the object language "points" and "lines" are used subject to exact rules; in ordinary discourse they are not. Nor should there be any confusion between the mathematical use of the word "point" (as an element of a mathematical system) and its use in a perceptual statement, since there would be a conflation of empiric and non-empiric concepts. Körner has pointed out that to do so would be to overlook the difference in kind between the two concepts; the former is an exact concept and the latter an inexact concept.[1]

The only alternative to giving 'concrete and intelligible meanings' to the signs of the object language is to give another mathematical 'meaning' to each of the signs. This entails that the semantic rules of translation from the sentences of the object language to the metalanguage will effect a translation between two axiomatic systems. There is no way of deciding whether there are two distinct points, when "point" is used as in the translations in the meta-theory of Axioms I-III. To ask 'Are there two points, distinct from one another?' is to ask what Carnap calls an internal question.[2.] By an internal question Carnap means a question which may be answered by reference to a linguistic framework, in this case the framework being an axiomatic system of geometry, since a

---

1. Körner, The Philosophy of Mathematics 1960, pp.58-62, 101-111
2. Carnap, 'Empiricism, semantics andontology' in Semantics and the Philosophy of Language, ed. Linsky.1952, p.209 et seq.

linguistic framework for "points" and "lines" considered as physical entities would not be satisfactory, as I have shown.

Other languages, and in particular, the system S which Carnap constructs[1] do not need an axiomatic system for their meta-language. Carnap's 'rules of truth' for the system S allow the truth-value of any sentence of S to be found without recourse to any axioms of the meta-theory. But this is so because the 'rules of truth' are also translation rules which translate the sentences of S into sentences about the physical world. Thus, rule 4, "$pr_2(in_1)$ is true if and only if Chicago is a harbour" allows the establishment of "$pr_2(in_1)$ is true" on the basis of observation. For formalised mathematical languages, on the other hand, no observation will provide sufficient grounds for asserting sentences of the type "$Ex_1Ex_2(x_1 \neq x_2)$".

I think I have said sufficient to show the importance of having axioms in the meta-theory when the object language is mathematical in content (in contrast to an object language of 'physical object' content as in system S above). On this point, however, I think Tarski would agree, for he writes:

---

1. For the system S, see of this thesis.

'Corresponding to the three groups of primitive expressions, the full axiom system of the meta-theory include three groups of sentences: (1) axioms of a general logical kind; (2) axioms which have the same meaning as the axioms of the science under investigation or are logically stronger than them, but which in any case suffice (on the basis of the rules of inference adopted) for the establishment of all sentences having the same meaning as the theorems of the science investigated; finally (3) axioms which determine the fundamental properties of the primitive concepts of a structural-descriptive type.'[1] (The underlining is my own)

Even though Tarski admits the need for some axioms of the second type, it is hard to understand on the basis of the previous discussion in this section how logically stronger axioms may gain admission into the meta-theory whilst not admitted into the object theory. It is at this point that I disagree with Tarski for I do not understand how the stronger logical axioms are to be justified. It is relevant to consider what Tarski says on the subject of these axioms.

---

1. CTF, section 4 pp.210-211

He writes:

' ... we are here interested exclusively in those deductive sciences which are not 'formal' in a quite special meaning of this word. I have, moreover, brought forward various conditions - of an intuitive not a formal nature - which are satisfied by the sciences here investigated: a strictly determinate and understandable meaning of the constants, the certainty of the axioms, the reliability of the rules of inference. An external characteristic of this standpoint is just the fact that, among the primitive expressions and the axioms of the meta-theory the expressions and axioms of the second group occur (of (2) above). For as soon as we regard certain expressions as intelligible, or believe in the truth of certain sentences, no obstacle exists to using them as the need arises.'[1] (The underlining is my own).

To be fair to Tarski, however, it should be pointed out that the sciences he investigates are taken from general logic and not from any geometric system. For geometry, although it is not strictly impossible to believe in the truth of certain sentences, such a belief would constitute a very weak

---

1. CTF, section 4, p.211, footnote 1.

foundation for the use of those sentences in the meta-theory. Even if the case of 'there exist at least two distinct points' seems pathological, other sentences may be brought forward which are less so. For example, if some axiomatised Euclidean geometry were formalised by means of some procedure analogous to Tarski's for the calculus of classes with the single exception of the axiom corresponding to the parallel postulate, then it would be clear that belief in the truth of the parallel postulate would be irrelevant. For, if the axioms of that system are independent, and in particular if the parallel postulate is independent of the other axioms (as in the system given by Veblen[1.]), the sentence that is the translation of the postulate in the object language is unprovable but will still be a sentence of that object language (providing that the rules of sentence formation for the language allow for its construction). In this case, as in the case of Desargues' Theorem considered previously, an appeal would have to be made to an axiom in the meta-theory that has no counterpart in the object theory. It is not possible to appeal to a <u>belief</u> in the truth or falsity of

---

1. Veblen, 'Foundations of Geometry', in <u>Monographs on Topics from Modern Mathematics</u>, 1955, pp.3-49.

this sentence to determine its truth, for the existence of alternative geometries, Lobashevsky's or Riemann's, allow a contradictory sentence to replace the parallel postulate and still retain the consistency of the axiom set. In this case no help is given by an appeal even to applicability since both Euclidean and non-Euclidean geometry may be used in physics.

I shall now give a brief summary of the preceding paragraphs on the subject of geometry. For 'true but unprovable' sentences to occur in the formalised language of a geometry, that geometry needs to be incomplete and also a logically stronger axiom set must be included in the meta-theory. The question then arises of how to justify the stronger axioms. An appeal cannot be made to an intuitive belief in the truth of these axioms; the 'intelligibility' of the translations in terms of the ordinary use of the words occurring in the meta-theory was found to be illusory; perceptual points and lines as translations of the expressions of the object theory were unsatisfactory.

Returning to the calculus of classes in which Tarski has shown the existence of 'true but unprovable' sentences, the argument loses a little of its weight, since here, axioms were introduced in to the meta-theory which were members of a set taken from a general mathematical logic. In this case, there may be some defence in appealing to 'intelligibility' or 'belief in the truth of certain sentences', since logical principles could be said to be self-evident. Such a view of logic would be a little naive in the light of the history of mathematical philosophy. Even the more sophisticated notion of the 'analyticity' of logical truths and Carnap's 'L-true' have their critics.[1.]

For the calculus of classes, extra axioms were needed in the meta-theory in order that the sentences α and γ_i could be proved to be true. The latter demanded that some axiom of infinity, or a modified version of that axiom be included in the meta-theory M so that the existence of an infinite number of individuals could be proved. But the axiom of infinity or the statement that there are an infinite number of individuals has itself been doubted by Russell:

'From the fact that the infinite is not self-contradictory, but is also not demonstrable logically, we must conclude that nothing can be known _a priori_ as to whether the number of things in the world is finite or infinite. ... The axiom of infinity will be true in some possible worlds and false in others; whether it is true or false in this world, we cannot tell.'[2]

1. vide, Quine, From a Logical Point of View, 1953 pp.45-46. also M.G. White, The Analytic and Synthetic' in John Dewey: Philosopher of Science and Freedom, 1950, pp.316-330

2. Russell, Introduction to Mathematical Philosophy, 1930 p.143

There are still other difficulties involved in the assumption that there are an infinite number of individuals. Kneale writes:

'What are the individuals of which Russell speaks, and how can we tell whether there are infinitely many of them? Russell says that he intends to refer to those things, whatever they may be, which can be named by logically proper names and cannot occur in propositions except as subjects. But he admits that it is difficult to indicate such things directly, and he even suggests that there may possibly be none because everything which appears to be an individual is in fact a class or complex of some kind.[1.]

As in the case of the geometries considered above, there is little to be hoped from an appeal to a belief in the truth of the additional assumptions made in the meta-theory; Tarski is not more explicit than Russell about the term 'individual'.

I shall now summarise the conclusions of this section. It was found by examining the proofs of some of the theorems in the last chapter that the exitence of 'true but unprovable sentences' could only be deduced in the meta-theory by the assumption of additional axioms which, though statable in the vocabulary of the object language, were not assumed there. Since

---

1. Kneale, W.M. The Development of Logic p.669.

these axioms were principles of general logic, there seemed no satisfactory reason why they should not appear in the object theory as well as the meta-theory. Also, it was found that for geometrical theories the existence of true but unprovable sentences would be extremely artificial since additional axioms in the meta-theory would be hard to justify. The calculus of classes was put in the same predicament by the assumption of the axiom of infinity which was necessary for the deduction that the sentences $\bar{\gamma}_1$ are true.

Section 4

I

In the last section I described Tarski's work on the semantic conception of truth for a certain formalised language - the calculus of classes. In the last chapter of that section I gave my criticisms of Tarski's definition based for the most part on the apparent difference between what is counted as the calculus of classes in the object language and what is counted as that calculus in the metalanguage. In particular, I disagreed with Tarski's conclusion that for this language there could be true but unprovable sentences.

The remaining sections of Tarski's paper deal with the problem of defining the semantic conception of truth for other languages. Firstly he succeeds with the language he calls the calculus of relations, defining "true" in terms of satisfaction analogous to the definition for the calculus of classes. He then deals with the logic of many-termed relations which resembles the second-order predicate calculus. Again a definition is given.

The last language with which he deals is what he calls 'the general theory of classes', a language resembling that of Russell's in the Principia Mathematica but without the axiom of reducibility and the symbols for many-termed relations.[1] The language considered

1. Instead of the axiom of reducibility, there are an infinite number of axioms which Tarski calls pseudo-definitions. Many-termed relations can be introduced by the Kuratowski-Wiener device of classes of ordered pairs, which are in turn defined as classes of classes.

is then a set-theoretic language comparable to those of Russell, Quine, Zermelo, etc. which were discussed in the first two sections of this thesis.

Translating Tarski's language into a more familiar one, the axioms he chooses are:

1. $(Ex^{n+1})(y^n)(y^n \epsilon x^{n+1} \equiv p)$ (axiom schema)

where p does not contain x^{n+1} free.

2. $(x^n)(y^{n+1})(z^n) \ (x^n \epsilon \ y^{n+1} \supset (z^n \epsilon \ y^{n+1} \vee \ x^n \neq z^n))$

3. $(Ex^2)(Ey^1)[y^1 \epsilon x^2 \ . \ (y^1)(y^1 \epsilon x^2 \supset (E z^1)(y^1 \subset z^1 \ . \ z^1 \epsilon x^2))]$

(axiom of infinity[1])

The superscripts indicate type-levels.

For this language Tarski attempts to construct a definition of truth in terms of satisfaction, but finds that such a definition is obstructed in the meta-theory by the lack of variables of higher type than any in the object language.

'In the language with which we are now dealing variables of arbitarily high (finite) order occur: consequently in applying the method of unification it would be necessary to operate with expressions of 'infinite order'. Yet neither the meta-language which

---

1. This translation has been made in terms of class membership so that it can be seen as set-theoretic. Tarski's own formulation allows two readings of X(Y): 'X has as an element the object Y, or the object Y has the property X' (CTF p.243) Quine's objections to the notation X(Y) in connection with Hilbert and Ackermann, Principles of Mathematical Logic, would apply equally to Tarski's notation and the dual reading of it which Tarski gives. vide. W. Quine, Set-Theory and its Logic. Cambridge, Mass. 1963. In connection with this axiom of infinity see the appendix of this thesis.

forms the basis of the present investigations, nor any other of the existing languages, contains such expressions. It is in fact not clear what intuitive meaning could be given to such expressions.'[1]

He then enquires whether the difficulties he encounters in trying to define the notion of truth are accidental or if they are a consequence of the language studied. In other words: is it logically impossible to construct a definition of truth for this language which satisfies convention T? He then gives a much simplified account of Gödel's theorem and concludes that no matter how a class of expressions is defined in the meta-language, this class must have members which are not in accord with condition α of convention T. I.e. Suppose a class, Tr, of expressions is defined, then there must be an expression in the object language such that:

$X \notin Tr$ if and only if p.

'X' is here the name of an expression and 'p' the translation of that expression into the meta-language. As a result no definition may be given for the semantic concept of truth which does not contravene convention T.

---

1. CTF. p.244

In the post-script to his paper, Tarski abandons the idea that to talk of expressions of infinite order has no clear meaning. Instead, by utilising the theory of transfinite ordinal numbers he then talks of allowing in the meta-language variables of order greater than any in the object language. For the object language under investigation where the variables that occur run through all finite types the meta-language needs a variable of order ω. A definition of truth which satisfies convention T can then be given successfully. As a consequence of this definition sentences of the following kind may be proved:

$X \in$ True if and only if p.

where 'X' is the name of a sentence of the object language which translates into the meta-language as 'p'. The class of provable sentences may be defined without the use of these variables, and Gödel's theorem gives for this class (Pr) the result noted on the previous page that there is an expression of the object language such that:

$X \notin$ Pr if and only if p

where 'X' is the name of that expression and 'p' its translation into the meta-language. Since it also follows that if $X \in$ Pr then $X \in$ Tr and if $(\sim X) \in$ Pr then $(\sim X) \in$ Tr, the expression designated by 'X' must belong to the class of true sentences but not to the class of provable sentences.

There are for this language, if Tarski's reasoning is correct, true but unprovable sentences. The proof of this assertion rests on Gödel's theorem applied to the general theory of classes. Gödel, independantly of Tarski, came to the same conclusion that such sentences exist, but Gödel relied on a naive notion of truth rather than any technical concept of the kind Tarski defined. From Gödel onwards, it has been taken for granted that there are arithmetical sentences which are true although they are unprovable in a formalisation of that arithmetic. Since the arithmetic notions of 0 and successor can be defined within set-theory it has been assumed that Gödel's theorem carries over to set-theory[1]: there are sentences of set-theory which are true but cannot be proved in

---

1. e.g. M. Dummet, 'In view of the fact that Gödel's theorem applies to any system which contains arithmetic, there would be an arithmetical statement expressible but not provable in this system, which we could recognise to be true.' M. Dummett, 'The Philosophical Significance of Gödel's Theorem',

'Every axiomatic theory, rich enough to contain a formalisation of arithmetic, is either inconsistent or contains a formula such that neither it nor its negation is provable within the theory and such that its truth can be demonstrated by extra-theoretic arguments.' S. Körner, 'On the relevance of Post-Gödelian Mathematics to Philosophy', Problems in the Philosophy of Mathematics, ed. I. Lakatos, Amsterdam 1967, p.124

a formalisation of set-theory.

In the last section I criticised Tarski's definition because it led to the odd conclusion that for the language considered in that section there were true but unprovable sentences. The fault in Tarski's argument, I maintained, was due to the different sets of axioms which seemed to be employed in the two languages: the meta-language and the object language.

For this new language, however, I shall use an entirely different argument. I shall argue that Tarski cannot assume that certain implications of Gödel's theorem on the incompletability of arithmetic carry over to the set-theory in which that arithmetic is expressed. I shall begin by examining Gödel's theorem for arithmetic in some detail.

Gödel[1] establishes a correspondence between the expressions of the formalised arithmetic language and sequences of such expressions with the natural numbers. He then shows that certain metamathematical relations between expressions of theobject language hold if and only if an arithmetic relation holds between the numbers corresponding to the expressions. The number corresponding to an expression of the object language is called

---

1. K. Gödel, "Uber formal unentscheidbare Sätze der Principia Mathematica und verwandte Systeme I Monatchefte für Mathematik und Phyick, vol. 38 pp.173-198; the following is a paraphrase of Gödel's argument, following Kleene, Introduction to Metamathematics, Amsterdam 1952.

the Gödel number of that expression. Consider the following metamathematical relation between an expression of the object language and a sequence of expressions of the object language.

P(A, B): A(x) is a well-formed formula with free variable x and B is a sequence of well-formed formulas such that B is a proof of A($\underline{p}$) where $\underline{p}$ is the numeral of the object language expressing the Gödel number of A(x).[1]

To this metamathematical relation there corresponds an arithmetic relation that holds between the Gödel numbers of A and B, when and only when the metamathematical relation holds. If the arithmetic relation is denoted by 'R' the above equivalence is expressed by 'P(A,B) if and only if R(g(A),g(B))' where g(A) is the Gödel number of A.

Gödel introduces the concdpt of numerical expressibility. An n-term arithmetic relation $F(x_1,\ldots x_n)$ is numerically expressed in the object language if and only if there is a well-formed formula of the object language with n free variables ($\underline{F}(x_1,\ldots,\underline{x}_n)$) such that if $F(a_1,\ldots,a_n)$ holds then $\vdash \underline{F(a_1,\ldots,a_n)}$ and if $F(a_1,\ldots,a_n)$ does not hold then $\vdash \sim\underline{F}(\underline{a}_1,\ldots\underline{a}_n)$ for each n-tuplet of numbers.[2] Since the definition of the arithmetic

1. The object language has terms which function as numerals e.g. 0, F(0), F[F(0)], F{F[F(0)]} ,........ (Gödel of cit.p.177) and these express the numbers 0,1,2,3....... Thus, the number 3 is expressed by F{F[F(0)]}
2. From here on expressions of the object language are underlined whenever confusion could arise Between expressions of the object language and expressions of the meta-language.

relation R is primitive recursive and since all primitive recursive relations can be numerically expressed in the object language, it follows that there is a well-formed formula of the object language $\underline{R}(\underline{y},\underline{z})$ such that

if, for two numbers $a_1,a_2,R(a_1,a_2)$ holds then

$\vdash \underline{R}(\underline{a}_1,\underline{a}_2)$ and

if, for two numbers $a_1,a_2,R(a_1,a_2)$ does not hold

then $\vdash \sim\underline{R}(\underline{a}_1,\underline{a}_2)$

The incompleteness theorem may now be proved.

Suppose $\vdash (\underline{y})\sim\underline{R}(\underline{p},\underline{y})$ where p is the Gödel number of $(\underline{y})\sim\underline{R}(\underline{x},\underline{y})$ then there would be a sequence of well-formed formulas $\underline{B}$ which would be a proof of the well-formed formula $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$. I.e. $P((\underline{y})\sim\underline{R}(\underline{x},\underline{y}), \underline{B})$. Therefore, $R(g[(\underline{y})\sim\underline{R}(\underline{x},\underline{y})], g(\underline{B}))$ by the equivalence of the arithmetic and metamathematical relations. Let $g(\underline{B}) = q$. Then, since R is numerically expressed by $\underline{R}, \vdash \underline{R}(\underline{p},\underline{q})$. From the predicate calculus incorporated in the object language it follows that $\vdash (E\underline{y})\ \underline{R}(\underline{p},\underline{y})$ and thence $\vdash \sim(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$. Assuming the object language to be consistent, it follows that it is not the case that $\vdash (\underline{y})\sim\underline{R}(\underline{p},\underline{y})$.

Having established that it is not the case that $\vdash (\underline{y})\sim\underline{R}(\underline{p},\underline{y})$ if the object language is consistent, it follows that no sequence of well-formed formulas is a proof of $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$. Thus, it is not the case that $R(p,1)$ nor

$R(p,2)$, nor $R(p,3)$ .... (Again from the equivalence of P and R). Since R is numerically expressed by $\underline{R}$ it follows that

$$\vdash \sim \underline{R}(\underline{p},\underline{1}), \quad \vdash \sim \underline{R}(\underline{p},\underline{2}), \quad \vdash \sim \underline{R}(\underline{p},\underline{3}), \ldots\ldots$$

If the assumption that the object language is ω-consistent is made, it is immediate that it is not the case that $\vdash \sim (\underline{y}) \sim \underline{R}(\underline{p},\underline{y})$. On the assumption that the object language is ω-consistent, it can be seen that neither $(\underline{y}) \sim \underline{R}(\underline{p},\underline{y})$ nor $(\underline{y}) \sim \underline{R}(\underline{p},\underline{y})$ can be proved. This is the incompleteness theorem in it syntactic sense.

Generally, however, more is claimed for the Gödel incompleteness theorem than this. There is also the consequence that there is a true arithmetic assertion that cannot be proved in the object language. For the truth of $(y) \sim R(p,y)$ can be seen from the truth of $\sim R(p,1), \sim R(p,2)$, etc. The well-formed formula which expresses this proposition in the object language is $(\underline{y}) \sim \underline{R}(\underline{p},\underline{y})$ which has just been shown to be unprovable. The existence of a true but unprovable sentence of the formalised language has been shown.[1] It is at this point that a leap is made in transferring this semantic implication of Gödel's theorem to a formalised set-theoretic language. It is claimed that arithmetic can be incorporated into abstract set-theory by means of a series of definitions, defining number in terms of sets. There are of course

1. This has been disputed Goddard "True and Provable" Mind, 1958 pp.13-31 Wittgenstein Remarks on the Foundations of Mathematics, Oxford 1956

many ways in which this may be done, and the method chosen will depend partially on the set-theory chosen.[1]

I shall suppose that a set-theory has been chosen and that in this set-theory a construction of arithmetic is attempted. A construction of arithmetic within set-theory is reckoned as successful if the definition of '0' and 'successor of' can be given in terms of sets only, in such a way as to preserve the laws of arithmetic.[2] Whatever definitions are given at least the five Peano axioms must be provable.[3] Three of these demand in addition to '0' and 'successor of' that the term 'number' be defined, for each of the following is a Peano axiom:

(i) 0 is a number

(ii) the successor of a number is a number

(iii) if P(0) and for each n,P(n)⊃P(successor of n) then for each n, if n is a number then P(n). (Schema)

It is necessary then to define 'number' as well as '0' and 'successor of'. (As the universe consists of sets, the extra clause in (iii) - the induction schema -' if n is a number'

1. Clearly a set-theory with type-distinctions will not allow a definition of number which has mixed types. For a comparison of set-theories see Quine "Set Theory and its Logic".

2. W. Quine, 'Set-theory and Its Logic', p.81

3. This is a minimal condition. It is also necessary to give a definition of '+' and '.', to satisfy the usual recursive definitions.

is necessary since otherwise the induction schema would imply that all sets had the property P which is not desired since in the set-theory there will be sets which are not numbers, although all numbers will be sets.)

If we suppose that the set-theory chosen is one which allows mixed types, the definitions may be given as follows:

(iv) 0 = df the nul set, i.e. $\emptyset$

(v) the successor of x, S(x), = df the set consisting solely of x, i.e. $\{x\}$[1]

(vi) the set of numbers = df the intersection of all sets containing 0 and closed with respect to the operation 'successor of'

i.e. $N = df \{x; (z)\ (0 \epsilon z.\ (y)[y \epsilon z \supset S(y) \epsilon z]) \supset x \epsilon z\}$

It is clear that the first two Peano axioms are satisfied by these definitions. The axiom schema of induction follows, for if P(0) and $(x)(P(x) \supset P(S(x))$ then P determines a set that contains 0 and is closed with respect to 'successor'. Now if $y \epsilon N$, y belongs to every such set and therefore P(y). It seems then as if the construction of arithmetic within set-theory is successful. Peano's axioms appear as provable theorems

---

1. essentially Zermelo's method 'Untersuchenden über die Grundlagen der Mengenlehre' _Mathematische Annalen_ 65

within the theory when definition (iv), (v) and (vi) are given.

Since Gödel's theorem was proved originally for a system of arithmetic which, apart from the propositional calculus and the predicate calculus of order ω, includes only the Peano axioms, it seems that the results of this theorem must carry over to the set-theory in which arithmetic has been constructed.[1]

However, Gödel's original theorem did not need a choice of terms to function as the natural numbers. The universe of discourse - the value of the variables - was limited to the numbers, the only constant terms being '0', and its successors. The Peano axioms included in the system are the three that do not mention number at all, i.e. $(x)(S(x) \neq 0)$, $(x)(y)(S(x) = S(y) \supset x=y)$ and the induction schema (iii).[2] Since there are only numbers which can be values of the variable, there is no set N consisting only of the natural numbers to define. The same may be said of the treatment Kleene gives.[3] His object language is the propositional calculus, the first-order predicate calculus, the same three Peano axioms as for Gödel's language,

---

1. K. Gödel, op.cit. pp. 177-178, 190-191
2. ibid. p.177
3. S. Kleene, op.cit. p.82.

recursive definitions of addition and multiplication and various axioms of identity. Clearly, there will be no necessity for a definition of number since here again there are only numbers admitted as possible values of the variable.

The problem now is to prove Gödel's theorem for a system in which there are other objects besides numbers. The set-theoretic construction of the numbers that has just been outlined is one such language. As before, a correspondence may be set up between the expressions of the language (and all sequences of such expressions) and the natural numbers. In this way the metamathematical predicate P and the arithmetic relation R correspond i.e. For any two expressions of the language, A and B, P(A,B) if and only if R(g(A),g(B)). The idea of numerical expressibility is slightly changed, for whereas in the formalised arithmetic case there was no choice as to which terms should be regarded as corresponding to the numbers 0,1,2, ..., a decision must be taken as to which sets in the language are to represent them. Following Zermelo I have taken $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots$ but it should be remembered that another choice was equally possible; von Neumann took $\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots$[1]. After

1. J. von. Neumann, "Zur Einführung der transfiniten Zahlen" <u>Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae</u> (sect. scient. math.) 1,1923 pp.199-208

this decision has been taken, the definition of numerical expressibility can be given. The arithmetic relation R is expressed by a well-formed formulas of the system with two free variables $\underline{R}(\underline{x},\underline{y})$ so that,

(i) if $R(a_1,a_2)$ holds for a pair of number a_1,a_2 then $\vdash \underline{R}(\underline{a}_1,\underline{a}_2)$

and (ii) if $R(a_1,a_2)$ does not hold for a pair of numbers a_1,a_2 then $\vdash \sim \underline{R}(\underline{a}_1,\underline{a}_2)$

Consider the well-formed formula $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$ where p is the Gödel number of the well-formed formula $(\underline{y})\sim\underline{R}(\underline{x},\underline{y})$ where $\underline{x}$ is a free variable. We are now in a position to prove that $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$ is unprovable. The proof is identical to the proof given previously for arithmetic. Thus, if the language is consistent, $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$ is unprovable. If now we attempt to show that $\sim(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$ is unprovable provided that the language is ω-consistent, the original proof fails. Although we can show that it is not the case that $R(p,1)$, nor $R(p,2)$, nor $R(p,3)$ etc. and thence that $\vdash\sim\underline{R}(\underline{p},\underline{1})$, $\vdash\sim\underline{R}(\underline{p},\underline{2})$, $\vdash\sim\underline{R}(\underline{p},\underline{3})$ etc. we may not say that therefore $\sim(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$ is unprovable for there are expressions in the language besides ϕ, $\{\phi\}$, $\{\{\phi\}\}$,..... (i.e. 0, 1, 2, ...). The only conclusion we may reach is that $\sim(\underline{y})[\underline{y\epsilon N} \supset \sim\underline{R}(\underline{p},\underline{y})]$ is unprovable which is not the negation of $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$. The undecidability of

$(\underline{y}) \sim \underline{R}(\underline{p},\underline{y})$ has not been shown. Choosing another formula will give us the incompleteness theorem. Suppose we choose the well-formed formula $(\underline{y}) [\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{x},\underline{y})]$ with free $\underline{x}$, then this formula will have a Gödel number, q, say, and we may show that both this formula with $\underline{q}$ substituted for the free variable $\underline{x}$ and its negation are unprovable provided that the language is ω-consistent. Suppose that $(\underline{y}) [\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y})]$ is provable, then there would be a number k such that R(q,k). Since R is numerically expressed by $\underline{R}$ it follows that $\underline{R}(\underline{q},\underline{k})$ is provable. Also, if x is a number $\vdash \underline{x} \epsilon \underline{N}$ (where '$\underline{x}$' is not a variable but the term representing the number x). Hence $\underline{k} \epsilon \underline{N}$ is provable. By the predicate calculus that is incorporated in the language $\vdash (E\underline{y}) [\underline{y} \epsilon \underline{N} . \underline{R}(\underline{q},\underline{y})]$. Therefore $\vdash \sim (\underline{y}) [\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y})]$ Assuming consistency $(\underline{y}) [\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y})]$ is unprovable. We may now prove that the negation of this formula is unprovable, provided that the set-theory is ω-consistent. As before, since $(\underline{y}) [\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y})]$ is unprovable, it is not the case that R(q,1), R(q,2) ... Hence $\vdash \sim \underline{R}(\underline{q},\underline{1})$, $\vdash \sim \underline{R}(\underline{q},\underline{2})$, ... and if the set-theory is ω-consistent $\sim (y) [\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y})]$ is unprovable.

An undecidable well-formed formula has been constructed on the assumption that the set-theory is ω-consistent. But ω-consistency for set-theory is not a predicate of set-theory of the same category as consistency. It demands for its definition that certain sets have been chosen to represent 0 and its descendents (via the successor function) as well as a set that contains just 0, 1, 2, etc. (the set denoted by 'N') Although ω-consistency is desirable, if not essential for any formal system of arithmetic, there is no such condition for set-theory. For if that set-theory with a particular choice of sets for the natural numbers turned out to be ω-inconsistent then this would suggest, not a rejection of the set-theory, but a rejection of the particular choice that we had made for the natural numbers. It may be the case that for every choice of sets for the natural numbers, the resulting theory would be ω-inconsistent, in which case there has been no undecidable formula constructed for the set-theory, although we may say that the set-theory is not suitable for the construction of arithmetic within it. I shall return to this subject later in this chapter.

Although ω-consistency is not of overriding importance for set-theory, consistency certainly is. Rosser has shown that Gödel's more stringent requirement that if a theory is ω-consistent then an undecidable formula exists, may be dropped for the weaker condition of consistency.[1] Following Kleene,[2]

---

1. J.B. Rosser, 'Extensions of some theorems of Gödel and Church', Journal of Symbolic Logic, 1936 vol.1, pp.87-91

2. M. Kleene, Introduction to Metamathematics, pp.208-209.

the metamathematical relation $\mathfrak{U}$ that the sequence of well-formed formulas B is a proof of the well-formed formula $\sim\underline{A}(\underline{p})$ where p is the Gödel number of the well-formed formula $\underline{A}(\underline{x})$ with free variable $\underline{x}$ is equivalent to an arithmetic relation T between the Gödel numbers of A and B. I.e. U(A, B) if and only if T(g(A), g(B)). For a formal system of arithmetic the relation T is numerically expressed by a well-formed formula $\underline{T}$. Consider the well-formed formula $(\underline{y})[\underline{R}(\underline{x},\underline{y}) \supset (E\underline{z})(\underline{z} < \underline{y} \,.\, \underline{T}(\underline{x},\underline{z})]$.

This is a well-formed formula with one free variable. Suppose its Gödel number is p. Then it can be shown that for the formalised arithmetic within which we are working the formula

(i) $(\underline{y})[\underline{R}(\underline{p},\underline{y}) \supset (E\underline{z})(\underline{z} < \underline{y} \,.\, \underline{T}(\underline{p},\underline{z}))]$ is undecidable if the theory is consistent. For suppose that the above well-formed formula is provable, then for some k, R(p,k) and, on the assumption of consistency $\sim(\underline{y})\left[\underline{R}(\underline{p},\underline{y}) \supset (E\underline{z})(\underline{z} < \underline{y} \,.\, \underline{T}(\underline{p},\underline{z})\right]$ is unprovable. Therefore it is not the case that T(p,0), T(p,1), T(p,2), ..., T(p,k) Hence $\vdash\sim\underline{T}(\underline{p},\underline{0})$, $\vdash\sim\underline{T}(\underline{p},\underline{1})$, $\vdash\sim\underline{T}(\underline{p},\underline{2})$, ..., $\vdash\sim\underline{T}(\underline{p},\underline{k})$. From the axioms and definitions of the arithmetic it follows that $\vdash(\underline{z})[\underline{z} < \underline{k} \supset \sim\underline{T}(\underline{p},\underline{z})]$

Also, $\vdash\underline{R}(\underline{p},\underline{k})$. From the predicate calculus incorporated in the language it follows that

$\vdash(E\underline{x})[\underline{R}(\underline{p},\underline{x}) \,.\, (\underline{z})(\underline{z} < \underline{x} \supset \underline{T}(\underline{p},\underline{z}))]$ and hence

$\vdash\sim(\underline{y})[\underline{R}(\underline{p},\underline{y}) \supset (E\underline{z})(\underline{z} < \underline{y} . \underline{T}(\underline{p},\underline{z}))]$ which is the negation of the well-formed formula that was assumed provable. Therefore if the formalised arithmetic is consistent $(\underline{y})[\underline{R}(\underline{p},\underline{y}) \supset (E\underline{z})(\underline{z} \leqslant \underline{y} \,.\, \underline{T}(\underline{p},\underline{z}))]$ is unprovable.

Assume that the negation of the last well-formed formula is provable, then for some number k, T(p,k) holds. Therefore $\vdash \underline{T}(\underline{p},\underline{k})$ and by the definitions and axioms of the formalised arithmetic we have (ii) $(\underline{y})[\underline{y} \geqslant \underline{k} \supset (E\underline{z})(\underline{z} \leqslant \underline{y} \,.\, \underline{T}(\underline{p},\underline{z}))]$. From the result at the end of the last paragraph, we know that it is not the case that R(p,0), R(p,1), R(p,3), ..., R(p,k) Hence, $\vdash \sim\underline{R}(\underline{p},\underline{0}), \vdash \sim\underline{R}(\underline{p},\underline{1}), \ldots, \vdash \sim\underline{R}(\underline{p},\underline{k})$. From the arithmetic again, (iii) $\vdash (\underline{x})\ \underline{x} \leqslant \underline{k} \supset \sim \underline{R}(\underline{p},\underline{x})$ Combining (ii) and (iii) and since $\vdash (\underline{y})[\underline{y} \leqslant \underline{k} \vee \underline{y} \geqslant \underline{k}]$ it follows that $\vdash (\underline{y})[\underline{R}(\underline{p},\underline{y}) \supset (E\underline{z})(\underline{z} \leqslant \underline{y} \,.\, \underline{T}(\underline{p},\underline{z}))]$. But we had supposed that the negation of this formula was provable. Therefore, assuming consistency, it follows that $\sim(\underline{y})\ \underline{R}(\underline{p},\underline{y}) \supset (E\underline{z})(\underline{z} \leqslant \underline{y} \,.\, \underline{T}(\underline{p},\underline{z}))$ is unprovable.

For the arithmetic languages of Gödel and Kleene, the assumption of consistency entails the existence of undecidable well-formed formulas. As with the previous Gödel result, it does not follow automatically that this undecidability carries over to a set-theory in which arithmetic can be constructed. It must be remembered that the universe now consists of sets, some of which represent numbers and others that do not. An extra difficulty arises at this point for the Rosser proof since the well-formed formula contains the symbol '$\leqslant$'. In order for the proof to be forthcoming a definition of '$\leqslant$' is required for sets. The properties this must have for the proof to go through may be found by examining the proof above. They are as follows:

(a) if $\vdash \underline{A}(\underline{0}), \vdash \underline{A}(\underline{1}), \ldots\ldots, \vdash \underline{A}(\underline{k})$ where k is a number, then $\vdash (\underline{x})(\underline{x} \leqslant \underline{k} \supset \underline{A}(\underline{x}))$

(b) whenever k is a number $\vdash (\underline{y})(\underline{y} \leqslant \underline{k} \ \underline{v}\ \underline{y} \geqslant \underline{k})$.[1.]

Quine's definition of '$\leqslant$' for sets is when inverted (Quine's definition of 'N' is the inverted version of the definition of 'N' given on page)

$\underline{x} \leqslant \underline{y} = \text{df } (\underline{z})[(\underline{x}\epsilon\underline{z}.\ ((\underline{y})\underline{y}\epsilon\underline{z} \supset \underline{S}(\underline{y})\epsilon\underline{z})) \supset \underline{y}\epsilon\underline{z}]$.[2]

This definition fails to meet condition (b) for whenever $\vdash \underline{k}\epsilon\underline{N}$ and $\vdash \underline{m}\notin\underline{N}$ it is not the case that $\vdash \underline{k} \geqslant \underline{m}$ v $\underline{k} \leqslant \underline{m}$, thus blocking the derivation from (ii) and (iii). The definition may be modified, however, in such a way as to satisfy (a) and (b). The following definition does satisfy both conditions.

$$\underline{x} \leqslant \underline{y} = \text{df } (\underline{x}\epsilon\underline{N}.\underline{y}\notin\underline{N}) \text{ v } (\underline{z})\ \ (\underline{x}\epsilon\underline{z}.[(\underline{y})\ (\underline{y}\epsilon\underline{z} \supset S(\underline{y})\epsilon\underline{z})] \supset \underline{y}\epsilon\underline{z})$$

This definition of '$\leqslant$' does not connect all sets (unless all sets belong to $\underline{N}$!) but it does connect a set which belongs to $\underline{N}$ to every set. With this definition of '$\leqslant$' the Rosser

---

1. Assuming that '$y \geqslant x$' is defined as '$x \leqslant y$'.
2. W. Quine, Set-Theory and its Logic, p.77

proof goes through as it stands and there is therefore an undecidable well-formed formula for set-theory. This time, unlike transferring the original Gödel proof to set-theory, there is no problem about the necessity of there being a class $\underline{N}$ which includes all and only the natural numbers.

If Quine's definition of '$\leqslant$' were to be used in set-theory rather than the modified version, then a proof of Rosser's theorem can be constructed by modifying the well-formed formula that is to be shown undecidable. If we replace in the following well-formed formula the free variable by the set representing the Gödel number of the well-formed formula we have a formula which can be shown undecidable. The formula is

$$'(\underline{y})[\underline{y} \epsilon \underline{N} \supset (\underline{R}(\underline{x},\underline{y}) \supset (\exists \underline{z})(\underline{z} \leqslant \underline{y} . \underline{T}(\underline{x},\underline{z})))]'.$$

Suppose that the Gödel number of the above formula is m, then consider the formula when '$\underline{x}$' is replaced by '$\underline{m}$'. Suppose the resulting formula is provable. Then R(m,k) for some number k. Thence $\vdash \underline{k} \epsilon \underline{N} . \underline{R}(\underline{m},\underline{k})$. Also, if the language is consistent it is not the case that the following hold:- $\nvdash T(m,0), \nvdash T(m,1), \ldots\ldots, T(m,k))$. From the definition of '$\leqslant$' it follows that $\vdash (\underline{z})(\underline{z} \leqslant \underline{k} \supset \sim \underline{T}(\underline{m},\underline{k}))$. Combining by the predicate calculus, we may deduce that the negation of the formula we supposed was provable is also provable. Assuming consistency the well-formed formula $(\underline{y})[\underline{y} \epsilon \underline{N} \supset (\underline{R}(\underline{m},\underline{y}) \supset (\exists \underline{z})(\underline{z} \leqslant \underline{y} . \underline{T}(\underline{m},\underline{z})))]$

is unprovable.

Now we may show that the negation of the last well-formed formula is unprovable. Suppose that it is provable, then $T(m,k)$ for some number k. From numerical expressibility and the predicate calculus it follows that

(iv) $(y)[y \epsilon N \supset (y \geqslant k \supset (Ez)(z \leqslant y . T(m,z)))]$. Also, if the language is consistent it follows that none of $R(m,0), R(m,1), \ldots\ldots, R(m,k)$ hold. Therefore,

(v) $\vdash (\underline{y})(\underline{y}\epsilon\underline{N} \supset (\underline{y} \leqslant \underline{k} \supset \sim \underline{R}(\underline{m},\underline{y})))$. Since $\vdash (\underline{k}\epsilon\underline{N} . \underline{y}\epsilon\underline{N}) \supset (\underline{y} \leqslant \underline{k} \ \underline{v} \ y > \underline{k})$ and $\vdash \underline{k} \epsilon \underline{N}$ we may combine (iv) and (v) to give $\vdash (\underline{y})[\underline{y}\epsilon\underline{N} \supset (\underline{R}(\underline{m},\underline{y}) \supset (E\underline{z})(\underline{z} \leqslant \underline{y} . \underline{T}(\underline{m},\underline{z})))]$

Thus, if the language is consistent then

$\sim(\underline{y})[\underline{y}\epsilon\underline{N} \supset (\underline{R}(\underline{m},\underline{y}) \supset (E\underline{z})(\underline{z} \leqslant \underline{y} . \underline{T}(\underline{m},\underline{z})))]$ is unprovable.

Again, in this proof there was no necessity that $\underline{N}$ should be a set that consisted of those and only those sets which represented the natural numbers.

Analogously, we may show that the following two well-formed formulas are undecidable:

$(\underline{y})(\underline{T}(\underline{r},\underline{y}) \supset (E\underline{z})(\underline{z} \leqslant \underline{y} . \underline{R}(\underline{r},\underline{z})))$ where r is the Gödel number of the same well-formed formula with the free variable '$\underline{x}$' replacing '$\underline{r}$' and the definition of '$\leqslant$' is the modified version above.

$(\underline{y})(\underline{y}\epsilon\underline{N} \supset (\underline{T}(\underline{n},\underline{y}) \supset (E\underline{z})(\underline{z} \leqslant \underline{y} . \underline{R}(\underline{n},\underline{z}))))$ where n is the Gödel number of the same well-formed formula with the free variable '$\underline{x}$' replacing '$\underline{n}$', and '$\leqslant$' is defined as the inverse of Quine's definition.

There is for set-theory no lack of well-formed formulas that can be shown to be undecidable. (Clearly, there are an infinite number, since, to the above undecidable formulas we may prefix any theorem of the language as an antecedent of a conditional the consequence of which is one of the undecidable formulas.) We cannot doubt that Gödel and Rosser's undecidability results do carry over to set-theory. But this has only syntactic import. Tarski has alleged that the incompleteness results also have semantic implications for set-theory[1]. It is just this that I wish to deny.

There are convincing reasons for saying that any intended formalisation of arithmetic must contain formulas that are unprovable even though they express true propositions when interpreted as arithmetic propositions, i.e. when the formal variables are taken as ranging over the natural numbers and the formal operations interpreted as arithmetic operations. The first Gödel formula that we have shown to be unprovable is $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$. When this is interpreted as expressing an arithmetical proposition we can see that the interpretation is true. For, as we have shown $\sim R(p,n)$ holds for each natural

1. See beginning of this section,

number n. Therefore $(y)\sim R(p,y)$ is true and this is expressed in the formal system by $(\underline{y})\sim\underline{R}(\underline{p},\underline{y})$. To show this more rigourously a definition of a formula being true is needed, but this can be given easily in terms of the satisfaction of formulas in the domain of natural numbers in the same way as for any set of formulas of the first-order predicate calculus.[1]

The Rosser formula fares similarly, for we have shown that $(\underline{y})(\underline{R}(\underline{p},\underline{y}) \supset (E\underline{z})(\underline{z} \leq \underline{y}.\ \underline{T}(\underline{p},\underline{z})))$ is unprovable (on the assumption of consistency). That is, there is no sequence of formulas such that this sequence is a proof of the well-formed formula $(\underline{y})(\underline{R}(\underline{x},\underline{y}) \supset (E\underline{z})(\underline{z} \leq \underline{y}.\ \underline{T}(\underline{x},\underline{z})))$ when the free variable is replaced by the Gödel number of the formula, namely p. From the equivalence of the metamathematical predicate P and the arithmetic predicate R we see that for each natural number n it is not the case that $R(p,n)$ and hence that it is the case that for each natural number n if $R(p,n)$ then there exists a number z such that $z \leq y$ and $T(p,z)$. The formula that expresses this in the formalised arithmetic is the unprovable formula at the beginning of this paragraph. Here again we have an unprovable but true formula of formalised arithmetic or an unprovable formula that expresses in the formalised arithmetic a true arithmetic proposition.

---

1. see A. Church, *Introduction to Mathematical Logic*, pp.174-175 and 227-228.

The problem now is to see if there are any corresponding results for set theory. We know that the set theory incorporates arithmetic in as much as Peano's postulates are deducible in it for a certain sub-system of sets. We also know that any formalised language in which Peano's postulates can be derived contains a formula that is unprovable but which expresses a true arithmetic proposition. But it does not follow that there is in formalised set theory a formula that is unprovable but which expresses a true set-theoretic proposition.

The position is made clearer if we consider those examples of undecidable formulas which we discussed earlier. The set theoretic version of Gödel's formula is $(\underline{y})(\underline{y} \in \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y}))$ and the two set theoretic versions of Rosser's formula are

$(\underline{y})[\underline{y} \in \underline{N} \supset (\underline{R}(\underline{m},\underline{y}) \supset (\exists \underline{z})(\underline{z} \leqslant \underline{y} . \underline{T}(\underline{m},\underline{z})))]$ and

$(\underline{y})[\underline{R}(\underline{p},\underline{y}) \supset (\exists \underline{z})(\underline{z} \leqslant \underline{y} . \underline{T}(\underline{p},\underline{z}))]$.

When discussing the Gödel formula, we found that $\underline{R}(\underline{q},\underline{0})$, $\underline{R}(\underline{q},\underline{1})$,.... all numerically expressed true propositions of arithmetic. Can we say that they also express true set-theoretic propositions? Trivially that is so, for they are also all provable. Hence they must be true in any model of the axioms, i.e. whenever an interpretation of the axioms is given for the axioms in which they are all true, then with this interpretation all theorems are true. Clearly the Gödel formula is not true in this sense, since its undecidability ensures that there are models for the axioms in which it is true and others in which it is false.[1] What must be presupposed in saying that the Gödel formula expresses a true set-theoretic proposition is that there is some one interpretation of the axioms in which the formula is true, that there is some intended model of the formalised set theory just as in the case of formalised arithmetic there is an intended model.

Let us suppose that there is such an intended model. Then in this model $\sim\underline{R}(\underline{q},\underline{0})$, $\sim\underline{R}(\underline{q},\underline{1})$, ... will express true set-theoretic propositions, for the aforesaid reason. If it is asserted that in this model $(\underline{y})(\underline{y}\in\underline{N} \supset \sim\underline{R}(\underline{q},\underline{y}))$ also expresses a true set-theoretic proposition then this could only be the case if the set in the model corresponding to $\underline{N}$ consisted of those and only those sets which correspond in the model to the sets that represent the natural numbers in the formalised set theory. That is, the set in the interpretation that corresponds to $\{\underline{x};\ (\underline{z})\ [\underline{0}\in\underline{z}.(\underline{w})(\underline{w}\in\underline{z} \supset \underline{S}(\underline{w})\in\underline{z})] \supset \underline{x}\in\underline{z}\}$ consists only of the sets corresponding to $\underline{\emptyset}$, $\{\underline{\emptyset}\}$, $\{\{\underline{\emptyset}\}\}$, $\{\{\{\underline{\emptyset}\}\}\}$, ... For if in the model there was a set a such that $a\in N$ and $a\neq\emptyset$, $a\neq\{\emptyset\}$, $a\neq\{\{\emptyset\}\}$,[2] etc. then we should be unjustified in saying

---

1. Derived straightforwardly from Gödel's completeness theorem of the predicate calculus.
2. The set of the model corresponding to a particular set of the formal theory is here written identically except for the absence of underlining.

that since $\sim\underline{R}(\underline{q},\underline{0})$, $\sim\underline{R}(\underline{q},\underline{1})$, ... each express true propositions of the model so must $(\underline{y})(\underline{y}\in\underline{N} \supset \sim\underline{R}(\underline{q},\underline{y}))$. In other words, to say that $(\underline{y})(\underline{y}\in\underline{N} \supset \sim\underline{R}(\underline{q},\underline{y}))$ expresses a true proposition is to presuppose a model in which N consists of those and only those sets ϕ, $\{\phi\}$, $\{\{\phi\}\}$, etc. and that this model is, in some sense, the intended or proper model for the formalised set theory.

Exactly the same remarks can be made about the Rosser formula $(\underline{y})[\underline{y}\in\underline{N} \supset (\underline{R}(\underline{m},\underline{y}) \supset (\exists\underline{z})(\underline{z}\leq\underline{y}.\underline{T}(\underline{m},\underline{z})))]$ because the antecedent of the hypothetical is the formula $\underline{y}\in\underline{N}$.

The third formula is more difficult to deal with since it does not contain the antecedent clause $\underline{y}\in\underline{N}$ explicitly. (It may, of course, occur implicitly in that if the function $\underline{R}(\underline{x},\underline{y})$ were spelled out in full $\underline{y}\in\underline{N}$ might occur as an antecedent clause in this expansion. The above remarks would then apply directly to this formula.) To say that the formula expresses a true proposition would then be to say that in the intended model whenever $R(p,x)$ is true for the set x then $(\exists z)(z\leq x.T(p,z))$ is true. Whether or not this is so can only be decided when the intended model is known.

To talk of the truth of an undecidable formula of a formalised set theory is to presuppose a model which we think of as the intended model of the theory. What one may question is whether one can talk of proper models of set theory in any clear manner. If one could talk of a unique intended model then it would be proper to talk of the truth (or falsity) of the above formulas.

Recently there has come into vogue the term 'standard model' to describe such a model. Myhill, for example, claims that there is only one standard model of set-theory.[1] Further, he claims that in the denumerable models of set-theory, the existence of which is assured by the Löwenhdim-Skolem theorem, when the predicate letter 'ϵ' is assigned it no longer represents the relation 'is a member of' since the standard model contains an indenumerable field of sets. 'Class-membership certainly has a vast non-denumerable field'.[2] As a corrective to this one should bear in mind that there are set-theories in which the membership relation has a denumerable field. Would it then be proper for Myhill to say that such theories do not contain the relation of class membership?

The Löwenheim-Skolem theorem assures us that there are models of a formalised set theory in the domain of natural numbers. I think a case could be made out for saying that such a model is 'non-standard' in the sense that an arithmetic predicate would then be the interpretation of 'ϵ'. e.g. If the axiom of infinity is ommitted from the Zermelo-Fraenkel axioms the

1. J. Myhill 'The Ontological Significance of the Löwenheim-Skolem Theorem' Academic Freedom, Logic and Religion, Philadelphia 1953, p.68.
2. ibid.

membership relation $x \epsilon y$ can be interpreted as the relation that the quotient of dividing y by the xth power of 2 is an odd number.[1] There is a sense here in which the arithmetic relation cannot be said to be the relation of class membership. But this sense is just that the relation holds between arithmetic entities, natural numbers, rather than, say, between numbers and sets of them. Certain formulations of the propositional calculus can in a like manner be construed as formulations of a partial arithmetic. Myhill's worry that set theory will turn out to be some complicated arithmetic relation[2] if the standard model is forgotten is misplaced. It would be as sensible to worry about the propositional calculus turning out to be a simple arithmetic theory.

But Myhill goes too far in saying that any denumerable model of a formalised set theory does not contain the relation of class membership. Various 'inner' models of certain set-theories are known.[3] One wonders if Myhill would say that these too do not contain the notion of class-membership, since the model contains only <u>some</u> of the sets of the whole theory.

There is another difficulty which Myhill overlooks and this is the difficulty of a preassigned interpretation of the predicate

1. due to Ackermann, mentioned in H. Wang, <u>A Survey of Mathematical Logic</u> p.392.
2. J. Myhill, op.cit. p.69
3. Gödel's proof of the consistency of the axiom of choice relies on such a model.

letters occurring in the formulas of the formal system. Myhill's definition of a standard model of a system is relative to interpretations being given to some of predicate letters of the system. Certain of the predicate letters of the formal system are given an interpretation and then a standard model is defined as a model in which those predicate letters receive the preassigned interpretation. Now such a definition is legitimate only if we can specify the preassigned interpretation. In order to talk of the standard model of set-theory Myhill says that we must first assign to the predicate letter 'ϵ' of the formalised set theory the relation of class membership. But in order to specify this relation, we succeed only in giving a set of conditions which the relation fulfills. This set of conditions in turn can receive odd interpretations; it is no more safe from the consequences of Gödel's completeness theorem than the formal system itself. On the subject of such preassigned interpretations Wang has likewise said that the explicit specification of the preassignment meets insuperable difficulties.[1]

If there is no model which could be regarded as the standard model of a formalised set-theory or, in the phraseology of three

1. H. Wang, 'On Denumerable Bases of Formal Systems', in Mathematical Interpretations of Formal Systems, ed. A. Heyting, Amsterdam, 1955 p.72

pages back, <u>the</u> intended model, then it is difficult to see how we could talk sensibly of the <u>truth</u> of some formula of the formal system which is unprovable.

In spite of these difficulties, it might be thought that even if we could not specify completely the intended or standard model we might be able to specify <u>some</u> of the conditions which a model must fulfill in order to be called standard. Thus, as we noted above, the standard model must include $(y)(y \epsilon N \supset \sim R(q,y))$ as one of its true propositions if we are to say that $(\underline{y})(\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y}))$ expresses a true proposition. It follows that N must contain only ϕ, $\{\phi\}$, $\{\{\phi\}\}$, ... This means that <u>N</u> receives as a translation in the model 'the set consisting solely of ϕ, $\{\phi\}$, ...' Now, suppose that we use Tarski's method for translating the formal sentences of the language into an informal language (belonging to the metalanguage). If we are considering Zermelo's set theory, for example, we may do so since the variables belong to one type. There are no difficulties for such a theory caused by the variables of the formal system belonging to infinitely many types. The concept of satisfaction can be successfully defined and closed formulas of the system will receive translations in terms of 'for all sets' and 'there is a set such that'. It is clearly that the import of the translation of <u>N</u> will be 'the intersection of all sets z that contain the nul set are such that if they contain a set y, they also contain the unit set whose only number is y. The translation will not be 'the set consisting solely of the nul set, the unit set of the nul set, ...'. Furthermore, the translation could not be 'the set consisting solely of $\emptyset$, $\{\emptyset\}$, ...', since no formula of a formalised set theory contains the means for expressing dots, or the phrase 'and so on' or 'etc'. The description is not expressible in the formal system and so any proposition which contains this description will not be directly expressible in the formal system. The most that we can hope for is that we can express it indirectly by some other description. This is what we attempt to do when we define N by means of an intersection.

Nothing, however, has been achieved by this translation. If the metalanguage is formalised then the translation of the formal system is once more Zermelo's set theory. The problem is the same as that which occurred in section 3. We have effected a translation of a formal language into what might be described as 'realistic' terms. The existential operator is not, in the translation, just a mere symbol but is used to assert the existence of some particular set. In order to segregate the set theoretic propositions in the translation which are true from those which are false we should have to have some standard set theory which we could use as a guide for deciding which of the propositions of the translation held and which did not. Until this standard set theory is specified, at least in part, there is no way of so classifying the propositions of the translation.

The propositions that are of most interest in this context are the undecidable sentences of the formal system. The translation of $(\underline{y})(\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y}))$ will be 'for all sets y, if y belongs to the intersection of all sets that contain y and are such that if they contain a set w, they also contain its unit set, then $\sim R(q,y)$' (where $\sim R(q,y)$ is the translation of $\sim \underline{R}(\underline{q},\underline{y})$). Whether this translation is a true proposition depends on whether the intersection does contain just ϕ, $\{\phi\}$, etc. A proof that this set does contain just these sets presupposes the existence of a set which contains just these. For suppose that every set which contained ϕ, $\{\phi\}$, ... also contained some additional member. Then there is no assurance that ϕ, $\{\phi\}$, etc. are the only sets common to all sets that contain them.[1] That there is a set consisting only of ϕ, $\{\phi\}$, etc. we could take as a true proposition of the standard model. Then it would follow that the intersection set contained only ϕ, $\{\phi\}$, etc. and, hence, that, for all sets y, if y belongs to the intersection set then $\sim R(q,y)$. Therefore, $(\underline{y})(\underline{y} \epsilon \underline{N} \supset \sim \underline{R}(\underline{q},\underline{y})$ expresses a true proposition of the model.

1. L. Henkin mentions this argument in hid discussion of the non-isomorphic character of the Peano axioms, 'Completeness in the Theory of Types', Journal of Symbolic Logic, 1950, p.89.

We shall find that this has unwelcome consequences. Quine[1], discussing ω-inconsistency and its implications for the system NF, says that if his system turned out to be ω-inconsistent it would only mean that a set of the system was wrongly chosen as the set containing $\underline{0}$, $\underline{1}$, $\underline{2}$ etc. NF contains the set $\underline{Nn}$ defined as $\{x; (z)[0 \epsilon z.(w)(w \epsilon z \supset S(w) \epsilon z)] \supset x \epsilon z\}$. The only difference between this set and the intersection set I have been using is the different definition that the successor of a set has in NF. Quine supposes that there could be some formula of the formal system such that $\psi(\underline{0})$, $\psi(\underline{1})$, $\psi(\underline{2})$, ... are all provable and also $(\exists \underline{x})(\underline{x} \epsilon \underline{Nn} . \sim\psi(\underline{x}))$. This situation, should it occur, would mean that $\underline{Nn}$ must contain other sets than $\underline{0}$, $\underline{1}$, $\underline{2}$, etc. The notion of ω-inconsistency loses its importance (in set theory) because we may be able to choose another set which will contain $\underline{0}$, $\underline{1}$, $\underline{2}$, ... but no $\underline{x}$ such that $\sim\psi(\underline{x})$, e.g. $\{\underline{x}; \underline{x} \epsilon \underline{Nn} . \psi(\underline{x})\}$. This process of continual refinement maynot end, in which case, Quine says, we may say that the system is <u>numerically insegrative</u>, i.e. the system fails to contain a proper translation of 'x is a natural number'.[2]

Reinterpreting Quine's argument for the translations put forward here, if there is a formula $\psi(\underline{x})$ such that $\psi(\phi)$, $\psi(\{\phi\})$, etc. are all provable and $(\exists \underline{x})(\underline{x} \epsilon \underline{N} . \sim\psi(\underline{x}))$ is

1. W. Quine, 'On ω-Inconsistency and the so-called axiom of Infinity' <u>Journal of Symbolic Logic</u>, 1953.
2. ibid.

provable then $\underline{I}$ should not be translated as the set consisting solely of ϕ, $\{\phi\}$, etc. But if we accept the idea of an intended model in which there is a set consisting solely of ϕ, $\{\phi\}$, etc. and Tarski's method for translating the formal language into 'realistic' terms, the set corresponding to $\underline{N}$ in the model <u>will</u> consist of just ϕ, $\{\phi\}$, etc.

Let us suppose that there is a formula of the formal system $\underline{\psi}(\underline{x})$ which has the above property. Now if we are prepared to say of the formula $(\underline{y})(\underline{y}\in\underline{N} \supset \sim \underline{R}(\underline{q},\underline{y}))$ that it expresses a true proposition in the intended model we must be equally prepared to say that $(\underline{y})(\underline{y}\in\underline{N} \supset \underline{\psi}(\underline{y}))$ expresses a true proposition of the intended model. In this case we must say that at least one of the axioms is false in the intended model, since, if they were all true in the intended model then so too would be all the formulas derived from them, one of which is $\sim(\underline{y})(\underline{y}\in\underline{N} \supset \psi(\underline{y}))$

This would be a surprising result because when we build up the axiom system we choose axioms which correspond as closely as possible to our 'intuition' of sets. We may not succeed in capturing all the 'correct' axioms but at least, we feel, we have not chosen any of the 'incorrect' ones.

But there seem to be worse consequences than this. If the axioms contain at least one which is false in the intended model, then it may be the case that $\underline{\psi}(\underline{a})$ where $\underline{a}$ is one of the sets ϕ, $\{\phi\}$, $\{\{\phi\}\}$ is false in the intended model. It may be remembered that the reason for considering such formulas to be true in the intended model was that they must be true in all interpretations in which the axioms were true. Since it turns out that the axioms are not true for the intended model, we have no good grounds for saying that each of $\psi(\phi)$, $\psi(\{\phi\})$, ... etc. are true. It might be thought, however, that there could be independent grounds for saying that they are each true, e.g. by inspecting the translations and seeing if they are true of the intended model. But there is good reason to suppose that $\psi(x)$ would be more complex than any of the axioms. That is, it would be harder to tell whether it was true of some particular set in the model than to tell whether the axioms are true in the intended model. It may, for example, contain more quantifiers than any of the axioms and be notationally longer than them. Consider the case of $\underline{R}(\underline{q},\underline{y})$. The expanded version of this will not even be surveyable in Wittgenstein's sense. Now, if $\psi(\underline{a})$ does not express a true proposition of the intended model where a is one of the sets ϕ, $\{\phi\}$, $\{\{\phi\}\}$, etc. we no longer have grounds for saying that $(\underline{y})(\underline{y}\epsilon\underline{N} \supset \psi(\underline{y}))$ expresses a true proposition of the intended model, for our grounds for saying this was that $\psi(\underline{a})$ expressed a true proposition of the intended model. In fact, if $\psi(\underline{a})$ were to express a false proposition of the intended model, $(\underline{y})(\underline{y}\epsilon\underline{N} \supset \psi(\underline{y}))$ would express a false proposition. For $\sim\underline{\psi}(\underline{a})$ would express a true proposition and therefore $(\exists\,\underline{y})(\underline{y}\in\underline{N}\,.\sim\underline{\psi}(\underline{y}))$ would be true, i.e. $(\underline{y})(\underline{y}\in\underline{N}\supset\psi(\underline{y}))$ would express a false proposition of the intended model. This is in direct contradiction to the initial position when we said that it expressed a true proposition.

Such a consequence is intolerable. What were the suppositions on which it rested? One was that such a formula existed. Now we could reject this, but to do so would be foolhardy. Formal systems have a habit of producing the most unlikely consequences. It is not beyond the bounds of possibility for such a formula to turn up. There is no guarantee that one will not, for if we could prove that a formal set theory is ω-consistent we would have a proof of its consistency. We know that we can have no proof of consistency which employs only those methods available in the formal theory. Any methods that we employ which are stronger than those in the system may themselves be inconsistent.

The other supposition was that we should accept that $(\underline{y})(\underline{y}\in\underline{N}\supset\underline{\psi}(\underline{y}))$ as expressing a true proposition in the intended model. We expressed doubts about the legitimacy of intended or standard models since they can never be completely specified. The notion of an intended model is unclear. Since the above argument produces a contradiction on the assumption of the formal language being ω-inconsistent (relative to N) if we allow the notion of an intended model to settle the truth and falsity of the sentences of the formal language, it would be a wiser course to abandon speaking of a formula expressing a true (or false) proposition of the intended model.

Once we do so, we can no longer talk of an unprovable formula expressing a true proposition of set-theory. Thus the claim that there are true but unprovable formulas of any formalised set theory can no longer be made.

There is no mystery about this. If we bear in mind that a set theory is only committed to what it says exists, thdre remain propositions that are not expressible by means of any formalised set theory. One such proposition is that there exists a set which contains only the nųl set and sets generated from it by means of the operation of forming the unit set. As we know from Henkin's work[1] it will always be possible to add to the axioms of any set in which Peano's axioms are derivable (relative to some set $\underline{N}$ of the theory) the set of axioms $\underline{a}\in\underline{N}$, $\underline{a}\neq\underline{0}$, $\underline{a}\neq\underline{1}$, $\underline{a}\neq\underline{2}$, ... without inconsistency. Indeed Skolem has given a model of Peano's axioms which, though denumerable, is not isomorphic with the natural numbers.[2] The consequence of Gödel's and Henkin's completeness theorems is that there are

1. L. Henkin, 'Completeness in the Theory of Types'
2. T. Skolem, 'Peano's Axioms and Models of Arithmetic', Mathematical Interpretations of Formal Systems, pp.1-14. (This paper utilises results published by Skolem in 1934).

propositions which are not even expressible in a formalised set-theory.[1]

---

1. R. Goodstein makes the remark that what Gödel's theorem shows for a formalised arithmetic is not that there is a true but unprovable formula but that the universal quantifier does not express 'for all'. R. Goodstein, 'The Significance of Incompleteness Theorems', British Journal for the Philosophy of Science, vol. xiv, 1963.

II

Recently certain axioms of set theory have been shown to be independent. Cohen[1] has produced proofs that show the axiom of choice is not derivable from the other axioms of Zermelo's axioms. Together with Gödel's result[2] that the axiom of choice is consistent with the other axioms, its independence is established. Cohen has proved that a simiiliar result holds for Cantor's continuum hypothesis.

The independence of both the axiom and the hypothesis shows clearly that we are free to choose either (or their negations) as axioms for set theory. Opposed to such a view, the realist argues that there are sets (in some none too clearly defined sense) and that the mathematician's job is to describe them and their behaviour. Consequently either the axiom of choice is true of this reality or it is not; if it is true we should adopt it as an axiom, if it is false we should adopt the negation of the axiom. Yet even if we accept the existence of this reality which set theory is supposed to be describing what kind of evidence can point either to its truth or its falsity? Both Cohen[3] and Godel[4] are realists and both suggest that the

1. P. Cohen, <u>Set Theory and The Continuum Hypothesis</u>, New York, 1966
2. K. Gödel, 'The Consistency of the Axiom of Choice and of the Generalised Continuum Hypothesis', <u>Proceedings of the National Academy of Sciences U.S.A.</u> 1938
3. P. Cohen, op.cit. p.151
4. K. Gödel, 'What is Cantor's Continuum Hypothesis?', <u>American Mathematical Monthly</u>, 1947

continuum hypothesis or its negation may be derived from some other proposition which we can see describes the reality. But now we would want to know how this 'higher' axiom can be seen to be a description of this reality. Cohen mentions another argument which he thinks might be used by future generations to show that the continuum hypothesis is 'obviously false'[1] This is that Zermelo's axiom of power sets gives sets that cannot be reached by means of the other axioms of set theory. Again one wants to know how it could be seen that such sets cannot be reached by such means. Eventually realists would have to fall back on the self-evidence of such axioms or at least on some metaphor or analogy. But self-evidence as a criterion of truth has obvious drawbacks[2] and no metaphor can compel us to accept that it is the most appropriate metaphor.

We are free to chose the continuum hypothesis or its negation as we like. Whichever way is chosen the result is that the rules for the use of 'is a member of' are specified further in the field of abstract sets. In section 1 I said that the construction of abstract set theory involved the setting up of rules for the use of 'is a member of' and 'set'. Our choice

1. P. Cohen, op.cit. p.151
2. see, for example, S. Korner, The Philosophy of Mathematics, p.135 et seq.

of rules would be guided by the use of 'set' and 'is a member of' as they commonly occur outside of set theory. Clearly we want a set theory which will have application, particularly in the area of natural number or real number theory. Thus we will be guided by the needs of these disciplines. But, then, as we have seen, the creation of this set theory creates its own problems. The syntactic paradoxes arise at precisely the point where the notion of 'set of' is replaced by the notion of an abstract set. The failure of the axiom of comprehension indicated only that care must be taken in setting up the rules for the use of 'is a member of' in this field. The different, non-equivalent axiomatic set-theories that have been set up each replace the axiom of comprehension by a set of existential axioms. Which one is the correct one? The question makes no sense without there being some set theory with which we are comparing all the different axiomatic set theories. In the last chapter ~~the difficulties inherent in holding theories and in the last chapter~~ we saw the difficulties involved in such an approach. There is nothing to force us to use one set theory rather than another except for ease of applicability or aesthetic preference.

Each axiom of abstract set theory chosen is not so chosen because of its 'truth'; rather, each chosen axiom reflects a <u>decision</u> to use 'is a member of' and 'set' in a certain way.

We may use some metaphor to explain why one has made this particular decision, but one is always free to reject the metaphor. We can see this at a stage earlier than the continuum hypothesis, particularly in the case of the axioms of foundation and replacement.[1] In Zermelo's set theory (excluding the axiom of foundation) it is not possible to show that there is no set which belong to itself, nor is it possible to show that non-grounded classes do not exist.[2] We may argue that it is clear that the members of each set may contain members and that these members may contain members and so on. But we may feel that there must be a layer which is fundamental and contain no members. Hence we adopt an axiom which will stop such sets arising. That is, we make a decision that no such sets exist, we do not find out that no such sets exist. For what would count as a proof that no such sets exist? Again the axiom of replacement can be justified only by recourse to such arguments as, given that there is an infinite set Z and the axiom of power set producing the sets UZ, UUZ, UUUZ, etc. why should there not be a set consisting of all of these. The analogy here is with the axiom of infinity, since this gives us a set consisting if all the sets $\emptyset$, $\{\emptyset\}$, $\{\{\emptyset\}\}$, etc. But there is nothing to force

1. As formulated in Fraenkel and Bar-Hillel, Foundations of Set-Theory pp.85-91
2. In the sense that it is not inconsistent to suppose that there are such sets.

us to accept that such a set exists. When we accept the axiom we make a decision that such a set exists; we make a decision to use 'is a member of' and 'set' in a way which was not determined by the other decisions we had taken.

The axiom of infinity which asserts the existence of a set with infinitely many members can also be rejected. Strict finitists would reject it. Of course such a rejection would mean that we could not construct mathematics in such a set theory but there is no reason for supposing that mathematics must be constructible in set theory. There are some who prefer that set theory dhould be capable of containing mathematics. Quine, for example, prefers a homogeneous universe to a heterogeneous one and will naturally prefer a set theory which will explicate numbers in terms of sets. But there is no logical necessity for a set theory to contain mathematics.

However, if we do reject the axiom of infinity we are on the point of departing from the normal use of 'is a member of' for we do talk of the set of all natural numbers and the natural numbers are infinite. We are parting company with the ordinary use of 'set' and 'is a member of' which we take as a guide for constructing our set theory. Whereas the ordinary use of the phrases gave us no guide with the other axioms it does so here. Even so, if we take a set theory in which there are no individuals, e.g. the Zermelo-Fraenkel axioms, the axiom of infinity is stated in terms of the nul set and sets generated out of the nul set by some set operation. Now talk of the nul set itself reflects a decision to use 'set' in a certain way even though such talk is so familiar that we tend to forget this. We have decided that we can talk of a set which has no members which is perhaps unjustified by any 'ordinary' use of 'set'. Whether there exists a set which contains this set and all unit sets obtained from it is not then determined by the normal use of 'set' or 'is a member of' since the existence of the nul set itself is not so ddtermined. Similar remarks apply to the other axioms of Zermelo-Fraenkel set theory. But once we have accepted this extension of 'set' then we have no choice but to accept the axioms of sum-set, pairing etc., if we are using the pre-set-theoretic use of 'is a member of' and 'set' as our guide.

When we speak of some proposition of set theory being true, we are tacitly understanding a particular set theory in which that proposition is a theorem. No sense can be made of the question: 'I know it is a theorem but is it true?' The only sense such a question could have would be: 'I know it is a theorem of the set theory S_1 but is it also a theorem of set theory S_2'. It is our decision to use 'set' and 'is a member of' in a particular way which determines which propositions containing only such phrases are true.

Appendix

Tarski's axiom of infinity as given in CTF (p.243) seems to contain an error. As stated there it could be translated as: There is a non-empty set z such that if $x \epsilon z$ there is a proper sub-set of x which is also a member of z. This clearly is not what Tarski intended for he says that such an axiom 'guarantees the existence of infinitely many individuals'. As stated by Tarski, the set postulated by the axiom of infinity cannot contain even one individual, since individuals do not possess proper sub-sets.

We may correct Tarski's axiom by altering the negation bar in two places. Then the axiom will conform to Quine's interpretation of the axiom as given by him in Set Theory and its Logic, p.280. Quine explicitly states that this is Tarski's axiom but Quine's axiom can be interpreted as: there is a non-empty set z such that if $x \epsilon z$ then there is a set of which x is a proper subset which also belongs to z. With the alterations given above Tarski's axiom can indeed be given this interpretation.

Bibliography

Only books referred to in the text are listed.

Beth, E. The Foundations of Mathematics, Amsterdam, 1959.

Black, M. 'Russell's Philosophy of Language' in The Philosophy of Bertrand Russell, ed. P. Schilpp, La Salle, pp.229-255.

——— 'The Semantic Definition of Truth', Analysis 1948 pp.49-63.

Bochvar, D. 'To the Question of Paradoxes of the Mathematical Logic and Theory of Sets' Mat. Sbornik 15 pp.365-384.

Boyer, C. The Concepts of the Calculus, 1949.

Carnap, R. Introduction to Semantics, Cambridge, Mass. 1942

——— 'Empiricism, Semantics and Ontology', in Revue Internationale de Philosophie, 1950, pp.20-40.

——— Meaning and Necessity, Chicago, 1956.

Church, A. Introduction to Mathematical Logic, Princeton, 1956.

Cohen, P. Set Theory and the Continuum Hypothesis, New York, 1966.

Curry, H. Foundations of Mathematics and Logic, New York, 1963.

——— Introduction to Mathematical Logic

Dummet, M. 'The Philosophical Significance of Gödel's Theorem'.

Einstein, A. Relativity, London, 1920.

Fitch, F. 'Self-Reference in Philosophy' Mind, 1946, pp.64-73.

Fraenkel, A. and Bar-Hillel, Y. Foundations of Set Theory, Amsterdam, 1958.

Fraenkel, A. Abstract Set Theory Amsterdam 1961.

Goddard, L. '"True" and "Provable"' Mind 1958, pp.13-31

Gödel, K. 'On Formally Undecidable Propositions of Principia Mathematica' (1931) trans. B. Meltze, 1962.

——— 'The Consistency of the Axiom of Choice and of the Generalised Continuum Hypothesis', Proceedings of the National Academy of Sciences U.S.A. 1938.

——— 'Russell's Mathematical Logic' in The Philosophy of Bertrand Russell, ed. P. Schilpp, La Salle, 1944. pp.125-153.

——— 'What is Cantor's Continuum Hypothesis?' American Mathematical Monthly, 1947.

Goodstein, R. 'Proof by Reductio ad Absurdum', Mathematical Gazette, 1948. pp.198-204.

——— Mathematical Logic, Leicester, 1957.

——— 'The Significance of Incompleteness Theorems' British Journal for the Philosophy of Science, vol.XIV, 1963.

Henkin, L. 'Completeness in the Theory of Types', Journal of Symbolic Logic, 1950, pp.81-91

Heyting, A. Intuitionism - an Introduction, Amsterdam, 1956.

——— 'Some Remarks on Intuitionism' in Constructivity in Mathematics, Amsterdam 1959 pp.69-71.

Hilbert, D. 'On the Infinite' in Philosophy of Mathematics ed. P. Bonacerrof and H. Putnam, Oxford 1964.

Hilbert, D. and Ackermann, W. Principles of Mathematical Logic (1928) New York 1950.

Hintikka, J. 'Identity, Variables, and Impredicative Definitions' Journal of Symbolic Logic, 1956, pp.225-245.

——— 'Vicious Circle Principle and the Paradoxes', Journal of Symbolic Logic, 1957, pp.245-249.

Huntington, E. 'Sets of Independent Postulates for the Algebra of Logic', Transactions of the American Mathematical Society, 1904 pp.288-309

Kleene, S. Introduction to Metamathematics, Amsterdam, 1952

Kneale, W. and Kneale, M. The Development of Logic, Oxford, 1962.

Kolmogorov, A. 'On the Principle of Excluded Middle' (1925) in From Frege to Gödel, ed. J. von Heijenoort, Cambridge, Mass. 1967 pp.420.

Körner, S. The Philosophy of Mathematics, London 1960

——— 'On the Relevance of Post-Gödelian Mathematics to Philosophy', Problems in the Philosophy of Mathematics ed. I. Lakatos, Amsterdam, 1967.

Kreisel, G. 'Informal Rigour and Completeness Proofs' in <u>Problems in the Philosophy of Mathematics</u>, ed. I. Lakatos, Amsterdam, 1967.

Luschei, E. <u>The Logical Systems of Lesniewski</u>, Amsterdam 1962

Myhill, J. 'The Ontological Significance of the Löwenheim-Skolem Theorem' <u>Academic Freedom, Logic and Religion</u>, Philadelphia 1953 pp.57-70

Nagel, E. and Newman, J. <u>Gödel's Proof</u>, London 1958

Neumann, von. J. 'Zur Einführung der transfiniten Zahlen', <u>Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae</u> I 1923 pp.199-208

Nidditch, P. <u>Elementary Logic of Science and Mathematics</u> London, 1960.

Peter, R. 'Rekursivitat und Konstruktivitat' in <u>Constructivity in Mathematics</u> Amsterdam 1959, pp.226-233

Prior, A. 'The Run-about Inference Ticket', <u>Analysis</u> 1960 pp.38-39.

Quine, W. 'New Foundations of Mathematical Logic', (1937) in <u>From a Logical Point of View</u>, Cambridge, Mass 1953 pp.80-101.

——— Mathematical Logic, New York, 1940.

——— 'Whitehead and the Rise of Mathematical Logic' (1941) in <u>Selected Logical Papers</u>, New York, 1966 pp.3-36

——— 'On ω-Inconsistency and the so-called Axiom of Infinity' Journal of Symbolic Logic 1953 pp.119-124.

——— 'On Frege's Way Out' Mind, 1955, 145-159

——— Word and Object New York, 1960.

——— 'Carnap and Logical Truth', in Studies in Logic and Language Dordrecht, 1962, pp.39-63.

——— Set Theory and its Logic, Cambridge, Mass. 1963.

Ramsey, F. 'The Foundations of Mathematics ed. R. Braithwaite, London, 1931.

Rosser, J. 'Extensions of some Theorems of Gödel and Church', Jourhal of Symbolic Logic, 1936 pp.87-91.

Russell, B. 'Letter to Frege' (1902) in From Frege to Gödel ed. J. von. Heijenoort, Cambridge, Mass. 1967, pp.124-125.

——— 'On Some Difficulties in the Theory of Transfinite Numbers and Order Types' Proceedings of the London Mathematical Society, 1907, pp.29-53.

——— 'Mathematical Logic as based on the Theory of Types', (1908) in Logic and Knowledge, ed. R. March, 1956 pp.59-102

——— Introduction to Mathematical Philosophy, London 1919

——— An Enquiry into Meaning and Truth, London 1940

Russell, B. and Whitehead, A. Principia Mathematica Cambridge, 1913

Ryle, G. Dilemmas, Cambridge, 1954

——— 'Heterologicality', Analysis 1961.

Sierpinski, W. Cardinal and Ordinal Numbers, Warsaw 1958.

Skolem, T. 'Einige Bernerkungen zur axiomatischen Begründung der Mengenlehre', Matematikerkongressen i Helsinkfors den 4-7 Juli 1922, Den femte Skandinavska matematikerkongressen, Redogomelre, Akadimiska Bokhandeln, Helsinki, 1923.

——— 'Peano's Axioms and Models of Arithmetic', Mathematical Interpretations of Formal Systems, ed. A. Heyting Amsterdam 1955 pp.1-14

Sommers, F. 'Types and Ontology' Philosophical Review, 1963 pp.327-363.

Strawson, P. 'Truth' Analysis, 1949, pp.83-97

——— Introduction to Logical Theory, London 1952.

Sxmielew, W. Review of Bochvar, Journal of Symbolic Logic, 1946, p.129.

Tarski, A. 'Der Wahrheitsbegriff in den formalisierten Sprachen' Studia Philosophica, 1935, pp.261-405.

——— 'The Establishment of Scientific Semantics', in Logic, Semantics and Metamathematics, tr. J. Woodger Oxford, 1956, pp.401-408.

——— 'The Concept of Truth in Formalised Languages' in Logic, Semantics and Metamathematics pp.152-278

——— 'The Semantic Conception of Truth', Philosophy and

Phenomenological Research, 1944, pp.341-375.

Tennesson, H. 'Permissible and Impermissible Locutions' Studies in Logic and Language Dordrecht, 1962, pp.220-233

Thomson, J.F. 'On Some Paradoxes' in Analytic Philosophy, 1962, pp.104-119.

Tucker, J. 'The Formalisation of Set Theory', Mind, 1963, pp.500-518.

——— 'Philosophical Argument', Supplementary Volume XXXIX 1965, The Aristotelian Society.

——— 'Constructivity, Consistency and Natural Languages' Proceedings of the Aristotelian Society, 1967 pp.145-168

Veblen, O. 'Foundations of Geometry' in Monographs on Topics of Modern Mathematics ed. J. Young, New York, 1955 pp.3-49

Waismann, F. 'Verifiability' in Essays on Logic and Language, ed. A. Flew, Oxford, 1951

——— The Principles of Linguistic Philosophy, ed. R. Harre London 1965.

Wang, H. 'Formalisation of Mathematics' Journal of Symbolic Logic 1954, pp.241-266.

——— 'On Denumerable bases of Formal Systems' Mathematical Interpretations of Formal Systems ed. A. Heyting Amsterdam, 1955.

——— 'Ordinal Numbers and Predicative Set Theory' (1959) in A Survey of Mathematical Logic, Peking 1963.

——— A Survey of Mathematical Logic, Peking 1963.

Weiss, P. 'The Theory of Types' Mind, 1928 pp.338-348

White, M. 'The Analytic and Synthetic' in John Dewey: Philosopher of Science and Freedom, New York, 1950 pp.316-330

Wittgenstein, L. Remarks on the Foundations of Mathematics, ed. G. von Wright, trans. G. Anscombe, Oxford 1956.

Wright, von. G. 'The Heterological Paradox' Societas Scientiarum Fennica Commentationes Physico-Mathematicae 1960, pp.1-28.

Zermelo, E. 'Untersuchenden über die Grundlagen der Mengenlehre I Mathematische Annelen 1908, pp.261-281.