As we have said, P is a formal language. This means, in part, that its signs and formulae are considered to be completely meaningless. Examining a language in this way facilitates the study of its syntax and prevents semantic issues from obtruding. It is possible, however, to assign a meaning to each and every sign of P and hence to every formula of P as well. This process is usually called 'interpreting' the language. By 'standard interpretation of P ' we shall mean the result obtained by attaching to each of the signs of P the meaning that is regularly attached to it. While the second and third versions of Gödel's proof that will be presented require no mention of an interpretation of P the first does (this is one of its distinguishing features). Let us now turn to this first version and try to determine the mechanisms on which Gödel's proof is based.

§2 Version I

As has already been indicated this version of the proof is not so much to be taken as a slightly modified duplication of it but rather as an heuristic device pointing to some of the mechanisms underlying the undecidability results.

From what has been said so far it should be clear that to every formula of P there corresponds both a unique number, the formula's Gödel number, and a unique proposition, its standard interpretation. This circumstance sets the stage for Gödel's project: what Gödel shows is that there exists a

39035

National Library
of CanadaBibliothèque nationale
du CanadaCANADIAN THESES
ON MICROFICHETHÈSES CANADIENNES
SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR

Robert John Gaizauskas

TITLE OF THESIS/TITRE DE LA THÈSE

Contradictions and Incompleteness: A Re-examination
of Gödel's Result

UNIVERSITY/UNIVERSITÉ

Carleton University

DEGREE FOR WHICH THESIS WAS PRESENTED/

GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE

M.A.

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ

1978

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE

Stephen Talmage

Permission is hereby granted to the NATIONAL LIBRARY OF
CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and neither the
thesis nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈ-
QUE NATIONALE DU CANADA de microfilmer cette thèse et
de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la
thèse ni de longs extraits de celle-ci ne doivent être imprimés
ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE

25/8/78

SIGNED/SIGNÉ

Robert Gaizauskas

PERMANENT ADDRESS/RÉSIDENCE FIXE

556 Westminster Ave.

Ottawa Ont.

K2A 2U3

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

CONTRADICTIONS AND INCOMPLETENESS:
A RE-EXAMINATION OF GÖDEL'S RESULT

by

Robert John Gaizauskas

A thesis submitted to the Faculty of
Graduate Studies and Research in
partial fulfillment of the requirements,
for the degree of Master of Arts.

Department of Philosophy
Carleton University
Ottawa, Ontario
April, 1978

© 1978 Robert John Gaizauskas

The undersigned recommend to the Faculty of
Graduate Studies and Research acceptance of
the thesis

"Contradictions and Incompleteness: A Re-
Examination of Gödel's Result" submitted by
Robert John Gaizauskas, B.A.,
in partial fulfillment of the requirements
for the degree of Master of Arts.

.....
Thesis Supervisor,

.....
Chairman
Department of Philosophy

What we call truth is an error insufficiently experienced, not yet drained, but which will soon age, a new error, and which waits to compromise its novelty.

Emile Cioran, A Short History of Decay.

Ξεῖνε, φιλητῶς εἰμί. λόγων ὁ ψευδόμενός με
ᾤλεσε καὶ νυκτῶν φροντίδες ἐσπέριοι.

- the epitaph of Philetas of Cos.

ABSTRACT

Three somewhat differing versions of Gödel's argument for the existence of undecidable formulae in the formal language P are presented in deductive form. Two immediately apparent objections to the claim that Gödel's result regarding P implies the incompleteness of every language capable of expressing arithmetic are examined and both are found not to be sustainable. An alternative metalanguage to the one Gödel uses to investigate P is described and it is argued that this metalanguage is to be preferred to the one Gödel uses. This metalanguage employs a concept of entailment according to which contradictions do not entail. Gödel's argument is shown not to be reproducible in this alternative metalanguage and it is concluded that his argument does not show P to be incomplete and, indeed, that there are no good reasons for believing P to be incomplete.

ACKNOWLEDGEMENTS

I should like to acknowledge, firstly and foremostly, Stephen Talmage to whom I am indebted for some of the ideas forming the core of this thesis, for constant and never unhelpful criticism, and for the encouragement without which I should never have finished. His acute intellect and passionate interest in philosophy have made it a privilege to work with him. Many thanks are also due, however, to my close friends -- Jan, Jim, Sharon, Bill, Michael, Kate -- and to my parents who, during the course of this project, have been tolerant of so many lapses on my part -- the result of an obsession which must have been very hard to understand. Further, I should like to thank both Christine Wirta who, in the face of pressure and a lot of nasty notation, typed the manuscript splendidly and Mrs. Joan Ricketts who drew every scopeline occurring in it. Finally, on a lighter note, I should like to thank the Melitta Coffee Company for keeping me alive and functioning for the last several months required to complete this thesis.

TABLE OF CONTENTS

INTRODUCTION	1
CHAPTER I GÖDEL'S ARGUMENT	6
§1 Preliminaries	7
§2 Version I	9
§3 Version II	12
§4 Version III	17
§5 The Three Versions: Comparison and Contrast	24
CHAPTER II GÖDEL'S RESULT AND THE COMPLETENESS OF ARITHMETIC	26
§1 The Truth of the Undecidable Formula	27
§2 The Definition of the Concept of Truth.	35
CHAPTER III CONTRADICTIONS AND ENTAILMENT: AN ALTERNATIVE METALANGUAGE	50
§1 Contradictions and Material Entailment.	53
§2 Natural Entailment	59
§3 Natural Entailment and the Conditional.	64
§4 The Liar Paradox	70
§5 Russell's Solution to the Liar Paradox.	79
CHAPTER IV GÖDEL'S ARGUMENT AND THE METALANGUAGE M_N	83
§1 The Three Versions Revisited	84
§2 The Analyticity of the Premises of Version II	88
§3 The Status of the M-Undecidable Formula	95
§4 Tarski's Argument and M_N^0	97

§5 Conclusions	101
APPENDIX I	104
BIBLIOGRAPHY	107

INTRODUCTION

In his famous paper on incompleteness¹ Gödel purports to demonstrate that P , a particular formal language capable of expressing arithmetic, is incomplete. It is on the basis of this result that the more general claim that arithmetic is incomplete is frequently asserted. Now Gödel's argument is itself presented in a language, generally called simply the metalanguage, which I shall call M . My intention, in this paper, is to show that Gödel's result concerning the incompleteness of P depends crucially on certain features of M . This I shall do by arguing that in a metalanguage M_N , differing somewhat from M , Gödel's argument is not reproducible. Thus we see that the completeness or incompleteness of a given formal language is relative to the language used to prove its completeness or incompleteness. Our conclusion, therefore, may be put as follows: while Gödel's argument does indeed show P to be M -incomplete it does not show P to be M_N -incomplete.

Given that we can distinguish at least two languages, M and M_N , which can be used to investigate P the further question arises as to whether one of them is to be preferred. I propose to argue not only that M_N is an alternative meta-

¹Kurt Gödel, 'Some Metamathematical Results on Completeness and Consistency, On Formally Undecidable Propositions of Principia Mathematica and Related Systems I, and On Completeness and Consistency', in From Frege to Gödel, ed. Jean van Heijenoort (Cambridge, Mass.: Harvard University Press, 1967), pp. 592-617.

language in which Gödel's argument is not reproducible but also that M_N is, in a sense that I shall establish, the preferred language for investigating P. Now it seems reasonable to assign the general designation 'complete' or 'incomplete' (as opposed to 'X-complete' or 'X-incomplete') only to a language which has been shown to be complete or incomplete in the preferred metalanguage. Adopting this convention I argue that given 1) the non-reproducibility of Gödel's argument in M_N and 2) the absence of any other considerations supporting the claim that P is M_N -incomplete, there are no grounds for believing that P is incomplete (showing that Gödel's argument is not reproducible in M_N is not, of course, sufficient to show that P is not M_N -incomplete).

The argument for these results takes the following form: The paper is divided into four chapters. The purpose of the first chapter is twofold: firstly, to provide a rough intuitive understanding of what is going on in Gödel's argument; secondly, to provide a deductive presentation of Gödel's argument which will make the structure of the argument completely clear (so far as I am aware the argument has never been presented in this manner before). The second chapter deals with a number of objections to the claim that Gödel's argument implies the incompleteness, not just of the one language capable of expressing arithmetic which he investigates, but of every language capable of expressing arithmetic and consequently of arithmetic itself. While it turns out that these objections cannot be sustained there are a number of

good reasons for discussing them: 1) they are objections that would occur to anyone trying to see how Gödel's argument implies the incompleteness of arithmetic and hence demand to be discussed 2) these objections provide a valuable contrast to the different sort of objection which I pursue in Chapters III and IV 3) they allow me to introduce an argument of A. Tarski's which, since it is closely related to Gödel's argument, falls, quite interestingly, within the scope of what I have to say about such arguments in Chapter IV. Because the approach of the second chapter which questions whether Gödel's argument must hold of all languages capable of expressing arithmetic turns out to be bankrupt, the possibility is suggested, at the outset of the third chapter, of avoiding Gödel's conclusion by adopting an alternative language for investigating languages capable of expressing arithmetic. The rest of the third chapter is devoted to introducing and justifying, on the grounds of accordance with intuition, the adoption of the language M_N . Amongst the consequences of adopting this new language is a new and, I believe, superior solution to the Liar paradox. This solution is presented in the third chapter both in order to further justify the adoption of M_N and to work out some features of M_N which, given the similarities between the Liar paradox and Gödel's argument, turn out to have significance for the latter as well. The fourth and final chapter demonstrates how Gödel's argument is not reproducible in M_N ; incidentally, Tarski's argument, presented in Chapter II, is also shown not to be reproducible in M_N .

I should like to make here some comment about the formalism and method of proof that will be used throughout. I have chosen to use a method of subordinate proof because I find it to be the most perspicuous of the proof procedures with which I am acquainted. My constant reference for this method has been Fitch's book Symbolic Logic² which I have followed as closely as I felt was practicable. In Chapters I and II the deductive versions of Gödel's and Tarski's arguments employ Fitch's scope line procedure fully but allow lines to succeed each other in a fashion which is not strictly in accordance with Fitch's rules. I justify these lapses on several counts: 1) Rigorous adherence to his rules would have made the proofs unmanageably long. 2) Putting all formulae in a form amenable to treatment according to the rules of Fitch's system would have obscured their connection with the original texts from which they were drawn (much of Gödel's argument, for instance, involves tacit quantification which would have to be made explicit in order for Fitch's rules governing quantification to apply). Since the deductive versions of Gödel's argument will almost certainly not be understandable apart from the original texts it was thought best not to unduly impede cross reference. Thus the line references to the far right in the proofs in Chapters I and II indicate the lines from which the line on which the reference occurs was inferred, but no attempt is made to indicate the

²Frederic Brenton Fitch, Symbolic Logic (New York: The Ronald Press Company, 1952).

rule of inference in accordance with which the inference was made. In Chapter III, on the other hand, I think I have followed strictly Fitch's procedures, indicating in every case both the lines from which each formula is inferred and the rule according to which the inference is made. Occasionally principles of inference are introduced which Fitch does not mention but these are clearly labelled and should raise no difficulties.

CHAPTER I

GÖDEL'S ARGUMENT

By Gödel's argument I mean the proof presented in 'On Formally Undecidable Propositions of Principia Mathematica and Related Systems' for his first major result--Theorem VI--which, if translated out of its arithmetized form, asserts roughly that there are formulae of P (P is the formal language Gödel is investigating) such that if P is consistent neither these formulae nor their negations are provable in P .¹ The argument for this assertion takes the form of demonstrating that a formula X of P with certain specific characteristics can be constructed.

The purpose of this chapter is to provide both a rough intuitive understanding of Gödel's argument and a deductive presentation of that argument which will put it in a form readily accessible to analysis.

One of the chief difficulties involved in fulfilling the second of these goals is in dealing with the diversity of symbols and approaches found in existing accounts of Gödel's argument: one author proves the result for a language of order 1, uses set notation and exhibits a taste for Gothic German script; another proves the result for languages of order ω , avoids set notation, uses only Roman script, and so on. These divergences make it extremely difficult, without developing a

¹Gödel, pp. 607-609.

totally new account, to capture all the relevant details through the exposition of just one author, or by adopting just one symbolization. Consequently, what I propose to do here is to present, in deductive form, three 'versions' of a proof for Gödel's result. The first, actually, is sufficiently different from Gödel's proof as to only somewhat misleadingly be called a version of it. Its purpose is to demonstrate clearly and in an intuitive way how the formula Gödel specifies is undecidable. The second version has, in part, a historical motivation: it is a reproduction, in deductive form, of Gödel's own proof. The third version claims the virtue of abstractness or generality: it shows how undecidability is the result of certain general logical relations.

In order to establish his result Gödel has to carry out certain preliminaries, most significantly the construction of the formalized language P and the provision of a method for arithmetizing it. I shall discuss these in the next section.

§1 Preliminaries

Gödel provides an exhaustive and exact description of P -- signs, definitions, axioms, and rules of inference -- which is too involved to be repeated here. This description concerns itself solely with the signs of P and the rules for their combination, ignoring completely the meanings these signs may have. P is thus a formal language in two senses: 1) its signs and the rules concerning their use are exhaus-

tively specified. 2) its signs are left entirely devoid of meaning in order that their syntactic properties only may be considered. Once the formal language P has been fully characterized a method of arithmetizing it is presented. A unique mapping of numbers onto signs of P and onto groups of signs of P is supplied. This mapping makes it possible both to uniquely assign to every formula of P (a particular sort of group of signs of P) a specific number and in turn to determine, of any number whether it is the number of any P -formula and, if so, of which. Furthermore, this scheme of mapping allows syntactical properties of formulae to be expressed as number properties of the numbers assigned to the formulae. Thus, for example, since provability is a purely syntactic notion in P , i.e. a P -formula is or is not provable depending on the nature and relations of the signs of which it is composed and regardless of the meaning which may be attached to them, provability can, by the device of the mapping scheme, be expressed as a number property holding of those numbers which are numbers of provable formulae of P . In particular, if we agree to call the formula of P which will be shown to be undecidable X , and agree that the number assigned to it by the mapping scheme (its Gödel number) is N , and let the property of being f be the number property possessed by a number if and only if it is the Gödel number of a provable formula of P , then the assertion that N is not f is logically equivalent to the assertion that X is not provable in P (X is not P -provable).

As we have said, P is a formal language. This means, in part, that its signs and formulae are considered to be completely meaningless. Examining a language in this way facilitates the study of its syntax and prevents semantic issues from obtruding. It is possible, however, to assign a meaning to each and every sign of P and hence to every formula of P as well. This process is usually called 'interpreting' the language. By 'standard interpretation of P ' we shall mean the result obtained by attaching to each of the signs of P the meaning that is regularly attached to it. While the second and third versions of Gödel's proof that will be presented require no mention of an interpretation of P the first does (this is one of its distinguishing features). Let us now turn to this first version and try to determine the mechanisms on which Gödel's proof is based.

§2 Version I

As has already been indicated this version of the proof is not so much to be taken as a slightly modified duplication of it but rather as an heuristic device pointing to some of the mechanisms underlying the undecidability results.

From what has been said so far it should be clear that to every formula of P there corresponds both a unique number, the formula's Gödel number, and a unique proposition, its standard interpretation. This circumstance sets the stage for Gödel's project: what Gödel shows is that there exists a

formula X of P whose Gödel number is N and whose standard interpretation is ' N is not f ', where the property of being f , as before, is the number property possessed by those numbers which are numbers of provable formulae of P . Since it is the case that if a P -formula is provable the proposition corresponding to it is true, then if X is provable, N is not f or, equivalently, X is not P -provable. Thus X cannot be P -provable. A few further steps suffice to show that not- X is also not P -provable.

The deductive demonstration which follows is based on a version of the proof first suggested by S. Talmage. It is nothing more than an exhibition of how a certain P -formula whose Gödel number is N , and to which corresponds the proposition that N is not f , is not provable and how its negation also is not provable. This demonstration assumes that such a P -formula exists and can be specified -- an assumption which the versions presented in the next two sections will prove to be the case.

In the following 'S-proposition corresponding to X ' denotes the proposition expressed by the P -formula X if it is interpreted in the standard way. The first three of the following eight premises specify the nature of the formula X . The remaining five are of the form of theorems and could have been introduced as rules of the system rather than as explicit premises.

1	X is a P-formula	hyp
2	N is the Gödel number of X	hyp
3	The S-proposition that corresponds to X is the proposition that N is not f	hyp

4	A P-formula is P-provable if and only if its Gödel number is f	hyp
5	If a proposition is P-provable then it is true	hyp
6	The proposition that q is true if and only if q	hyp
7	A S-proposition is P-provable if and only if the P-formula to which it corresponds is P-provable	hyp
8	A S-proposition corresponding to a P-formula is true if and only if the S-proposition corresponding to the negation of the P-formula is not true	hyp
9	X is P-provable if and only if its Gödel number is f	1, 4
10	X is P-provable if and only if N is f	2, 9
11	X is P-provable	hyp
12	The S-proposition corresponding to X is P-provable	1, 7, 11
13	The proposition that N is not f is P-provable	3, 12
14	The proposition that N is not f is true	5, 13
15	N is not f	6, 14
16	N is f	10, 11
17	If X is P-provable then N is not f and N is f	11, 15, 16
18	It is not the case that both N is not f and N is f	Law of Non-Contradiction
19	X is not P-provable	17, 18
20	The Gödel number of X is not f	1, 4, 19
21	N is not f	2, 20
22	The proposition that N is not f is true	6, 21
23	The S-proposition that corresponds to X is true	3, 22
24	The S-proposition corresponding to not-X is not true	8, 23
25	The S-proposition corresponding to not-X is not P-provable	5, 24

26	Not-X is not P-provable	7, 25
27	X is not P-provable and not-X is not P-provable and the S-proposition corresponding to X is true	19, 23, 26
28	There is a P-formula which is not P-provable and whose negation is not P-provable and whose corresponding S-proposition is true	27

§3. Version II

The preceding version shows clearly why it is that a P-formula with certain characteristics will be undecidable. It does not show us, however, how to prove that such a formula exists or how to identify one. To do this we must turn to a fuller proof. At this point, therefore, I should like to present a deductive version of Gödel's original proof. In order to do so I have found it necessary to formalize, to some extent, the metalanguage in which his proof is given. This formalization has not been performed systematically or exhaustively; rather just enough has been done to allow the proof to be expressed in symbolic terms. A description of this formalization will be found below.

The following demonstration presupposes all of Gödel's symbolism, all the apparatus of P, and all of the forty-five recursive definitions leading up to the syntactic definition of provability. Consequently, this demonstration is not meant to be intelligible independently of Gödel's work; its purpose is to present his proof in a deductive form so as to make it accessible for analysis later on.

One small divergence between Gödel's Theorem VI and what is proved here should be noted before the proof is presented: Gödel shows that there are formulae² which are not k-provable and whose negations are not k-provable, where a formula is k-provable if 1) it is the immediate consequence of the set of formulae containing all the formulae of a specified set k of formulae and all the axioms of P or 2) it is one of the members of the set consisting of k and the axioms. In our demonstration we shall consider the case in which k is empty. Thus 'provable' shall apply only to axioms and to their immediate consequences.

There are seven premises in our proof. Let me describe each of them briefly as an aid to exposition. The first premise states that P is ω -consistent. The second is the definition of the particular relation used to construct the undecidable formula. The third states that the relation and the function from which the relation defined in the second premise is built are recursive. The fourth says that any relation obtained from a recursive relation by substitution of recursive functions for variables is recursive. The fifth is Gödel's

²Throughout the paper I have adopted the procedure of referring to notions of the arithmetized metalanguage by underlining the term which designates that notion in the non-arithmetized metalanguage. Thus, for example, while, with reference to P, 'formula' denotes a certain sort of group of signs of P, 'formula' denotes the class of numbers which are associated, through the Gödel numbering scheme, with those and only those numbers which are formulae of P. Gödel uses small capital lettering to effect the same distinction; Mostowski uses double-spaced lettering. I have chosen to use underlining throughout due to typographical limitations and in order to promote uniformity.

Theorem V.³ The sixth says simply that a formula is provable if and only if there exists something which has the relation of 'being a proof of' to the formula which it is asserted is provable. The seventh says that if something is a proof of the generalization of a given formula with respect to a given variable then the formula resulting from substituting for the variable with respect to which the formula is generalized the numeral denoting the number which is the proof of this generalization is provable.

The following symbols have been introduced in order to properly formalize Gödel's proof:

1. The predicate letter 'C' for 'is recursive'. (Recursivity is fully defined by Gödel.)
2. Signs for variables: (i) small Roman letters -- 'f', 'g' -- for functions. (ii) capital Roman letters -- 'R', 'S', 'Q' -- for relations. N.B. The Greek letter ' λ ' will serve as the functional and relational abstract. Thus ' $\lambda f(x)$ ' and ' $\lambda Q(x,y)$ ' will name a function and a relation respectively while ' $f(x)$ ' and ' $Q(x,y)$ ' will denote the value the function $\lambda f(x)$ takes for the argument x and the proposition that the relation $\lambda Q(x,y)$ holds between x and y respectively. (iii) capital Roman letters -- ' X_n ', ' Y_n ' -- for n -tuples of numbers. (i.e. ' X_n ' stands for the n -tuple $(x_1, \dots, x_n)$)

³Gödel, pp. 606-607.

The proof:

key: $r = \text{Sb}(q \frac{19}{Z(17\text{Gen}q)})$

1	$-(\exists a) ((n) \text{Bew}(\text{Sb}(a \frac{v}{Z(n)})) \cdot \text{Bew}(\text{Neg}(v\text{Gena})))$	hyp
2	$\lambda Q(x, y) \equiv \lambda x \text{BSb}(y \frac{19}{Z(y)})$	hyp
3	$C(\lambda(xBy)) \cdot C(\lambda(\text{Sb}(x \frac{y}{w})))$	hyp
4	$(\lambda R) (((\lambda R \equiv \lambda \text{Sub}(S \frac{x_1, \dots, x_n}{f_1, \dots, f_n})) \cdot C(\lambda S))$ $\cdot C(\lambda f_1) \cdot \dots \cdot C(\lambda f_n)) \supset C(\lambda R))$	hyp
5	$(\lambda R) (C(\lambda R) \supset (\exists r)(x_n) ((R(x_1, \dots, x_n)$ $\supset \text{Bew}(\text{Sb}(r \frac{u_1, \dots, u_n}{Z(x) \dots Z(x_n)}))$ $\cdot ((\bar{R}(x_1, \dots, x_n)$ $\supset \text{Bew}(\text{Neg}(\text{Sb}(r \frac{u_1, \dots, u_n}{Z(x_1), \dots, Z(x_n)}))))))$	hyp
6	$(n) (\text{Bew}(n) \equiv (\exists m) m\text{Bn})$	hyp
7	$\text{Bew}(x\text{Gen}y) \supset (n) \text{Bew}(\text{Sb}(y \frac{x}{Z(n)}))$	hyp
8	$C(\lambda Q(x, y))$	2, 3, 4
9	$Q(x, y) \supset \text{Bew}(\text{Sb}(q \frac{17}{Z(x)} \frac{19}{Z(y)}))$ $\therefore \bar{Q}(x, y) \supset \text{Bew}(\text{Neg}(\text{Sb}(q \frac{17}{Z(x)} \frac{19}{Z(y)})))$	5, 8
10	$\frac{x\text{B}(\text{Sb}(y \frac{19}{Z(y)})) \supset \text{Bew}(\text{Sb}(q \frac{17}{Z(x)} \frac{19}{Z(y)}))}{x\text{B}(\text{Sb}(y \frac{19}{Z(y)})) \supset \text{Bew}(\text{Neg}(\text{Sb}(q \frac{17}{Z(x)} \frac{19}{Z(y)})))}$	2, 10
11	$y = 17\text{Gen}q$	hyp
12	$\frac{x\text{B}(17\text{Gen}r) \supset \text{Bew}(\text{Sb}(r \frac{17}{Z(x)}))}{x\text{B}(17\text{Gen}r) \supset \text{Bew}(\text{Neg}(\text{Sb}(r \frac{17}{Z(x)})))}$	10, 11
13		10, 11

14	Bew(17Genr)	hyp
15	$(\exists n) nB(17Genr)$	6, 14
16	$(\exists n) Bew(Neg(Sb(r \frac{17}{Z(n)})))$	13, 15
17	$(n) Bew(Sb(r \frac{17}{Z(n)}))$	7, 14
18	$(\exists n) Bew(Neg(Sb(r \frac{17}{Z(n)})))$ $\cdot (n) Bew(Sb(r \frac{17}{Z(n)}))$	16, 17
19	$(\exists a)((n) Bew(Sb(a \frac{17}{Z(n)})))$ $(\exists n) Bew(Neg(Sb(a \frac{17}{Z(n)})))$	18
20	$Bew(17Genr) \supset (\exists a)((n) Bew(Sb(a \frac{17}{Z(n)})))$ $\cdot (\exists n) Bew(Neg(Sb(a \frac{17}{Z(n)})))$	14-19
21	$\overline{Bew(17Genr)}$	1, 20
22	$(n) \overline{nB(17Genr)}$	6, 21
23	$(n) Bew(Sb(r \frac{17}{Z(n)}))$	12, 22
24	$Bew(Neg(17Genr))$	hyp
25	$(n) Bew(Sb(r \frac{17}{Z(n)})) \cdot Bew(Neg(17Genr))$	23, 24
26	$(\exists a)((n) Bew(Sb(a \frac{17}{Z(n)})))$ $Bew(Neg(17Gena))$	25
27	$Bew(Neg(17Genr)) \supset (\exists a)((n) Bew(Sb(a \frac{17}{Z(n)})))$ $\cdot Bew(Neg(17Gena))$	24-26
28	$\overline{Bew(Neg(17Genr))}$	1, 27
29	$\overline{Bew(17Genr)} \cdot \overline{Bew(Neg(17Genr))}$	21, 28

This demonstration shows that there exists a P-formula, specifically that P-formula whose Gödel number is 17Genr, which is not P-provable and whose negation is not P-provable. It does not show, however, as the first demonstration did, why it is that the undecidable formula is true. In order to show that this is the case we have to introduce into this proof the apparatus required to talk about the interpretations of the P-formulae named by the numbers occurring in the proof. Since this process is quite complex and since the pattern of proof required to show that 17Genr is true is identical to the one pursued in the preceding version it will not be reproduced here. There is no doubt, though, that the undecidable formula is true. Of this Gödel himself was aware.

§4 Version III

The following demonstration is a deductivized version of the proof Mostowski gives of his Theorem I in Chapter VI of his book Sentences Undecidable in Formalized Arithmetic.⁴ This theorem states, that 'If K is a closed consistent set of matrices then the set $G \setminus K$ is not K-definable.'⁵ While again, as with Gödel, I shall not suppose my remarks to be intelligible independently of the original I shall make some effort to explain the central terms.

⁴Andrej Mostowski, Sentences Undecidable in Formalized Arithmetic (Amsterdam: North-Holland Publishing Company, 1957), pp. 88-89. Due to the impossibility of reproducing Gothic type here I use the Roman equivalents where Mostowski uses Gothic capitals. As I indicate the meanings of all letters as I introduce them no confusion should arise.

⁵Mostowski, p. 88.

Let me first explain the content of Theorem I. 'Matrix' is the term Mostowski uses to refer to arithmetized sentential functions and the sentences that arise when the free variables in such functions are bound or replaced by constants. G is the set of all sentences of (S) (' (S) ' is the name of the formal language Mostowski considers. It is simpler than Gödel's P for it possesses only variables of the first order.). A set of matrices K is called 'closed' if $K = T_K$, where T_K is the set of matrices which are K -provable in Gödel's sense. (Here we shall consider cases in which K is, at least potentially, a non-empty class.) A relation R is said to be K -definable if there is a matrix ϕ in which exactly the k free variables $a_1, \dots, a_k$ occur such that for arbitrary integers $n_1, \dots, n_k$ the following two conditions are satisfied: If $R(n_1, \dots, n_k)$ then $\phi(D_{n_1}, \dots, D_{n_k})$ is in K and if $\bar{R}(n_1, \dots, n_k)$ then $\neg \phi(D_{n_1}, \dots, D_{n_k})$ is in K . (' D_n ' denotes the numeral associated through the arithmetization scheme with the sign of (S) standardly interpreted as the n -th natural number.) Membership in a set may be considered a unary relation and hence it is also possible to speak of sets as K -definable. Thus, what Theorem I says is that any set K of matrices which is K -provable will either contain sentences for which it cannot be K -proven that they satisfy any matrix in (S) chosen to express the relation of being a K -provable sentence or will not contain sentences for which it cannot be K -proven that they do not satisfy any matrix in (S) chosen to express the relation of being a K -provable sentence.

This explanation is perhaps somewhat confusing and does not immediately show how Theorem I bears on undecidability. For this reason let me amplify this discussion of Theorem I with particular reference to undecidability. Let $K = T$ where T is the set of all provable matrices of (S) . Since T is closed and consistent it follows, according to Theorem I, that the set of provable sentences of (S) is not T -definable. That is, for any matrix which we might think expresses the unary relation 'is provable' one of the following will be the case: either 1) if this relation does in fact hold of certain numbers then the sentence obtained by replacing the free variables in ϕ by the numerals associated with the numbers of which the relation holds will not be in T or 2) if this relation does not hold of certain numbers then the denial of the sentence obtained by replacing the free variables in ϕ by the numerals associated with the numbers of which the relation does not hold will not be in T . Mostowski puts it more concisely thus: 'either there are integers in T for which we cannot prove in (S) that they satisfy the matrix (ϕ) , or there are integers outside T for which we cannot prove in (S) that they do not satisfy the matrix.'⁶ In either case we must infer that there are sentences which are intuitively true but unprovable in (S) . (This formulation is perhaps a bit misleading for it is not completely accurate to say that there are integers in T for which we cannot prove in (S) that they satisfy the matrix. To be completely precise we should say that there are integers in T for which we cannot prove in (S)

⁶Mostowski, p. 90.

that the numerals associated with these numbers satisfy the matrix.)

Mostowski goes on, after proving Theorem I, to prove more explicitly that there are sentences θ of (S) such that neither θ nor $\neg\theta$ is in T and he identifies some such sentences.⁷ But, since Theorem I figures as the crucial element in each of these further demonstrations, we shall content ourselves here with the intuitive explanation given above about how undecidability follows from Theorem I and instead turn to presenting a deductivized version of the proof of Theorem I which, as we have said, essentially contains the undecidability result.

Before we can proceed with this demonstration a number of preliminaries must be carried out. First, a word should be said about the frequent passage from talk about sets to talk about relations and vice versa. Mostowski, it seems to me, is not always careful in distinguishing the two and, though no serious consequences result, this does lead to a certain amount of confusion. All the results concerning K -definability are proved of relations and functions yet we find him talking, without acknowledging the fact, of sets being K -definable. In his first chapter he has introduced the notation ' $\lambda xyz \dots (\dots x \dots y \dots z \dots)$ ' to denote the relation holding between the integers x, y, z if and only if they satisfy the condition $\dots x \dots y \dots z \dots$ stated in the brackets;⁸ yet in Chapter VI we find him talking about the set $\lambda n (\dots n \dots)$.⁹

⁷ Mostowski, pp. 90-96.

⁸ Mostowski, p. 18:

⁹ Mostowski, p. 88.

In order to avoid confusions of this sort I shall adopt the following rules. Sets will be introduced through letters which are designated as denoting sets. 'K-definability' will be applied only to relations and functions. Where H , say, is a set we shall replace the expression ' H is K-definable' with what I suppose was intended, viz., ' $\lambda n(n \in H)$ is K-definable'. Expressions employing the relational abstract notation -- ' $\lambda n(\dots n \dots)$ ' -- will be referred to only as relations and never as sets. (I shall extend Mostowski's use of ' λ ' and use it as was done in the preceding proof. Where no confusion will arise concerning what the arguments are in a given function or relation, ' $\lambda n \dots n_k R(n_1, \dots, n_k)$ ', for instance, will be abbreviated to ' $\lambda R(n_1, \dots, n_k)$ '.

As a second preliminary to our presentation of Theorem I we must introduce the formalism necessary to present it in a deductive form. The following symbolization has been adopted:

1. K is a set of matrices; G is the set of all sentences of (S) ; H is a set.
2. θ is a sentence; ϕ is a matrix -- $\phi(a_1, \dots, a_n)$ is a matrix with n free variables.
3. R is a relation -- $R(n_1, \dots, n_k)$ is a relation with k arguments; F is a function -- $F(n_1, \dots, n_m)$ is a function with m arguments.
4. r_k is the predicate 'is K-definable'.
5. S is the substitution function where $S(i, p, b)$ is the result of substituting p for the free variable a_i through-

out b.

6. N_k is a k-tuple of integers.

Thirdly, as with the version presented in the preceding section I shall briefly characterize the premises involved. The first is a condition holding on the proof, the remaining six are theorems which either have been proven by Mostowski or can be easily proven. The first premise states that K is a closed and consistent set of matrices. The second claims that if the relation λR is K -definable then so is the relation $\lambda \bar{R}$. The third claims that for all integers and all sets an integer is a member of the complement of a set if and only if it is not a member of the set itself. The fourth claims that the function S is K -definable. The fifth claims that if λR is a K -definable relation with k arguments and λF is a K -definable function with m arguments then the relation $\lambda R(n_1, \dots, n_{j-1}, F(m_1, \dots, m_m), n_{j+1}, \dots, n_k)$ is K -definable. The sixth is simply the formal definition of K -definability. The seventh claims that if ϕ is a matrix with one free variable then the result of substituting the numeral associated with n for the free variable in ϕ is a sentence and it also makes the notational remark that $\phi(D_n)$ is equivalent to $S(1, D_n, \phi)$.

We are now in a position to prove Theorem I which states, as will be recalled, that if K is a closed, consistent set of matrices then the set $G \wedge K$ is not K -definable (the relation $\lambda \theta (\theta \in G \wedge K)$ is not K -definable).

1	$K = T_K \cdot (0) (0 \in T_K \vee (0 \rightarrow 1 \approx 1+1) \bar{\in} T_K)$	hyp
2	$r_K(\lambda n(n \in H)) \supset r_K(\lambda n(n \in -H))$	hyp
3	$(n) (H) (n \in -H \exists n \bar{\in} H)$	hyp
4	$r_K(\lambda S)$	hyp
5	$r_K(\lambda R(n_1, \dots, n_k)) \cdot r_K(\lambda F(m_1, \dots, m_m))$ $\cdot r_K(\lambda R(n_1, \dots, n_{j-1}, F(m_1, \dots, m_m), n_{j+1}, \dots, n_k))$	hyp
6	$r_K(\lambda R) \supset (\exists \phi(a_1, \dots, a_k)) (N_k) ((R(n_1, \dots, n_k)$ $\supset \phi(D_{n_1}, \dots, D_{n_k}) \in K) \cdot (\bar{R}(n_1, \dots, n_k)$ $\supset - (D_{n_1}, \dots, D_{n_k}) \in K)$	hyp
7	$(\phi(a_1)) (S(1, D_n, \phi) \in G)$ $\cdot (\phi(a_1)) (\phi(D_n) \in S(1, D_n, \phi))$	hyp
8	$r_K(\lambda \theta(\theta \in G \wedge K))$	hyp
9	$r_K(\lambda \theta(\theta \in -G \wedge K))$	2, 8
10	$r_K(\lambda \theta(S(1, D_n, n) \in -G \wedge K))$	4, 9
11	$r_K(\lambda \theta(S(1, D_n, n) \bar{\in} G \wedge K))$	3, 10
12	$(\exists \phi(a_1)) ((S(1, D_n, n) \bar{\in} G \wedge K \supset \phi(D_n) \in K)$ $\cdot (S(1, D_n, n) \in G \wedge K \supset - \phi(D_n) \in K))$	6, 11
13	$(\exists \phi(a_1)) ((S(1, D_n, n) \bar{\in} G \wedge K \supset S(1, D_n, \phi) \in G \wedge K)$ $\cdot (S(1, D_n, n) \in G \wedge K \supset -S(1, D_n, \phi) \in G \wedge K))$	7, 12
14	$n = \phi$	hyp
15	$S(1, D_n, n) \bar{\in} G \wedge K \supset S(1, D_n, n) \in G \wedge K$	13, 14
16	$S(1, D_n, n) \in G \wedge K$	15

The differences between versions II and III are subtler and more difficult to explain briefly. Gödel argues from the property of recursive relations expressed in premise 5 of version II and from the fact that the relation 'x is a proof of y' is a recursive relation. Mostowski, on the other hand, argues from the fact that the unary relation 'is provable' is not recursive, as he had defined recursivity, and shows that undecidability is an almost immediate consequence of this fact. In the section following his proof of Theorem I Mostowski claims that the second of the three proofs he gives, based on Theorem I, for the existence of undecidable sentences is 'approximately the original proof of Gödel'.¹⁰ Since it is not immediately obvious how this is so and since we have not presented the proof in question here anyway (rather 'only' a proof of Theorem I on which the latter proof depends) we shall rest content with noting the general procedural difference mentioned above and not pursue our comparison at greater length. One final point about a similarity the two proofs exhibit should, however, be made: both rely crucially on the fact that the relation 'x is a proof of y' is recursive while the relation 'x is provable' is not.

¹⁰Mostowski, p. 92, n. 5.

CHAPTER II

GÖDEL'S RESULT AND THE COMPLETENESS OF ARITHMETIC

The significance which is generally claimed for Gödel's result is that it shows elementary arithmetic to be incomplete.¹ For, since P contains undecidable formulae and can be interpreted as elementary arithmetic it follows that if P is the formal language adopted to express arithmetic then arithmetic is incomplete. Notice, however, that while P is sufficient to express elementary arithmetic it is not necessary, i.e. it is not the only language capable of doing so. That other languages could perhaps serve to express elementary arithmetic immediately raises the question of whether each of these other languages will contain undecidable formulae and hence whether any language for arithmetic could be complete. If one wishes to call into question the claim that Gödel's result implies the incompleteness of arithmetic an examination of alternative languages capable of expressing arithmetic will clearly be one approach. In this chapter I should like to investigate this alternative object language approach and to this end I shall examine two arguments which might be thought to cast doubt on the claim that Gödel's argument shows elementary arithmetic to be incomplete. After analysis it will become clear that both these arguments are inadequate to reject this claim.

¹Ernest Nagel and James R. Newman, Gödel's Proof (London: Routledge & Kegan Paul Ltd., 1959), p. 6 and pp. 98-9; and, Jean van Heijenoort, 'Gödel's Theorem', The Encyclopedia of Philosophy, 1967, Vol. 3, p. 348.

§1 The Truth of the Undecidable Formula

The first argument which might be thought to cast doubt on inferring the incompleteness of arithmetic from the incompleteness of P can be put as follows: while the undecidable formula whose existence is demonstrated by Gödel (let us again call it X) is not P -provable it is yet true. The force of 'true' here appears to be this: though X cannot be decided in P it can be decided in a metalanguage rich enough to express the notions introduced in the premises of the versions of Gödel's proof given in the preceding chapter plus, in the case of the second and third versions, certain additional notions (specifically, the notion of interpretation); hence, ' X is true' means 'a proof of X may be provided in a suitable metalanguage'. Now, given the fact that the undecidable formula of P is decidable in a suitable metalanguage and given the fact that P is only a sufficient and not a necessary language for expressing arithmetic, it seems that all that has to be done to avoid the problem of the incompleteness of arithmetic is to discover a language capable of expressing arithmetic which is also a suitable language for deciding the undecidable formula of P . This should be easily done. For the language in which the undecidable formula of P becomes decidable is necessarily more comprehensive than P and consequently will be capable of expressing at least as much as P . Thus, if P is amended to P' by the addition of the requisite notions, notions present already in meta- P , then

formulae like X would become decidable, and while this would not establish the completeness of arithmetic it would appear to invalidate the inference from P 's being incomplete to the incompleteness of arithmetic. In conclusion, then, it might be argued that while Gödel's result is true enough for P it is not true for all languages capable of expressing arithmetic and that in consequence we must simply be careful in choosing a suitable language if we wish to avoid incompleteness.

This argument, however, is inadequate. To it a proponent of the claim that Gödel's result shows arithmetic to be incomplete may reply as follows: let us grant that X is decidable in an appropriate meta- P . What about this language, though? Is it complete? Perhaps Gödel's result can be reproduced in this amended language so that for any formula X undecidable in P but decidable in P' there will be a formula X' of P' which is undecidable in P' but perhaps decidable in P'' . If so, while any particular undecidable formula might be decidable in an appropriate formalized metalanguage, no language capable of expressing arithmetic would be constructible which did not contain a formula undecidable in that language.

The claim that any amending (in the sense of consistent supplementation) of P will result in a language in which undecidable formulae can again be shown to exist -- I shall call the extendability thesis. This thesis is not discussed at length by Gödel but is by Mostowski in his exposition of Gödel's theory.² Let us now consider Mostowski's argument for this

²Mostowski, pp. 107-112.

thesis. My account of his argument will be qualitative but I do not think it loses much for that. The introduction of all the symbolism required to formalize the argument would be too large a task to undertake here and, furthermore, Mostowski himself has only outlined the necessary formalism.

In order to present Mostowski's argument for the extendability thesis we must first elucidate the important distinction between the two classes of matrices T and Tr , defined by Mostowski in the course of his exposition. T is the class of all provable matrices of (S) ; Tr includes the class of all true sentences of (S) and also certain matrices which are not sentences and which have no special designation. (In terms of these classes the undecidability result may be put thus: there exists a sentence X of (S) such that $X \in T$ and $\bar{X} \in T$ and $X \in Tr$.)

Let us first consider how T is defined.³ Put concisely, to say that a matrix ϕ is provable is to say that a formal proof of ϕ exists where to say that g is a formal proof of ϕ is to say 1) that every line of g either is an axiom or follows by modus ponens from two preceding lines or is the result of substituting a numerical expression for a free variable in a preceding line and 2) that the last line of g is ϕ . It is clear that T can be defined strictly by reference to notions, and consequently via the symbols, found in (S) . That it can be so defined is usually expressed by saying that T is a syntactic notion.

³Mostowski, pp. 40-41.

Now consider Tr. In defining Tr Mostowski adopts Tarski's idea of defining truth in terms of the satisfaction of a given sentential function by given objects.⁴ The sense attached to the notion of satisfaction can be illustrated by this scheme: 'for every a we have a satisfies the sentential function 'x is white' if and only if a is white'.⁵ If a sentential function or matrix has n free variables both Tarski and Mostowski find it convenient to speak of infinite sequences either satisfying or failing to satisfy the sentential function where the n-th term of the sequence is the value assigned to the n-th variable wherever it occurs in the function. If a sentential function is a sentence it will, in (S), be of one of two forms or a combination of these two forms: 1) it will be an expression all the terms of which are constants, numbers in (S), and containing the equality sign as the only relation sign. An example of such a sentence is '(1 sum 1) mult 1 equ 1 sum 1'. 2) it will be an expression containing variables all of which are bound. Clearly in the case of sentences we cannot talk about their being satisfied by infinite sequences in quite the same way as we can about sentential functions containing free variables. These differences, however, do not require that we cease talking about satisfaction by infinite sequences when dealing with sentences.

⁴Alfred Tarski, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, trans. J.H. Woodger, (Oxford: Clarendon Press, 1956), pp. 152-278.

⁵Tarski, p. 190.

Rather the definition of satisfaction can be stated in a fashion which takes sentences into account. The definition is so constructed that 1) an infinite sequence, it turns out, satisfies a numerical equivalence if and only if the equivalence holds. Obviously the individual sequence has no bearing on whether the equivalence holds, so if it does then every infinite sequence will satisfy it. If it does not then no sequence will.

2) an expression containing only bound variables is said to be satisfied by an infinite sequence f if and only if every sequence differing from f in at most those places corresponding to the variables bound in the expression satisfies the function obtained from the expression by dropping its quantifiers.

For example, a sequence f satisfies the function $z \equiv (x_k)y$ if and only if every sequence differing from f in at most the k -th place satisfies y . Again, as with numerical equivalences, it becomes clear that if one sequence satisfies such an expression as, for example, z then every one will. For, supposing in our example that x_k is the only variable in y and that f satisfies z then since f and every sequence differing from it in at most the k -th place satisfy y then every other sequence will also satisfy it, for further differences in other than the k -th place will be irrelevant to the sequence's satisfaction of y .

Having clarified the notion of satisfaction we may now state Mostowski's definition of Tr : the set Tr is the set of sentential functions all of which are satisfied by every

infinite sequence of integers.⁶ Since Tr is defined by reference to the notion of satisfaction its definition contains talk about arbitrary sets of integers (the infinite sequences of values which the variables take on). As a result it cannot be defined solely in terms of the notions, and so through the symbols, found in (S) . This fact is usually expressed by saying that Tr is a semantic notion. The most significant consequence of Tr 's being a semantic notion is that if it is to be formally defined this can only be done in a language richer than (S) , i.e. in a language extended in the appropriate fashion.

We set out to present Mostowski's argument for the extendability thesis. Having distinguished the sets T and Tr we are now in a position to talk about it more specifically. Gödel's result, as we mentioned, can be expressed by saying that there exists a sentence X of (S) such that $X \in T$ and $\bar{X} \in T$ and $X \in Tr$. Since Tr cannot be formalized in (S) there is no danger of believing we have both demonstrated and not demonstrated X when we say that $X \in T$ and $X \in Tr$. Rather we have, as Mostowski says, just demonstrated the insufficiency of proof procedure in (S) . If we extend (S) in the appropriate way (by adding variables for sets of integers, and the corresponding quantificational apparatus, and the operation 'is a member of') we arrive at a new language (S_1) which contains sufficient formalism to define the class Tr . The sentence X , undecidable

⁶Mostowski, p. 63.

in (S), is now easily decided in (S₁) by a formal demonstration of its membership in Tr.

Lest it be felt, however, that the language (S₁) is a language capable of expressing arithmetic in which undecidable sentences cannot be shown to occur we must carefully scrutinize the sentence X. The standard interpretation of X, the undecidable sentence which Mostowski employs (Gödel's undecidable sentence is slightly different), is this: 'there exists no formal proof for the sentence m',⁷ where m is a determinate sentence of the language (S). What is important to recognize here is that the unprovable sentence itself contains reference to provability; and, of course, the provability referred to here is provability in (S). Now, in the system (S₁) there will be a new class T₁ which will take over the role T had in (S) as the syntactically defined class of all provable sentences. It should be possible, then, to construct in (S₁) a sentence X' identical to X except that any reference to T, i.e. to being provable in (S), must be changed to a reference to T₁, i.e. to being provable in (S₁). This done, there appears, prima facie, to be no reason why a proof of the undecidability of X' in (S₁), entirely analogous to the proof of the undecidability of X in (S), could not be produced. Further, it seems clear that any attempt to remedy this situation by appealing to a new notion of truth whose definition is formalizable in a new system (S₂) will be confuted by yet another new undecidable sentence X'', constructed on analogy

⁷Mostowski, p. 107.

with X and X' . This, roughly, is Mostowski's argument for the extendability thesis.

Before leaving this discussion of the extendability thesis a few loose ends in Mostowski's account might be pointed out both because they suggest interesting questions and because failure to point them out could lead to perplexity. First, Mostowski never indicates the relation between Tr and T_1 : are they equivalent? is one a subset of the other? do they only intersect? Second, what form will Tr_1 take? It is not clear whether Mostowski feels that the new variables in (S_1) are just to be sets of integers or whether they could also be sets of sets. Nothing he says seems to rule out this latter interpretation. Finally, Mostowski makes the interesting claim that 'a system (S_ν) in which all sentences of (S) would be decidable if exists at all (sic) must correspond to an enormous transfinite ordinal ν '.⁸ This seems to suggest either 1) that he believes there are undecidable sentences in (S) other than those which can be decided in (S_1) which will require a vastly amended system in order to be decided (the meaning of the subscript ν is never precisely specified) or else 2) that he believes that only in a vastly amended formal system will one be able to produce a completeness proof, as opposed to simply being unable to produce undecidable propositions. Neither of these claims is explained or supported.

At any rate the earlier argument seems undeniable: so long as the undecidable sentence itself involves the notion

⁸Mostowski, p. 112.

of being provable in the language in which it is expressed and so long as the undecidable sentence can only be decided by appealing to notions not contained in the formal language in which it is expressed then it appears that we are forced to admit that no formal language can be constructed in which undecidable sentences will not occur.

§2 The Definition of the Concept of Truth

The second argument which might be thought to cast doubt on inferring the incompleteness of arithmetic from that of P is not so much an outright counter-argument as it is simply a doubt about certain potential consequences of amending P, consequences other than those discussed in the preceding section. There we considered and rejected the objection that since the undecidable formula was true and hence decidable in a formal language capable of expressing the notion of true in P, there perhaps existed a language P' in which no undecidable formulae occurred. Here I should like to consider the possibility that in some amended formal language P* a definition of truth for P* could be expressed in R*. If this were possible then since, according to the earlier argument, every language which is the result of amending P will contain undecidable formulae it might be the case that we could show in P* that an undecidable formula of P* was true. But to do this would be to establish an antinomy: for on the one hand we would have shown, through an argument in meta-P*, that a formula of P*,

say X , was undecidable while on the other hand we would have shown in P^* that X was true." If this conflict between syntactic and semantic approaches could arise, if there could be a formula which was not provable in a language but was demonstrably true in that language, then serious doubt would be thrown on the whole conventional approach to metalanguage and in particular on Gödel's results. And why should it not be the case that in some formal language, of transfinite order if necessary, a definition of truth for the language is constructible in the language itself?

The question of constructing a definition of truth for a given language in that language itself is directly addressed by Tarski, though not in the present context of debate. His answer to this question forms part of his answer to the larger question about the possibility of giving a definition of truth for formalized languages in general. Answering this larger question is his project in his long article 'The Concept of Truth in Formalized Languages'.⁹ I shall now give a general description of this project in order to show how Tarski's argument excludes the possibility, outlined above, of syntactic and semantic notions coming into conflict in some formal language.

Tarski rejects the possibility of constructing a precise definition of truth for natural language, presenting arguments to show that a definition of truth in natural language leads unavoidably to the so-called semantical antinomies

⁹Tarski, pp. 152-278.

such as the Liar paradox. However, he feels that a precise definition of truth is possible for formal languages provided certain conditions are met.

The first condition (condition I) specifies the criterion according to which a definition of truth shall be called an adequate definition of truth. This criterion is expressed by what Tarski calls convention T:

Convention T. A Formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called an adequate definition of truth if it has the following consequences:

- (α) all sentences which are obtained from the expression 'x_εTr 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;
- (β) the sentence 'for any x, if x_εTr then x_εS' (in other words 'Tr ⊆ S').¹⁰

(Two explanatory comments should be made here: 1) 'S' denotes the class of all sentences of the language in question. 2) Tarski's conception of the metalanguage includes perhaps more than the intuitive conception of the metalanguage we have been relying on so far does. Tarski conceives the metalanguage as not only possessing names for all the signs of the language and the logical apparatus necessary to talk about the language; he also conceives it as containing an interpretation -- what is here spoken of as the 'translation' -- of every sentence in the language.)

In order to specify the further condition that must be met for a precise, formal definition of truth to be possible

¹⁰Tarski, pp. 187-188.

a characteristic of formal languages according to which they may be classified must be introduced and explained. Formal languages may be distinguished according to their order. The order of a language is equivalent to 'the smallest ordinal number which exceeds the orders of all variables occurring in this language'.¹¹ A variable is a sign of the language and the order of a sign is defined as 0 in the case that the sign is the name of an individual or a variable representing an individual and n in the case that the order of all the arguments in all the sentential functions in which the sign occurs as a sentence-forming functor is less than n and the order of at least one of them is exactly $n-1$. A sign is said to be of infinite order if it is 'a sentence-forming functor of only those sentential functions which have all their arguments of finite order, where, however, these orders are not bounded above by any natural number'.¹² The order ω is assigned '(t) o those signs of infinite order which are functors of sentential functions containing exclusively arguments of finite order ... A sign which is a functor in only those sentential functions in which the arguments are either of finite order or of order ω (and in which at least one argument of a function is actually of order ω), is of the order $\omega+1$ '.¹³

With the notion of order of a language clarified we can now state the second condition (condition II) which a lan-

¹¹Tarski, p. 270.

¹²Tarski, p. 269.

¹³Tarski, p. 270.

guage must meet in order for a formal definition of truth to be possible. This condition is, for us, the major result at which Tarski arrives in his paper. We state it as he has revised it in the postscript to his article:

- A. For every formalized language a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage with the help only of general logical expressions, of expressions of the language itself, and of terms from the morphology of language--but under the condition that the metalanguage possesses a higher order than the language which is the object of investigation.
- B. If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed.¹⁴

For a formal definition of truth to be possible, therefore, the condition that the metalanguage in which this definition is constructed be of a higher order than the language for which it is constructed is both necessary and sufficient. The claim that this condition is sufficient is of no special interest; rather, we are interested in the claim that it is necessary. What is Tarski's argument for this claim?

Actually, due to the positions adopted in the original article, as opposed to the postscript, Tarski does not argue directly for this claim. However, he does argue for a related claim and leads us to believe, in his postscript, that the earlier argument can be modified to support the claim we are interested in here. I shall attempt to perform this modification by altering the crucial theorem in the relevant respects and presenting a revised proof for it.

¹⁴Tarski, p. 273.

Theorem I': In any formalized language if the metatheory is of an equivalent or lesser order than the language under consideration then (α) in whatever way the symbol 'Tr', denoting a class of expressions, is defined in the metatheory, it will be possible to derive from it the negation of one of the sentences which were described in the condition (α) of the convention T; (β) assuming that the class of all provable sentences of the metatheory is consistent, it is impossible to construct an adequate definition of truth in the sense of convention T on the basis of the metatheory.¹⁵

Theorem I' effectively states the relation between the two conditions Tarski has imposed on formal languages for the purpose of defining the concept of truth: if condition II does not hold condition I will be violated. Therefore, since condition I is necessary condition II must hold.

Self-confessedly, Tarski's 'proof' of Theorem I is 'only a sketch of the proof ... one which is far from complete'. This sketchiness is such that I must admit to not being able to fill in all the gaps. Nevertheless, I feel that there is some value, aside from the immediate context (see Chapter IV), in presenting the proof: perhaps my keener-sighted readers will find no problem in supplying the missing steps in the argument. I shall do two things here. First, after the fashion of Tarski, I shall run through the argument in a rough, expository manner. Second, after the fashion of Chapter I, I shall attempt to present the argument in a deductive form.

As a prelude to the exposition a few introductory remarks serving to specify the differences between our proof and Tarski's and to forewarn the reader of some of the larger

¹⁵Cf. Tarski, p. 247. The proof follows on pp. 247-251.

gaps in the argument must be made: 1) In proving Theorem I Tarski was concerned only with establishing the result for one formalized language, the language of the general theory of classes. What he says later, however, commits him to Theorem I' which makes a claim about all formalized languages. We shall not attempt here to prove the result for all languages; rather, we shall prove the result for only one language -- P^0 , a language of order 0 -- and shall assume that a similar proof can be provided for any other formal language. 2) We are interested only in languages capable of expressing arithmetic; numbers, therefore, will be the basic elements in the languages we consider. Numbers have been treated in different manners: Russell and Tarski treat numbers as classes of classes of individuals -- objects, according to Tarski of the second order. Gödel and Mostowski treat numbers as simple individuals, i.e. as objects of the 0-th order. Since we have chosen to consider a language of order 0 we have no option but to treat them in this latter way. This marks a second divergence between our proof and the sketch Tarski provides. 3) In the proof it is necessary to assume that P^0 is capable of expressing any arithmetic function which is an operation on natural numbers corresponding to an operation on expressions of P^0 via a unique association of natural numbers and expressions of P^0 . The consequence of this feature of P^0 is that P^0 is capable of expressing an arithmetized meta- P^0 provided that this meta- P^0 is of equal or lower order than P^0 . Tarski was working with the language of the general theory of classes

and he simply assumed that it sufficed 'for the foundation of the arithmetic of the natural numbers'¹⁶ and consequently that it was capable of expressing the sort of arithmetic functions specified above. Can we be sure, though, that this vital assumption holds of P^0 ? Without proving this to be so, and I repeat that Tarski never proves that his assumption holds of the language of the general theory of classes, let me cite the following considerations in support of the supposition that the assumption in question does hold of P^0 : a) Since Tarski is committed to the truth of Theorem I', and presumably would argue its truth on the grounds that an argument similar to the one he provided for the language of the general theory of classes could be repeated for a language of any order, he is committed to the position that the assumption we are considering holds of P^0 . b) If, as we have agreed, we let the natural numbers occur as individuals in P^0 and specify basic arithmetical axioms as axioms of P^0 then P^0 becomes capable of serving as a foundation for 'that part of number theory which does not make use of the notion of limit'.¹⁷ Granted, since at no point does Tarski ever specify precisely what portion of arithmetic is sufficient for expressing the sort of arithmetic functions in question, this last consideration may not be conclusive. But, since a rigorous proof of the sufficiency of P^0 for expressing arithmetized meta- P^0 functions would require an exhaustive specification of meta- P^0 and of all

¹⁶Tarski, p. 249.

¹⁷van Heijenoort, p. 350.

the possible forms a definition can take in meta- P^0 , we shall simply operate on the supposition that any expression defined in meta- P^0 will, provided the metalanguage is of equal or lower order than P^0 , embody only those formal properties which, if the expression were arithmetized, would require only the arithmetical devices found in P^0 . 4) The preceding remark noted one weakness in the demonstration. Another, perhaps more serious, weakness must now be pointed out. As Tarski is well aware the demonstration itself is carried out in what he calls the meta-metatheory.¹⁸ However, supposedly for the sake of simplicity, he expresses himself in places as though the demonstration belongs to the metatheory. Thus, 'in particular', he says, 'instead of saying that a given sentence is provable in the metatheory, we shall simply assert the sentence itself.'¹⁸ I strongly doubt whether this achieves the desired simplicity: it is one thing to prove, on the basis of the meta-metatheory, a given sentence of the metalanguage; it is quite another to prove, on the basis of the meta-metatheory, that the same sentence is provable in the metatheory. At a number of points throughout Tarski's argument, therefore, while it is apparent how a sentence of the metalanguage follows from truths in the meta-metatheory, it is not at all clear how these truths guarantee that the sentence is provable in the metalanguage. Thus, in the course of the argument a number of principles which are not even acknowledged, let alone demonstrated, need to be relied upon. The relevance of this remark will become clearer as the

¹⁸Tarski, p. 249.

exposition unfolds.

The idea of the proof is something like this: it is possible, as with Gödel numbering in P , to assign through a one-to-one correspondence a unique natural number to every expression of P^0 . It is also possible to define, in meta- P^0 , an infinite sequence of expressions, ϕ , in which every expression of P^0 will occur exactly once. As a result, it becomes possible to correlate with every class of expressions of P^0 a class of natural numbers. Consequently, since 'Tr', as it is defined in the metalanguage, denotes a class of expressions of P^0 , there will correspond to this class a class of natural numbers.

A function v_k is defined in P^0 which asserts that x is equivalent to the number k , i.e. $v_k \equiv x = f^k(0)$, where $f^k(0) = 0^k$ -- the k -th successor of 0. Consider the expression $\bigcup_1 (v_n \cdot \phi_n) \bar{E} \text{Tr}$. Since this is an expression of meta- P^0 there will be a corresponding expression in the natural numbers which we may call $\psi(n)$ such that we can assert:

for any n , $\bigcup_1 (v_n \cdot \phi_n) \bar{E} \text{Tr}$ iff $\psi(n)$ 1.

Now, since, as we have noted in remark three above, P^0 is assumed to be sufficient to express arithmetic functions of an arithmetized P^0 of order no greater than P^0 , and since ' $\psi(n)$ ' is an arithmetic function of such an arithmetized meta- P^0 , it follows that ' $\psi(n)$ ' will be an expression of P^0 . Thus, ' $\psi(n)$ ' will be a term of ϕ , say the k -th term. That is, ' $\psi(n)$ ' = ϕ_k . If we substitute ' k ' for ' n ' in sentence 1. we get:

$\bigcup_1 (v_k \cdot \phi_k) \bar{E} \text{Tr}$ iff $\psi(k)$ 2.

According to Tarski's convention T, if 'Tr' has been adequately defined then sentences of the form ' $x \in \text{Tr}$ iff p ' will be generated for all sentences in P^0 where ' x ' denotes a structural-descriptive or other name of a particular sentence and ' p ' denotes the statement expressed by this sentence or any equivalent statement. Now, since ' $U_1(v_k \cdot \phi_k)$ ' denotes a sentence of P^0 we may assume that if 'Tr' is adequately defined in meta- P^0 then the sentence generated by applying the above schema to it will be a consequence of the definition of 'Tr'. In the schema ' $x \in \text{Tr}$ iff p ' we replace ' x ' by ' $U_1(v_k \cdot \phi_k)$ ' and ' p ', considering the meaning of the statement $U_1(v_k \cdot \phi_k)$, by 'there is an x such that $x=k$ and $\psi(x)$ ', or equivalently, ' $\psi(k)$ '. This yields:

$$U_1(v_k \cdot \phi_k) \in \text{Tr} \text{ iff } \psi(k) \quad 3.$$

Sentences 2. and 3. are contradictory. Consequently it is established that the negation of one of the sentences mentioned in condition α of convention T must be among the consequences of any definition of 'Tr' in meta- P^0 .

Thus, the first part of Theorem I' is established. In order to make it clear how the second part follows we must recall the comments made in remark 4. above to the effect that sentences simply asserted were to be taken more fully as assertions that these sentences were provable sentences of the metatheory. In particular, sentences 2. and 3. are to be taken as provable sentences of the metatheory. Consequently, if we assume that the class of all provable sentences of the metatheory is consistent it follows that since the supposition

that an adequate definition of truth is possible in a meta-theory of equal or lower order than P^0 has the consequence that a sentence and its negation are provable in the meta-theory, then an adequate definition of truth must not be possible in a metatheory of equal or lesser order than P^0 .

The difficulty mentioned in remark 4. now becomes clearer: it is apparent that given a unique number mapping between every expression of P^0 and a natural number, operations on expressions of P^0 defined in meta- P^0 are equivalent to functions of natural numbers and hence that sentence 1. holds; it is not so apparent that it is provable in meta- P^0 that any given operation on expressions of P^0 defined in meta- P^0 is equivalent to a function of natural numbers, especially where meta- P^0 is of order equal to or lesser than P^0 . To establish sentence 1. one need only adopt the principle that to every sentential function of meta- P^0 , there corresponds a function of natural numbers which holds if and only if the sentential function holds; to establish that sentence 1. is meta- P^0 -provable, however, one must adopt the principle that to every sentential function of meta- P^0 there corresponds a function of natural numbers such that it is meta- P^0 provable that the first holds if and only if the second does. In the deductive version which follows I have adopted the stronger principles where it was necessary to do so in order to preserve the deductive nature of the presentation.

I shall now present a deductive version of what I take to be Tarski's argument:

1	The class of all provable sentences of M^0 is consistent	hyp
2	There exists an adequate definition of truth in M^0	hyp
3	Every expression of P^0 occurs exactly once in an infinite sequence ϕ of expressions defined in M^0	hyp
4	To every sentential function of M^0 there corresponds a function of natural numbers which holds if and only if the sentential function holds	hyp
5	Every function of natural numbers which corresponds to a sentential function of M^0 such that the sentential function holds if and only if the natural number function does occurs as an expression of P^0	hyp
6	' $\bigcup_1 (\bigvee_n \cdot \phi_n) \in \text{Tr}$ ' is a sentential function of M^0	2
7	$(n) (\bigcup_1 (\bigvee_n \cdot \phi_n) \in \text{Tr} \text{ iff } \psi(n) \text{ is } M^0\text{-provable})$	4, 5
8	' $\psi(n)$ ' is an expression of P^0	6, 7
9	$(\exists k) (' \psi(n) ' = \phi_k)$	3, 8
10	$(\bigcup_1 (\bigvee_k \cdot \phi_k) \in \text{Tr} \text{ iff } \psi(k))$ is M^0 -provable	7
11	$(\bigcup_1 (\bigvee_k \cdot \phi_k) \in \text{Tr} \text{ iff } \psi(k))$ is M^0 -provable	2, 9
12	The class of all provable sentences of M^0 is consistent	1, reit
13	10 . 11 . 12	10, 11, 12 conj int
14	If there exists an adequate definition of truth in M^0 then 10 . 11 . 12.	2-12, imp int
15	There does not exist an adequate definition of truth in M^0	14, Law of Non-Cont
16	If the class of all provable sentences of M^0 is consistent then there does not exist an adequate definition of truth in M^0	1-15 imp int

The demonstration we have presented shows only that no definition of 'Tr' in a meta- P^0 , assumed to be of equivalent or lesser order than P^0 , can meet the conditions of convention T. As we have noted, there appears to be no reason why a similar proof could not be constructed for languages of any given higher order and hence we must conclude that it is impossible to construct for a formal language of any order a definition of truth in a metalanguage of equivalent or lesser order. This completes our rough indication of a proof for Theorem I'.

As we noted earlier, Theorem I' shows how condition II must hold if condition I is to hold, i.e. if an adequate definition of truth is to be possible. Thus with this rough proof of Theorem I' we have indicated how Tarski believes that a definition of truth is only possible in a metalanguage of at least one order higher than the object language. If this is so then the doubt we entertained at the beginning of this section is shown to be unfounded: it could never be the case that a formula undecidable in a given language could, at the same time, be demonstrably true in that language for, since a definition of truth for the language in question is not possible in the language in question the undecidable formula could never be shown to be true in the language.

Thus, both the arguments examined in this chapter which were thought to call into question the possibility of

inferring the incompleteness of arithmetic from the incompleteness of P are shown, in fact, to be specious. Both these arguments, as we remarked out at the outset of this chapter, make use of the idea that since P is not a necessary language for the expression of arithmetic then perhaps there is an alternative language capable of expressing arithmetic in which Gödel's result cannot be reproduced. The first argument does this by suggesting that since a language can be constructed in which the undecidable formula of P is decidable then perhaps this new language is itself complete. The second does it by suggesting that a language capable of expressing arithmetic might be constructed in which the notion of truth, usually considered to be a semantic notion, could be expressed as a syntactic notion, with the result that the demonstration of the existence of a true but unprovable formula would become paradoxical. Of course the rebuttal of these arguments by those who think that Gödel's result does imply the incompleteness of arithmetic does not by any means constitute a rigorous proof that there is no formal language capable of expressing arithmetic in which Gödel's result cannot be reproduced. But it does eliminate the most obvious ways of attempting to argue that such a language might exist. In the face of this unpromising situation the idea suggests itself that if we wish to continue to doubt that Gödel's result implies the incompleteness of arithmetic then perhaps we should abandon the project of looking for an alternative language for arithmetic in which Gödel's result would not be reproducible, and see instead if our doubt can be pursued in some other manner.

CHAPTER III

CONTRADICTIONS AND ENTAILMENT:

AN ALTERNATIVE METALANGUAGE

In the last chapter we studied two arguments which questioned whether Gödel's result implied the incompleteness of arithmetic. Both of these arguments made use of the fact that Gödel's P was only a sufficient language for expressing arithmetic and not a necessary one. This fact suggested that perhaps there was an alternative language capable of expressing arithmetic in which Gödel's result would not be reproducible. Searching for alternative object languages was shown, however, to be an unpromising approach and, consequently, at the end of the last chapter, we raised the question of whether there might be some other way to call into question the claim that Gödel's result implies the incompleteness of arithmetic.

I think that there is another way to do this. Gödel's proof proves something about an object language P , but his proof is itself carried out in a metalanguage, a metalanguage which we shall call M . If we wish to call the implications of his result into question then not only may we question the necessity of using P to express arithmetic, we may also question the necessity of using M to investigate P . Is the metalanguage in which Gödel carries out his proof the only one possible and, if not, is it the one which most highly recommends

itself? Of course Gödel's presentation of his argument is an informal one: he does not exhaustively specify the ~~signs~~ and rules of M . Nevertheless, we can isolate certain features of M and examine the reasons for their having been adopted.

One feature of M which is not unique to it but is found in general use in formal languages is the employment of what I shall call material entailment and of the rules of inference standardly associated with this entailment relation. By material entailment I mean roughly (a more precise definition will be given later) the entailment relation defined by the rules of inference of any standard system of quantificational logic, for example, the one set forth by Fitch. That M employs material entailment may be inferred from a number of facts: 1) The entailment relation holding in P is material entailment. Given the tradition that Gödel was working in there is no reason to suppose that he believed that this entailment relation was anything less than adequate for the metalanguage as well. 2) As will be shown below Gödel's argument depends essentially on one feature of material entailment. Thus we can be sure that the entailment relation used in M possesses at least this feature of material entailment; this is a sufficient ground, given that there appear to be no other divergences, for assuming that the entailment relation employed in M is material entailment.

What I should like to do in this chapter is introduce a modified concept of entailment and suggest how the standard inference rules have to be modified in order to go along with

it. The language obtained from M by altering it to include these modifications we shall call M_N . In the next chapter we shall investigate the possibility of reproducing Gödel's argument in M_N . Here we shall concentrate on developing, independently of any reference to Gödel's project, a language based on a non-standard entailment relation.

As well as introducing an alternative metalanguage M_N I should also like to argue that it is to be preferred, as a metalanguage, to M . Thus I shall argue that there are good reasons for adopting M_N over M , i.e. that there are good reasons for favouring the new concept of entailment introduced here to the standard concept of material entailment. These reasons are of two sorts: 1) There is the evidence of intuition: the new concept of entailment is more closely in accord with what we intuitively take entailment to be. 2) There are desirable consequences which follow from adopting the new concept of entailment: a new and superior solution to the Liar paradox (and perhaps to other paradoxes as well) becomes possible. 1) is the more important justification, and indeed I believe it to be incontrovertible, for adopting the new concept of entailment. In order to introduce the new concept of entailment, therefore, I shall argue that the standardly accepted concept of material entailment is, with respect to our intuitive concept of entailment, deficient; then, I shall argue that by modifying the standard concept in the appropriate fashion we can arrive at an intuitively more satisfactory concept of entailment. (As a consequence of this approach

the content of this chapter stands as something of intrinsic interest and hopefully is of worth regardless of its implications for Gödel's argument.)

§1 Contradictions and Material Entailment

Our dissatisfaction with material entailment stems from the role contradictions are allowed to play in systems employing this entailment relation. It is generally held that in such systems every proposition is entailed by a contradiction, i.e., that from the conjunction of a proposition and its negation any further proposition may be validly deduced. This principle may be arrived at in various ways:

- 1) It may be taken as a primitive rule. Fitch, whose method of subordinate proofs we have been using and will continue to use, is one who introduces this principle as a primitive rule:

Rule of negation elimination ("neg elim"). Every proposition q is a direct consequence of the pair of propositions p and $\bar{p}$. Thus from a pair of contradictory propositions every proposition may be regarded as following. This rule may seem odd to the reader unfamiliar with symbolic logic. He would perhaps wish to contend that we should assume that nothing follows from contradictory results, instead of assuming that everything follows. This rule, however, is probably best interpreted as meaning, "anything is true if a contradiction is true".

- 2) It may be derivative, resting on rules of inference which are taken as primitive. The following demonstrations show how if either of the rules of modus tollendo ponens and modus

¹Fitch, p. 54.

tollens (i.e. modus tollendo tollens) is taken to be primitive² then any proposition q can be shown to follow from a contradiction $p \cdot \bar{p}$.

1	$p \cdot \bar{p}$	hyp
2	p	1, conj elim
3	$p \vee q$	2, dis int
4	$\bar{p}$	1, conj elim
5	q	3, 4, modus tollendo ponens

1	$p \cdot \bar{p}$	hypo
2	p	1, conj elim
3	$\bar{q}$	hyp
4	p	2, reit
5	$\bar{q} \supset p$	3-4, imp int
6	$\bar{p}$	1, conj elim
7	$\bar{q}$	5, 6, modus tollens
8	q	7, double neg elim

3) It may, even more fundamentally, be seen to be based on the commonly held truth-functional definitions of the connectives ' $\vee$ ', ' $\cdot$ ', ' $\supset$ ', and ' $\neg$ ' and on the notion of material entailment, according to which one proposition in a proof entails another provided the latter can never be true if the former is false. We show here, by reference to a simple truth table,

²One logician who takes the latter as well as the former of these rules to be primitive is Benson Mates in his Elementary Logic 2nd. ed., (New York: Oxford University Press, 1972), p. 97.

how a contradiction $p \cdot \bar{p}$ materially entails any arbitrary proposition q ; i.e., we show, given the truth-functional definitions of ' $\cdot$ ' and ' $\bar{}$ ', how for any assignment of truth values to the components of $p \cdot \bar{p}$ it will never be the case that $p \cdot \bar{p}$ is true and any arbitrary proposition q is false.

p	q	p	$\cdot$	$\bar{p}$	q
T	T	T	F	F	T
F	T	F	F	T	T
T	F	T	F	F	F
F	F	F	F	T	F

Clearly $p \cdot \bar{p}$ can never be true and consequently it can never be the case that $p \cdot \bar{p}$ is true while some proposition, of which it is wondered whether it is entailed by $p \cdot \bar{p}$, is false. Each of the steps in the two demonstrations in 2) could, instead of relying upon the intuitive obviousness of the principles of inference to which they appeal, have been shown to be valid material entailments by truth tables similar to the one above. Thus, it is in the truth functional definitions of the connectives and in the notion of material entailment that we find the ultimate source, I believe, of the idea that a contradiction entails every proposition.

Almost everyone, in studying logic, is struck by the counter-intuitiveness of the procedure which allows one to validly deduce anything from a contradiction. This counter-intuitiveness may be seen simply as one instance of the counter-intuitiveness of material entailment in general which, while it ensures that a false conclusion never follows from true

premises, seems to fail to capture the true sense of entailment which underlies intuitive argumentation in natural language. One proposition materially entails another if there is no possible assignment of truth values to the former such that the former is true and the latter false and yet it is obvious that there may be no relation of 'following from' or 'being a reason for' holding between these two propositions. This sort of problem with material entailment, and the similar problem of the material conditional, has preoccupied many modern logicians, most notably C.I. Lewis³ and Anderson and Belnap.⁴ I have no wish to enter into the extensive current debate concerning these logicians' ideas. All I want to remark here is that the general concern with the possible inadequacy of material entailment may be characterized as a preoccupation with the relevancy of premises to conclusions in entailment relations. That is, what logicians have objected to when objecting to material entailment in general is that it is possible for there to be true material entailments whose premises bear absolutely no contentual connection whatsoever with their conclusions; and this seems to violate our intuitive sense of entailment which insists that the premise be a reason for believing the conclusion, where for one proposition to be a reason for believing another it is necessary that there be some contentual connection between the two propositions.

³Clarence Irving Lewis, 'Strict Implication', in Logic and Philosophy, ed. Gary Iseminger (New York: Appleton-Century-Crofts), pp. 67-76.

⁴Alan Ross Anderson and Nuel D. Belnap, 'Entailment', in Logic and Philosophy, ed. Gary Iseminger (New York: Appleton-Century-Crofts), pp. 76-108.

Certainly it is plausible to object to a concept of entailment which allows one to validly deduce anything from a contradiction on the grounds that a contradiction may not have any relevance to the proposition which is deduced from it. However, I think that there is another sort of objection, heretofore not explored, which is simpler and more plausible -- enabling us to avoid the manifold problems involved in defining 'relevance' in an acceptable fashion. This other objection is made on the grounds of what I shall call, for lack of a less pedantic term, the ratioannihilation of contradictories.

I should like to suggest, following S. Talmage, that our intuitive notion of entailment ('natural' entailment) is closely related to our notion of conclusive reason, so that to say p entails q is to say that conclusive reason for believing p is, or would be, conclusive reason for believing q . Consider also the principle, which I take to be self-evident, that if any sort of reason is provided for believing a proposition p and if p materially entails any number of further propositions, say q and r and s , then the reason provided for believing p provides equal reason, supposing no further evidence to be considered, for believing q and for believing r and for believing s . If we accept these two premises then it follows that a contradiction ought to entail nothing for the simple reason that a contradiction can provide no more reason for believing any given proposition than it can for believing any other. Given any proposition q which is 'validly deduced' from a contradiction p , $\bar{p}$, $\bar{q}$ can be equally

'validly deduced' showing that conclusive reason for believing $p \cdot \bar{p}$ can provide no more reason for believing q than for believing its contrary. Hence, since reason for believing $\bar{q}$ is reason for disbelieving q , $p \cdot \bar{p}$ can provide no net or conclusive reason for believing q . Indeed, $p \cdot \bar{p}$ does not even provide conclusive reason for believing p or for believing $\bar{p}$ for it provides equally as much reason for believing either's contradictory and therefore, as much reason for disbelieving as for believing either one of them. We might say that the two terms of the contradiction $p \cdot \bar{p}$ cancel each other out in terms of their capability for providing conclusive reason for believing anything else (hence the term 'ratioannihilation').

The fundamental opposition of terms in a contradiction which makes them completely incapable of providing conclusive reason for believing any further proposition is, I believe, a sufficient ground for adopting a concept of entailment which will ensure that contradictions can entail nothing. In the following section I define such a concept, which I call 'natural entailment'. This definition of 'natural entailment' can be seen as a purely stipulative definition. However, since I believe that the consequences of this definition are necessary consequences of a definition of natural entailment, i.e. of a definition of what it is for conclusive reason for believing one proposition to be conclusive reason for believing another, and since they also appear to me to approximate the consequences required for such a definition to be a sufficient definition of natural entailment, I shall take the definition

to be a substantive one. By no means, though, should it be assumed that this definition precludes additional qualifications.

§2 Natural Entailment

In this section I shall introduce a number of definitions which, for the most part, are based on definitions first put forward by S. Talmage. Following this, I shall explain how the definition of natural entailment proposed here has the consequence of disallowing contradictions from entailing. While the definitions presented here are phrased so as to apply to formulae of quantificational logic they are also to be understood, in the appropriate context, as applying to propositions of a propositional calculus. This practice is not really unusual: the term 'conjunction', for instance, is frequently used in this ambiguous fashion to denote now a particular sort of purely formal arrangement of signs, now a certain combination of propositions. Context eliminates the possibility of confusion and the deliberate ambiguity saves us from having to introduce a whole series of parallel definitions.

We introduce the following five definitions:

p materially entails q iff p is a formula or group of formulae of quantificational logic and q follows from p in accordance with any of the sets of entailment rules that are standardly proposed for this logic -- conjunction, disjunction, negation and quantification rules. (There is no need for us to exhaustively

specify such a set of rules here. As we have been using the predicate calculus system developed by Fitch, we shall take his definitions of the implication rules as a standard.)

p naturally entails q iff (i) p materially entails q (ii) p does not materially entail $\bar{q}$

An exact contradiction is the conjunction of some formula and its negation

A contradiction is any formula that materially entails an exact contradiction

A contradiction generator is any formula or group of formulae that materially entails an exact contradiction (this definition makes it essential that p , in the above definitions, be understood as standing either for a formula or for a group of formulae)

It is evident that in a deductive system employing natural entailment as defined above no proposition will be naturally entailed by a contradiction. This is so because any contradiction p will materially entail any proposition $\bar{q}$ as well as q and hence will fail to meet the second condition for natural entailment. Moreover, no proposition or group of propositions will naturally entail a contradiction because anything that materially entails a contradiction will itself be a contradiction generator and consequently will naturally entail nothing. Thus, for instance, neither r nor $\bar{r}$ will be naturally entailed by the contradiction $r \cdot \bar{r}$. Nor will the contradiction $r \cdot \bar{r}$ be naturally entailed by the propositions r and $\bar{r}$.

It is clear, therefore, that the above definition of natural entailment has the desired consequence of disallowing contradictions from entailing. In the preceding section, however, when we discussed the motivation for disallowing contradictions from entailing we were concerned only with arguing that what we have come to call exact contradictions ought not to entail. It is a consequence of the above definitions that not only do exact contradictions fail to naturally entail but so as well do those contradictions which are not exact contradictions and those contradiction generators which are neither exact contradictions nor contradictions (from this it follows that nothing can naturally entail a contradiction). Do the same arguments which led us to hold that exact contradictions ought not to entail suggest as well that those contradictions and contradiction generators which are not exact contradictions ought not to entail? I think that they do: it will be recalled that we held that an exact contradiction $r \text{ . . . } \bar{r}$ ought not to entail any proposition and, in particular, that it ought not to entail r or to entail $\bar{r}$ on the grounds that 1) entailment ought to be understood through the notion of conclusive reason -- p should be said to entail q iff conclusive reason to believe p provides conclusive reason to believe q and 2) if any sort of reason is provided for believing a proposition and if this proposition materially entails any number of further propositions then the reason provided for believing the initial proposition provides equal reason, supposing no further evidence to be considered, for

believing each of the propositions it materially entails. From 1) and 2) it follows that exact contradictions ought not to entail because, to take, as an example, the particular case of the relation between the contradiction and its components, conclusive reason for believing $r \cdot \bar{r}$ does not provide conclusive reason for believing r or for believing $\bar{r}$ since it provides equal reason for believing either's negation and hence equal reason for disbelieving each of r and $\bar{r}$. One may argue in a similar fashion concerning contradictions and contradiction generators which are not exact contradictions: if one accepts 1) the connection made between the notion of entailment and that of conclusive reason and 2) the principle of equal distribution of reason for believing a proposition over the propositions which it materially entails then nothing which materially entails an exact contradiction ought to entail any proposition. For, since any proposition which materially entails an exact contradiction materially entails every proposition there will be equal reason for believing every proposition and its negation and hence equal reason for believing and disbelieving every proposition. Consequently, any proposition which materially entails an exact contradiction can provide conclusive reason for believing no proposition. Thus, the same arguments which led us to hold that exact contradictions ought not to entail do suggest as well that those contradictions and contradiction generators which are not exact contradictions ought not to entail and we conclude, therefore, that the consequences of the above definitions are

fully in accord with the intuitions which prompted them.

As we have noted, neither r nor $\bar{r}$ will be naturally entailed by the contradiction $r \cdot \bar{r}$. Nor will the contradiction $r \cdot \bar{r}$ be naturally entailed by the propositions r and $\bar{r}$. Thus, the introduction of natural entailment into a formal system will restrict the scope of the standard conjunction elimination and introduction rules which in all cases allow a conjunct to be inferred from a conjunction and a conjunction to be inferred from the set of its conjuncts. This is not all, however: natural entailment demands restrictions on the scope of all the standard inference rules associated with material entailment. As a consequence, there ceases to be, in a natural entailment system, such a thing as a universally valid form of argument. All the inference rules have to be qualified to take into account the cases in which contradictions occur as components of propositional compounds. It may be objected that in many cases it is difficult to know for certain that none of the premises in our argument is a contradiction or contains a contradiction and that if natural entailment is adopted we will not know, in these cases, whether we have a valid argument. I submit, however, that this is how things should be: the validity of an argument should be conditional, depending on whether or not contradictions are employed in the deduction. If they are the argument should be rejected; if they are not, or if there is no good reason for believing that any of the premises is a contradiction or contains a contradiction, then the argument should be accepted as a valid

natural entailment, or as valid pending the demonstration of the existence of a contradiction amongst the premises.

§3 Natural Entailment and the Conditional

One significant consequence which we must accept if we adopt the natural entailment relation defined in the previous section concerns the conditional. The standard material conditional is not true only if the antecedent is true and the consequent is not. In a natural entailment system, however, the conditional is not true not only when its antecedent is true and its consequent false but also when its antecedent is a contradiction.⁵ How this is a consequence of adopting natural entailment is straightforward enough: a conditional can be seen as the consequence of a subordinate proof whose hypothesis or hypotheses is or are the antecedent of the conditional and whose conclusion is the consequent. The conditional will be true if and only if the consequent is entailed by the antecedent together with the axioms and perhaps other hypotheses from preceding subordinate proofs. Now, if we adopt an entailment relation which has the consequence of disallowing contradictions from entailing, it is clear that no conditional whose antecedent is a contradiction can be true

⁵In this sentence propositions are spoken of both as being not true and as being false. This difference in usage marks what is, in the following, an important distinction: the assertion that a given proposition is not true simply precludes the truth of that proposition; the assertion that a given proposition is false not only precludes the truth of that proposition but implies that the negation of that proposition is true.

because the antecedent, i.e. the premise of the subordinate proof whose consequence is the conditional, can entail nothing and so, in particular, cannot entail whatever the consequent of the conditional in question happens to be. Thus, no conditional whose antecedent is a contradiction can ever be true.

These remarks make it clear, a fortiori, that in a natural entailment system it will not be possible to establish anything by using proofs which employ subordinate proofs whose hypotheses are contradiction generators, i.e. whose hypotheses materially entail exact contradictions. It will surely be objected that the adoption of such a procedure will rule out use of the argument form known as reductio ad absurdum, one of the mainstays of mathematical proof procedure. Let us closely examine the reductio argument form to see if this need, in fact, be the case.

A reductio argument is usually of the following form:

1	$\bar{p}$	hyp
.	.	
.	.	
.	.	
n	$q \cdot \bar{q}$	
n+1	$\bar{p} \supset (q \cdot \bar{q})$	n, imp int
n+2	$-(q \cdot \bar{q})$	Law of Non-Contradiction
n+3	$\bar{\bar{p}}$	n+1, n+2, modus tollens
n+4	p	n+3, double neg elim

A set of premises is hypothesized and it is demonstrated that they lead to an exact contradiction. Hence a material condi-

tional may be introduced whose antecedent is the set of hypothesized premises and whose consequent is the exact contradiction. Since the denial of the consequent is assumed as an axiom, the denial of the antecedent follows immediately. This procedure clearly relies on using a contradiction as an hypothesis or, equivalently, as the antecedent of a conditional, a situation which natural entailment prohibits. Are we forced, if we adopt natural entailment, to deny the legitimacy of every reductio argument?

I believe that a way may be found both to preserve the results of reductio arguments and to hold true to our intuitions which balk at the thought of inferring propositions from contradictions.⁶ To illustrate my argument I should like to look at a simple example of a reductio argument:

1	$fa \cdot \overline{fb} \cdot a=b$	hyp
2	fa	1, conj elim
3	$a=b$	1, conj elim
4	fb	2, 3, Leibniz's Law
5	$\overline{fb}$	1, conj elim
6	$fb \cdot \overline{fb}$	4, 5, conj int
7	$(fa \cdot \overline{fb} \cdot a=b) \supset (fb \cdot \overline{fb})$	1-6, imp int
8	$\neg(fb \cdot \overline{fb})$	Law of Non-Contradiction
9	$\neg(fa \cdot \overline{fb} \cdot a=b)$	7, 8, modus tollens

As an alternative to the preceding proof consider the following:

⁶I am indebted to S. Talmage for the general idea underlying the solution here proposed to this problem.

1	$fa \ . \ a=b$	hyp
2	fa	1, conj elim
3	$a=b$	1, conj elim
4	fb	2, 3, Leibniz's Law
5	$\overline{fb}$	hyp
6	$\overline{fb}$	5, rep
7	$-(fa \ . \ a=b \ . \ \overline{fb})$	4, 6, contradiction avoidance

In this proof the same conclusion is reached as in the preceding proof. Here, however, it is not necessary to assume, i.e. to hypothesize, a contradiction. Instead, components of the contradiction $fa \ . \ a=b \ . \ \overline{fb}$ are assumed, components which are assumed not to be contradictions themselves, and from them the two parts of the exact contradiction derived in the previous proof are obtained separately. We then apply a new rule, the contradiction avoidance rule, which enables us to assert the negation of the contradiction. This rule can be stated as follows: if one proposition naturally entails another and if a third naturally entails the negation of the second then the negation of the conjunction of the first and third can be asserted $((p \supset q) \ . \ (r \supset \bar{q})) \supset -(p \ . \ r)$.

In the system we are proposing the contradiction avoidance rule will be primitive and not derivative as it is in most systems of material entailment. It might be objected, though, that this is not so, that our contradiction avoidance rule is superfluous since it can be shown to be derived from the Law of Non-Contradiction and modus tollens. Thus:

1	$p \supset q$	hyp
2	$r \supset \bar{q}$	hyp
3	$p . r$	hyp
4	$p \supset q$	1, reit
5	p	3, conj elim
6	q	4, 5, modus ponens
7	$r \supset \bar{q}$	2, reit
8	r	3, conj elim
9	$\bar{q}$	7, 8, modus ponens
10	$q . \bar{q}$	6, 9, conj int
11	$(p . r) \supset (q . \bar{q})$	3-10, imp int
12	$-(q . \bar{q})$	Law of Non-Contradiction
13	$-(p . r)$	11, 12, modus tollens

This demonstration, however, relies on a subordinate proof whose premise is a contradiction. And, as we noted earlier, such subordinate proofs can establish nothing for nothing is naturally entailed by the premise of such a proof. What is superfluous, in a natural entailment system, is not the contradiction avoidance rule but rather the Law of Non-Contradiction itself. For, so far as I am aware, the only significant use for this law is in reductio proofs of the standard form discussed above where, in conjunction with modus tollens, it can be used to assert the denial of premises which materially entail an exact contradiction. With the modification of the reductio argument form, however, and the introduction of the contradiction avoidance rule, the Law of Non-Contradiction is

no longer necessary.

It might be thought that in rejecting the Law of Non-Contradiction we are at the same time disposing of the only justification for our rule of contradiction avoidance. But this is not so. What we are claiming is disposable is a rule which enables one to assert the negation of any line of a proof of the form $p \cdot \bar{p}$. We are not denying and, in fact, strongly hold to the intuition which says that the conjunction of a proposition with its negation must be false. (We would, however, deny that this intuition must be supported by the argument that the conjunction of a proposition with its negation must be false because otherwise every proposition would follow: 1) This is not a good argument. It begs the question as to what is wrong with every proposition of the language following from one proposition. 2) The intuition which says that the conjunction of a proposition with its negation must be false is in no need of further support.. Contradictions are not the case for the very fundamental reason that if any proposition is the case it excludes the possibility of its negation's being the case. This is an utterly basic feature of logical space, one which determines the possibility of language and thought, one which cannot be denied without being supposed.)

Thus, generalizing from the example given above, we conclude that the change in the truth conditions of the conditional, which is a consequence of the adoption of natural entailment, does not invalidate the results of reductio argu-

ments. Rather, employing this altered conditional just forces us to rewrite reductio arguments in a form which, since they now rest on legitimate principles of natural entailment, is more intuitively satisfying.

§4 The Liar Paradox

So far the only reason offered for accepting a formal deductive system employing natural entailment is that it is in accord with those intuitions which reject the notion that a contradiction can entail a proposition or that a contradiction can occur as the antecedent of a true 'if...then...' proposition. This reason will remain the impetus for the programme but agreeable consequences of this natural entailment approach will lend additional support to the project. One such consequence is a new and, I feel, superior method of dealing with some of the standard logical paradoxes. The paradox with which I shall be concerned here is the Liar paradox but the approach developed in the preceding sections has relevance for the Russell paradox and possibly for others as well.

The threat of the Liar paradox stems from the fact that it seems to establish a contradiction in an intuitively unobjectionable fashion. What is viewed as threatening, though, is the contradiction, for it is feared that if it is established then every proposition in the language in which the Liar occurs will be proved. If we deny, however, that

anything is entailed by a contradiction then the Liar paradox becomes completely innocuous. It remains puzzling, to be sure, but its occurrence in a language need not lead us to take drastic measures in order to banish it.

The Liar paradox can be expressed in the following way:

Key: P_I -- the principle that if a true proposition says that p , then p

P_{II} -- the principle that if a proposition says that p , and p , then the proposition is true

1	a says that a is not true	hyp (empirically true statement)
2	a is true	hyp
3	a is not true	1, 2, P_I
4	If a is true then a is not true	2-3, imp int
5	a is not true	4, Law of the Excluded Middle
6	a is true	1, 5, P_{II}

or

1	a says that a is not true	hyp (empirically true statement)
2	a is not true	hyp
3	a is true	1, 2, P_{II}
4	If a is not true then a is true	2-3, imp int
5	a is true	4, Law of the Excluded Middle
6	a is not true	1, 5, P_I

It is supposed that 1 could be empirically verified. That is, it is supposed that there could be a set of conventions for assigning certain names to certain propositions (e.g. conventions concerning chalk marks on blackboards) such that if one were to inquire as to what proposition the name 'a' was attached to one would discover that it was the proposition 'a is not true' so that it would follow that 'a says that a is not true'.

In this proof the first hypothesis leads legitimately by material entailment to an exact contradiction, indicating that it is itself a contradiction. This being the case there is no need to worry that 'a says that a is not true', which apparently could be true, 'establishes' an exact contradiction; for, being a contradiction itself, it naturally entails nothing.

It will be objected that this is no solution to the paradox at all: even if we allow that 'a says that a is not true' is a contradiction and allow that it naturally entails nothing the paradox does not disappear, for the problem is that 'a says that a is not true' apparently can be true and yet as a contradiction it must necessarily be false.

The answer to this objection is, quite simply, that it is not the case that in a formal system based on natural entailment all contradictions need be false. In the discussion of the reductio argument (§2) we saw that in a natural entailment system the Law of Non-Contradiction would be replaced by a contradiction avoidance rule. The contradiction avoidance rule, however, does not allow us to automatically assert the

negation of every contradiction: only those contradictions which can be divided into propositional components in such a fashion that one group of the components, not itself a contradiction, leads to one half of an exact contradiction and some other group of the components, also not itself a contradiction leads to the other half of the exact contradiction can be denied. This result is completely in accord with the intuitions which prompted our adoption of the contradiction avoidance rule: an exact contradiction is false because of the fundamental opposition of its two terms, because of the fact that a proposition's being the case excludes the possibility of its negation's being the case. This sort of polarity is also present in contradictions which can be divided into propositional components such that one group of these components naturally entails one term of an exact contradiction while a different group naturally entails the other term. It is this possibility of division which warrants us in denying the contradiction. A contradiction which cannot be so divided, however, is not subject to the same reasoning concerning its rejection and indeed there seems to be no reason at all for rejecting such a contradiction (there need be no worry about not rejecting it since, as a contradiction, it naturally entails nothing). Let us make this distinction between sorts of contradictions clearer by introducing the following definitions:

a polylithic contradiction is one which is a conjunction, N , consisting of n conjuncts, each of them a (well-formed)

formula, such that there exists an exact contradiction materially entailed by N which meets the following conditions:

- (i) some subclass P of N containing p conjuncts ($p < n$) naturally entails one term of this exact contradiction
- (ii) some other subclass Q of N containing q conjuncts ($q < n$), none of which is also a component of P , naturally entails the other term of this exact contradiction

a monolithic contradiction is one which is not a polyolithic contradiction

(A note of explanation should be made here: consider an exact contradiction $r, \bar{r}$. Let us suppose that r is naturally entailed by P , a conjunction of any number of formulae, together with R (both P and R are assumed to be used in the proof); and, let us suppose that $\bar{r}$ is naturally entailed by Q , a conjunction of any number of formulae none of which, however, is a conjunct of P , together with R (both Q and R are assumed to be used in the proof). Now, it would appear that the contradiction P, Q, R fails to meet the criterion for being a polyolithic contradiction. This is only true, though, if R is not analytic. A formula will be said to be analytic iff (i) it is of, or can be put into, the form of a conditional and its antecedent naturally entails its consequent (generalized formulae will be called analytic iff they can be replaced by formulae in which the quantified variables are replaced by free variables and which are of the form of conditionals whose antecedents naturally entail their consequents) or (ii) it is naturally entailed by an analytic formula or the conjunction

of some number of analytic formulae. If R is analytic then $P \cdot Q \cdot R$ is a polythetic contradiction for P alone naturally entails r and Q alone naturally entails $\bar{r}$. This is so because the conjunction of an analytic formula and any non-contradictory formula naturally entails no more than the non-contradictory formula does alone.⁷ And this in turn is a consequence of the fact that an analytic formula follows from the axioms and/or from formulizations of the rules of inference; thus, given any analytic formula a and any non-contradictory formula b , b naturally entails $b \cdot a$.)

It follows immediately from the definitions of polythetic and monolithic contradictions and from the definition of the contradiction avoidance rule that only the negations of polythetic contradictions can be asserted in a formal system employing natural entailment; the negations of monolithic contradictions cannot, on the basis of the contradiction avoidance rule, be asserted in such systems. This situation provides a solution to our problem with 'a says that a is not true': since the contradiction 'a says that a is not true' is monolithic it does not follow, simply from its being a contradiction, that it is false and consequently the tension between its apparently being possible that 'a says that a is not true' is true, and the fact that 'a says that a is not true' is a contradiction, is resolved.

Does this mean that we are committed to there being true contradictions? I do not believe so. How could it come

⁷For a more detailed exposition of this point see Appendix I p. 104.

about that 'a says that a is not true' is true? It could be true only if either it were directly empirically verifiable or it followed from propositions which were either directly empirically verifiable or necessarily true. As we have noted 'a says that a is not true' is not a directly empirically verifiable proposition; it depends for its truth partly on assuming conventions about associating names with propositions as well as upon empirically verifiable facts about whether the physical signs representing the name of 'a is not true' and the physical signs representing the proposition itself conform to the naming conventions. Thus 'a says that a is not true' could be true only if it followed from a group of propositions some of which were directly empirically verifiable and some of which were necessarily true. But 'a says that a is not true' is a contradiction and since anything which materially entails it is itself a contradiction generator nothing can naturally entail it. Consequently, in a system of natural entailment, 'a says that a is not true' cannot be true (this, however, does not mean that it is false; we shall discuss this question below).

While 'a says that a is not true' is not true (though perhaps not false) what about the conjunction of empirical and necessary truths which materially entail it? Is this conjunction true? On the standard view of entailment this conjunction would be true if all of its conjuncts were true, the reason for this being that a set of conjuncts always entails their conjunction. But this conjunction is a contradiction,

since it materially entails the contradiction 'a says that a is not true', and so it follows that the set of conjuncts which materially entails it constitutes a contradiction generator and consequently the conjunction naturally entails nothing. Thus, as with 'a says that a is not true', the conjunction which materially entails 'a says that a is not true' cannot be true for the condition which must be met in order for it to be true cannot be met: it cannot be naturally entailed by its conjuncts. This leaves us with the question of whether the contradiction generator, equivalent to the set of conjuncts of the contradictory conjunction which materially entails 'a says that a is not true', is true. But this cannot be the case: for a set of propositions is neither true nor false. Finally, then, we arrive at this situation: 'a says that a is not true' is materially entailed by a contradiction generator, in this case a set of propositions, each of which may be true, which itself has no truth value. Both the conjunction of the propositions constituting the contradiction generator and 'a says that a is not true' are not true because to say that they were true would be to say that they were naturally entailed by a group of empirically and analytically true propositions; yet, being contradictions, nothing can naturally entail them.

We have argued 1) that 'a says that a is not true' need not be false because, as a monolithic contradiction, it is not subject to the contradiction avoidance rule and 2) that 'a says that a is not true' cannot be true because, as a contradiction, there is nothing to make it true. Have we, then,

with 2), argued ourselves back into a position where we must say that 'a says that a is not true' is false (i.e. that its negation is true)? To say that if a proposition is not true then it is false is to affirm the Law of the Excluded Middle. Should we take it that the Law of the Excluded Middle holds in the language which is used to express the Liar paradox? It will be remembered that the Law of the Excluded Middle was used in the derivation which showed the Liar to be a contradiction (p. 52). For the sake of the argument, therefore, let us continue to assume that $p \vee \bar{p}$ holds in the language chosen to express the Liar. Note the following, however: since $p \vee \bar{p}$ is materially equivalent to $p \supset p$ we shall not be able to hold an unqualified Law of the Excluded Middle in a natural entailment language. If p is a contradiction p cannot occur as the antecedent of a true conditional and hence, in particular, the conditional $p \supset p$ can never be true. Thus, in a system employing natural entailment the Law of the Excluded Middle is seen not to hold of contradictions and since, to return to the case at hand, 'a says that a is not true' is a contradiction, the Law will not hold of it. Consequently, we will not be able to infer from the Law of the Excluded Middle and from the fact that 'a says that a is not true' is not true to the conclusion that 'a says that a is not true' is false.

This, then, is our position: the statement 'a says that a is not true' is a monolithic contradiction. It is not true because there is nothing which could make it true. This, however, is not to say that it is false. Its falsehood cannot

be inferred from its not being true because even if the Law of the Excluded Middle is accepted it has to be limited, in a language which employs natural entailment, to propositions which are not contradictions. Since 'a says that a is not true' is a contradiction the Law will not hold of it and so its not being true is not a sufficient ground for inferring its falsehood. However, it could still be false: Its falsehood would be demonstrated if a proof of its falsehood which did not rely on the fact that 'a says that a is not true' is not true and on the Law of the Excluded Middle could be produced.

To sum up: By adopting a formal system employing a natural entailment relation defined as we have defined it, a position which is both in accord with our intuition and which allows a happy solution to the Liar paradox is arrived at. Lest it be felt that the solution to the Liar paradox offered here is of an ad hoc nature let me, before we return to further discussion of Gödel's argument, contrast it briefly with Russell's solution, one which appears to have been widely accepted.

§5 Russell's Solution to the Liar Paradox

Roughly speaking, Russell pinned the blame for the contradiction arising from the Liar on the fact that the Liar is a self-referential statement.

When a man says 'I am lying', we may interpret his statement as: 'There is a proposition which I am affirming and which is false'. All statements that

'there is' so-and-so may be regarded as denying that the opposite is always true; thus 'I am lying' becomes 'It is not true of all propositions that either I am not affirming them or they are true'; in other words, 'It is not true for all propositions p that if I affirm p , p is true'. The paradox results from regarding this statement as affirming a proposition, which must therefore come within the scope of the statement. This, however, makes it evident that the notion of 'all propositions' is illegitimate; for otherwise, there must be propositions (such as the above) which are about all propositions and yet cannot, without contradiction, be included among the propositions they are about. Whatever we suppose to be the totality of propositions, statements about this totality generate new propositions which, on pain of contradiction, must lie outside the totality. It is useless to enlarge the totality, for that equally enlarges the scope of statements about the totality. Hence there must be no totality of propositions, and 'all propositions' must be a meaningless phrase.⁸

The vicious self-reference which leads to the contradiction lurks in the expression 'all propositions', an expression which must, as a consequence, be banned. This banishment finds its formal outlet in the theory of types, wherein propositional functions are assigned to a type according to the order of the variables occurring in them. It is then legislated that propositional functions cannot contain as arguments functions of a higher order than themselves. As a result, expressions like 'all propositions' must be replaced by bounded expressions like 'all propositions of order n '.

Russell says

Wherever 'all propositions' are mentioned, we must substitute 'all propositions of order n ', where it is indifferent what value we give to n but it is essential that n should have some value. Thus

⁸Bertrand Russell, 'Mathematical Logic as Based on the Theory of Types', in From Frege to Gödel, ed. Jean van Heijenoort (Cambridge, Mass.: Harvard University Press, 1967), p. 154.

when a man says 'I am lying', we must interpret him as meaning: 'There is a proposition of order n which I affirm and which is false'. This is a proposition of order $n+1$; hence the man is not affirming any proposition of order n ; hence his statement is false, and yet its falsehood does not imply, as that of 'I am lying' appeared to do, that he is making a true statement. This solves the Liar.⁹

Without doubt this is a solution of sorts. But it surely is as ad hoc as any solution to this paradox could be. What Russell has done is to identify a unique feature of propositions like the one in the Liar paradox, namely, their self-referentiality, and then has constructed a formal language in which propositions with this characteristic cannot occur. While it is true that if propositions which refer to themselves are excluded from a language the paradox will be avoided, there is no consideration of whether this is the only way in which the paradox can be avoided nor is there any further rationale, i.e. beyond their leading to the paradox, for excluding self-referential propositions. (Put another way, if the paradox is taken seriously then Russell's solution begs the question.) It is for these reasons that I say that Russell's solution appears ad hoc.

Russell does not, therefore, present a genuine solution to the Liar or, I believe, related paradoxes. He does not explain why they ought not to occur, he simply ensures that they will not occur in his formal language. The solution presented above to the Liar paradox 1) is at least no more ad hoc than Russell's: in an analogous manner we single out a feature

⁹Russell, p. 166.

of the Liar -- its monolithicity (Russell fastened on to its self-referentiality) -- and construct a formal system in which this feature is rendered harmless (Russell so constructs his system that the troublesome feature cannot occur) 2) does go further towards genuinely explaining the paradox because the process of singling out monolithic contradictions and prohibiting them from occurring as hypotheses is supported by reasons -- sound intuitions regarding what follows from contradictions -- whereas Russell's elimination of self-referential sentences is not.

CHAPTER IV

GÖDEL'S ARGUMENT AND THE METALANGUAGE M_N

At the beginning of the preceding chapter we raised the possibility of using a metalanguage different from the one Gödel uses (M) in order to investigate P . The rest of that chapter was spent 1) in describing the key aspect in which one such potential alternative metalanguage (M_N) differs from the sort of standard formal language we may assume Gödel's M , were it formalized, to be, viz. in its employment of natural as opposed to material entailment, and 2) in justifying the adoption of M_N over M by arguing that (i) the concept of natural entailment is closer than is the concept of material entailment to what we intuitively take entailment to be and (ii) a new and superior solution to the Liar paradox becomes possible if a language using natural entailment is adopted.

In this chapter I should like to investigate the possibility of reproducing what Gödel proves about P but now using, instead of M , the metalanguage M_N . I shall show that in M_N the sort of proof produced in M cannot be provided, i.e. that the possibility of proving P to be incomplete depends on certain features of the metalanguage in which this proof is attempted. The question of the completeness of P is, therefore, shown to depend fundamentally on the choice of metalanguage used to investigate P . Thus, we can assert that while P is M -incomplete it is not necessarily M_N -incomplete.

Since M_N incorporates certain features which are superior to features of M it can be said to be the preferred metalanguage. If we agree that P will be called 'complete' or 'incomplete' according as to whether it is demonstrated to be complete or incomplete in the preferred metalanguage then we can assert that Gödel's argument does not show P to be incomplete. Hence, a fortiori, it does not show arithmetic to be incomplete.

Our decision to adopt a metalanguage based on natural entailment does not have significance for Gödel's argument alone. Tarski's argument, which we discussed in Chapter II, also turns out to be interestingly affected by the adoption of a metalanguage M_N^0 , which is the result of modifying M^0 through introducing natural in place of material entailment. We shall examine the effects that the adoption of the metalanguage M_N^0 has on his argument in the fourth section of this chapter.

This chapter is the final chapter in the paper. By way of closing I shall, in the last section, provide a summary of the major results which I claim to have established in the course of the paper.

§1 The Three Versions Revisited

What we shall do in order to show that if one uses M_N as the language for investigating P then one cannot prove, by means of Gödel's argument, that there are undecidable formulae in P is demonstrate that each of the three versions of the

proof presented in Chapter I relies on arguing from, i.e. hypothesizing, a monolithic contradiction. Since in M_N nothing can be established by conditional or subordinate proofs whose hypotheses are contradictions, and since monolithic contradictions, unlike polythetic ones, cannot be so divided as to be used along with the contradiction avoidance rule in the revised reductio argument form permissible in M_N , it follows that if Gödel's argument requires the hypothesizing of a monolithic contradiction then it cannot be presented in M_N .

I shall repeat the argument for the non-reproducibility of Gödel's result in M_N with specific reference to each of the three versions presented in Chapter I. One crucial assumption of this argument is, in each case, that the premises of Gödel's argument, with the exception of the one which I wish to maintain is monolithic, are all analytic. This assumption I shall attempt to justify fully only for the second version (showing that it holds of the other versions presents no additional difficulties). Since the justification requires the discussion of some points at length I shall postpone it and devote the whole of the next section to it; for now I shall simply make the assumption.

Consider first version I: The initial eight premises in this version are analytic (only the first seven shall interest us here). What about the status of the hypothesis made in line 11? Line 11 is the hypothesis of a reductio proof aimed at proving its negation. That is, line 11, along

with lines 1-7, materially entails a contradiction (line 17) and this fact is used to reject line 11, i.e. to assert that X is not P-provable. Now, since lines 1-7 are analytic it follows, according to the theory developed in Chapter III, that if line 11 does not meet the criterion for polytheticity then it is a monolithic contradiction.¹ It is indeed the case that 'X is not P-provable' is not a conjunction such that there exists an exact contradiction materially entailed by this conjunction, one term of which is naturally entailed by one group of the conjuncts and the other term of which is naturally entailed by a completely different group of the conjuncts. Consequently, we must conclude that line 11 is a monolithic contradiction. Since it is not possible to establish anything in M_N by using subordinate proofs whose hypotheses are contradictions and since line 11, as a monolithic contradiction, cannot be so divided as to be used in conjunction with the contradiction avoidance rule in a legitimate reductio proof, it follows that in M_N the proof given in version I would not be reproducible.

¹It is assumed here that there are no analytic contradictions. This assumption can be argued for as follows: If there were analytic contradictions then since every contradiction materially entails an exact contradiction every analytic contradiction would materially entail an exact contradiction. However, it is also true that every proposition materially entails an analytic proposition. Therefore, every proposition would materially entail an analytic contradiction. But then every proposition would also materially entail an exact contradiction and consequently every proposition would be a contradiction. Thus there could be no natural entailments in a language which contained an analytic contradiction. This seems to be a sufficient ground for ruling their occurrence out.

Consider next version II: A similar argument can be seen to apply here. I wish to maintain that 'Bew17Genr' is a monolithic contradiction. The argument for this conclusion is as follows: line 14 ('Bew17Genr') conjoined with lines 1-7 and line 11 materially entails an exact contradiction, (the conjunction of lines 1 and 19). Since lines 2-7 and 11 are analytic and line 1 is analytic if true (this will be demonstrated in the next section) it must be the case that if line 1 is true then line 14 is a contradiction. Further, if line 14 is not a conjunction such that there exists an exact contradiction materially entailed by this conjunction, one term of which is naturally entailed by one group of the conjuncts and the other term of which is naturally entailed by a completely different group of the conjuncts then we must conclude that if line 1 is true then line 14 is a monolithic contradiction. It is evident that line 14 is not such a conjunction and hence we conclude that if line 1 is true then line 14 is a monolithic contradiction. Again, since it is not possible to establish anything in M_N by using subordinate proofs whose hypotheses are contradictions and since line 14, which, if line 1 is true, is a monolithic contradiction, cannot be so divided as to be used in conjunction with the contradiction avoidance rule in a legitimate reductio proof, it follows that in M_N the proof given in version II would not be reproducible. (Of course if line 1 is false all of Gödel's proof is invalidated. Note, however, that if line 1 is false this does not imply that line 14 is

2 2

OF/DE

not a monolithic contradiction.)

Consider finally version III: The same argument holds again. Lines 2-7 and line 14 are analytic and line 1 is analytic if true. Lines 1-7 and 14 conjoined with line 8 materially entail an exact contradiction (the conjunction of the second component of the conjunction in line 1 and line 21) making line 8, if line 1 is true, a contradiction. Line 8 is not a conjunction such that there exists an exact contradiction materially entailed by this conjunction, one term of which is naturally entailed by one group of the conjuncts and the other term of which is naturally entailed by a completely different group of the conjuncts. It is, therefore, if line 1 is true, a monolithic contradiction. As we have argued regarding each of the other versions, since it is not possible to establish anything in M_N by using subordinate proofs whose hypotheses are contradictions and since line 8, which, if line 1 is true, is a monolithic contradiction, cannot be so divided as to be used in conjunction with the contradiction avoidance rule in a legitimate reductio proof, it follows that in M_N the proof given in version III would not be reproducible.

§2 The Analyticity of the Premises of Version II

As I have remarked, it is a crucial assumption of the argument that Godel's result cannot be reproduced in M_N that the premises of his argument, with the exception of the hypothesis which is claimed to be monolithic, are analytic. In this

section I shall justify this assumption with regard to version II. (The justification of this assumption for the other versions, as I have said, raises no additional difficulties.) That is, I shall argue that lines 2-7 and 11 of version II are analytic and that line 1 is analytic if true. Let me begin by discussing the straightforward cases of lines 2-7 and 11.²

1) 2 is a stipulative definition. It could be expressed as the conjunction of the two conditionals 'If X is the relation $\lambda Q(x,y)$, then X is the relation $\lambda xBSb(y \overset{19}{z}(y))$ ' and 'If X is the relation $\lambda xBSb(y \overset{19}{z}(y))$ then X is the relation $\lambda Q(x,y)$ '. In each of these conditionals the antecedent naturally entails the consequent. This is so because of our agreement or stipulation that $\lambda Q(x,y)$ will be the relation $\lambda xBSb(y \overset{19}{z}(y))$. Thus 1 is analytic.

2) Let us consider just one of the conjuncts of 3. What we say about it will apply equally to the other. $C(\lambda(xBy))$ says that the relation λxBy is recursive. This fact can be expressed by the conditional 'If X is the relation λxBy then X is recursive'. The proof that the antecedent of this conditional naturally entails the consequent is given by Gödel himself through the construction of the long list of recursive relations, the forty-fifth of which is the relation in question. Clearly there is no need to repeat that lengthy demonstration here.

3) 4 is already in the form of a conditional. That its consequent is naturally entailed by its antecedent follows at

²Refer to the definition of 'analytic' given above on pp. 74-75.

once, as Gödel says, from the definition of 'recursive'.

4) 5 is also already in the form of a conditional. It is a formalized expression of Gödel's Theorem V the proof of which suffices to show that the antecedent naturally entails the consequent.

5) As a biconditional 6 is the conjunction of two conditionals. That each of them is a natural entailment follows from the definitions of 'Bew' and 'B'.

6) Again, premise 7 is already in the form of a conditional. The antecedent says that the generalization of y with respect to x is provable. The consequent says that for any integer n the result of substituting the numeral associated with n for occurrences of the variable x in y will be provable. That the antecedent naturally entails the consequent follows from the definitions of 'Bew', 'Gen', and 'Sb'.

7) 11, like 1, is a stipulation, in this case the specification of a particular value for y . It can be expressed as the conjunction of two conditionals; the antecedent of each of these conditionals naturally entails its consequent on account of our agreement or stipulation that this be the case.

Let us now consider the question of whether line 1, which is the assumption that P is ω -consistent, is analytic if it is true. (We shall call this assumption 'A'.) It is important to see that the relevant question here is not that of the analyticity of A but of whether A is analytic if it is true: Gödel states his conclusion in a conditional form -- if P is consistent P is incomplete -- so that no demonstration

of A, a demonstration which we might investigate to see if its premises were natural entailments which naturally entailed A, is presupposed. In the absence of such a demonstration and in the absence of the need to provide one we must rather ask: is what we mean by hypothesizing A such that we can conclude that if A then A is analytic?

Our first question, then, must be: what does it mean to hypothesize A in meta-P? There are two possibilities:

1) Hypothesizing A can mean supposing that there is a finite proof of A in meta-P. If there is a finite proof of A in meta-P will A be analytic? Given that in the meta-P we select we shall insist upon (i) premises which are natural entailments and (ii) natural entailment rules, it follows that if there is a finite proof of A in meta-P then A will necessarily be analytic.

2) Hypothesizing A can mean supposing only that each of the finite conjuncts in the infinite conjunction equivalent to A is provable. Let us suppose that A is just the straight consistency assumption $(a) - (Bew(\forall Gena) \cdot Bew \text{ Neg}(\forall Gena))$. (This supposition is harmless for the same remarks will hold of the ω -consistency assumption. The straight consistency assumption has the advantages of being (i) notationally simpler and (ii) the one employed by Mostowski and Tarski to whose proofs the arguments presented here will also apply.) If A is understood in this fashion it can be expanded into an infinite conjunction as follows: We let $a_1, a_2, a_3, \dots$ be all the class signs of P; $g_1, g_2, g_3, \dots$ be the generalizations, respectively, of

$a_1, a_2, a_3, \dots$ with respect to the free variables occurring in them (hence $g_1, g_2, g_3, \dots$ form a subset of the set of sentential formulae of P); $p_1, p_2, p_3, \dots$ be all the possible sequences of well-formed formulae of P . 'xBy', as in Gödel, means 'x is a proof of y'. A, then, is the infinite conjunction

$$\begin{array}{l}
 p_1 B g_1 \supset p_1 \bar{B} \text{Negg}_1 \quad p_1 B g_1 \supset p_2 \bar{B} \text{Negg}_1 \quad p_1 B g_1 \supset p_3 \bar{B} \text{Negg}_1 \quad \dots \\
 p_2 B g_1 \supset p_1 \bar{B} \text{Negg}_1 \quad p_2 B g_1 \supset p_2 \bar{B} \text{Negg}_1 \quad p_2 B g_1 \supset p_3 \bar{B} \text{Negg}_1 \quad \dots \\
 p_3 B g_1 \supset p_1 \bar{B} \text{Negg}_1 \quad p_3 B g_1 \supset p_2 \bar{B} \text{Negg}_1 \quad p_3 B g_1 \supset p_3 \bar{B} \text{Negg}_1 \quad \dots \\
 \vdots \\
 p_1 B g_2 \supset p_1 \bar{B} \text{Negg}_2 \quad p_1 B g_2 \supset p_2 \bar{B} \text{Negg}_2 \quad p_1 B g_2 \supset p_3 \bar{B} \text{Negg}_2 \quad \dots \\
 p_2 B g_2 \supset p_1 \bar{B} \text{Negg}_2 \quad p_2 B g_2 \supset p_2 \bar{B} \text{Negg}_2 \quad p_2 B g_2 \supset p_3 \bar{B} \text{Negg}_2 \quad \dots \\
 p_3 B g_2 \supset p_1 \bar{B} \text{Negg}_2 \quad p_3 B g_2 \supset p_2 \bar{B} \text{Negg}_2 \quad p_3 B g_2 \supset p_3 \bar{B} \text{Negg}_2 \quad \dots \\
 \vdots \\
 p_1 B g_3 \supset p_1 \bar{B} \text{Negg}_3 \quad p_1 B g_3 \supset p_2 \bar{B} \text{Negg}_3 \quad p_1 B g_3 \supset p_3 \bar{B} \text{Negg}_3 \quad \dots \\
 p_2 B g_3 \supset p_1 \bar{B} \text{Negg}_3 \quad p_2 B g_3 \supset p_2 \bar{B} \text{Negg}_3 \quad p_2 B g_3 \supset p_3 \bar{B} \text{Negg}_3 \quad \dots \\
 p_3 B g_3 \supset p_1 \bar{B} \text{Negg}_3 \quad p_3 B g_3 \supset p_2 \bar{B} \text{Negg}_3 \quad p_3 B g_3 \supset p_3 \bar{B} \text{Negg}_3 \quad \dots \\
 \vdots
 \end{array}$$

Notice that this conjunction is trebly infinite: first, the conjunction of the finite conjuncts, each of which states that if a particular sequence of formulae is a proof of some sentential formula then a further, or perhaps the same, sequence of formulae is not a proof of the negation of that sentential formula, is infinite because there are an infinite number of sequences of formulae which are not proofs of the negations of the sentential formula in question; moreover since

for a single, but unspecified, sentential formula there are an infinite number of sequences of formulae any one of which could be a proof of it, there must be an infinite number of such infinite conjunctions; finally, since there are an infinite number of sentential formulae, A must be equivalent to an infinite conjunction each conjunct of which is equivalent to an infinite number of infinite conjunctions.

The fact that no demonstration of the provability of each conjunct of an infinite conjunction is possible is irrelevant: what is important is what we are to say about the hypothesis that each conjunct is provable. If hypothesizing A is taken to mean assuming that each of the finite conjuncts in the infinite conjunction is provable then is A, if true, analytic? Assuming that this infinite conjunction naturally entails each of its conjuncts and conversely, that the infinite set of conjuncts naturally entails the conjunction, then clearly A, if true, is analytic. For, if A is true each finite conjunct of the treble infinite conjunction equivalent to A is analytic. For example, if A is true then the conditional 'if p_1 is a proof of a_1 then p_1 is not a proof of Nega_1 ' is true. If this conditional is true then, by truth-functional equivalence, either p_1 is not a proof of a_1 or p_1 is not a proof of Nega_1 . But if p_1 is not a proof of a_1 it will be analytically true that p_1 is not a proof of a_1 and similarly if p_1 is not a proof of Nega_1 it will be analytically true that p_1 is not a proof of Nega_1 . Consider ' p_1 is not a proof of a_1 ': this statement can be expressed as the condi-

tional 'if X is the sequence of formulae p_1 then X is not a proof of a_1 ' or again, more fully, as 'if X is a certain sequence of formulae (here the sequence is specified) then X is not a sequence of formulae with certain characteristics (here the characteristics are specified)'. This latter expression can again be expressed more fully and this process carried on until we arrive at an expression where, supposing ' p_1 is not a proof of a_1 ' to be true, the antecedent of the conditional clearly naturally entails the consequent (a crude example: 'if X is the sequence a_2, a_3, a_4 , then X is not a sequence of formulae whose last term is a_1 '). Thus, if it is true, ' p_1 is not a proof of a_1 ' is analytic and the same holds for ' p_1 is not a proof of $\text{Neg}a_1$ '. Consequently, each finite conjunct of A, if it is true, is analytic. From the fact that each conjunct of A is analytic and from the fact that these conjuncts naturally entail the conjunction equivalent to A it follows that if A is true it is analytic.

Now, it might be objected that it is not proper to speak of an infinite conjunction's being entailed, that the term 'entailment' ought only be applied to relations which can be shown to hold by means of a demonstration consisting of a finite number of steps. To this it may be replied that (i) there seems to be no good reason for this restriction and (ii) if hypothesizing an infinite conjunction in meta-P is to be meaningful then an infinite conjunction must be capable of being entailed by the axioms and definitions of meta-P; that is, if we are going to allow that hypothesizing A in meta-P

can have the sense of hypothesizing an infinite conjunction then we must allow that it is possible to speak, with full propriety, of an infinite conjunction's being entailed.³

Thus, taking the hypothesis A in either of the two senses in which it can be taken we see that if A is true then it is analytic. This completes our demonstration, regarding version II of Gödel's proof, of the analyticity of lines 2-7 and 11 and of line 1 if it is true.

§3 The Status of the M-Undecidable Formula

The argument in the preceding two sections has shown that if the consistency assumption is true then 'Bew17Genr' is a monolithic contradiction. The consequence of this, as we have said, is that in M_N a conditional proof in which 'Bew17Genr' occurs as a hypothesis cannot be used to establish anything. Thus Gödel's argument is shown not to be reproducible in M_N . This leaves us, however, with the question of what the status of 'Bew17Genr' is in M_N and with the further question of what we can say about the formula 17Genr.

³What we are raising here is the question of the possibility of infinite proofs. This is an interesting question and deserves further study. Formal definitions of 'proof' usually preclude the possibility of infinite proofs. Gödel's definition of what constitutes a proof in P is a case in point. In fact, his whole project hinges crucially on the impossibility of such a thing as an infinite proof: for, were infinite proofs possible the Gödel numbering scheme, whereby a finite number can be assigned to the sequence of formulae which is a proof of a given formula, would become inadequate. That is, it would become impossible to express the notion of provability in an arithmetized metalanguage. This would render futile the attempt to construct an undecidable formula in the fashion in which Gödel purports to do so.

The remarks made in Chapter III about the status of 'a says that a is not true' apply directly to 'Bew17Genr'.⁴ As a monolithic contradiction (assuming the consistency assumption to be true) it cannot be true and yet this does not necessarily mean that it is false, since in M_N the Law of Non-Contradiction does not apply to contradictions. It is possible, though, that it could be shown to be false. If 'Bew17Genr' were shown, in M_N , to be false (this, of course, would have to be through a proof which did not involve hypothesizing it) then, by using that portion of Gödel's argument which follows the purported demonstration of 'Bew17Genr'; P could be shown to be M_N -incomplete and hence incomplete. Or rather, this could be shown provided that 'BewNeg17Genr' is not a monolithic contradiction -- a question I have not investigated. It does not seem likely, however, that a proof of 'Bew17Genr' which did not involve hypothesizing 'Bew17Genr' could be given.

If such a proof of the falsity of 'Bew17Genr' is indeed impossible then 'Bew17Genr' is neither true nor false. What are the implications of this eventuality for the decidability of 17Genr? If it can never be decided whether 'Bew17Genr' or 'Bew17Genr', which would be a consequence of 'Bew17Genr's' being neither true nor false, then it is certainly the case that no proof of 17Genr could ever be provided. It does not follow, as it might be thought, that in this case we

⁴See pp. 75-79.

have simply reinforced Gödel's conclusion that 17Genr is undecidable. For, while we are committed to saying that if ' $\text{Bew}17\text{Genr}$ ' is neither true nor false then no proof of 17Genr can be provided it does not follow that a proof of $\text{Neg}17\text{Genr}$ is impossible. Gödel's argument for ' $\text{BewNeg}17\text{Genr}$ ' requires the proof of ' $\text{Bew}17\text{Genr}$ '. If no such proof is possible then it remains open whether or not ' $\text{BewNeg}17\text{Genr}$ ' is provable and hence whether 17Genr is decidable.

We may state our position then as follows: P may or may not be M_N -incomplete. It will be so only if 1) a proof of ' $\text{Bew}17\text{Genr}$ ' which does not rely on hypothesizing ' $\text{Bew}17\text{Genr}$ ' can be produced and 2) ' $\text{BewNeg}17\text{Genr}$ ' is not a monolithic contradiction.

§4 Tarski's Argument and M_N^0

Are there implications for Tarski's argument if the meta-metalanguage in which he carries out his argument is replaced by one using natural entailment? I shall argue here that there are. Let us recall, in brief, the general structure of Tarski's argument: what he argues is that if the concept of truth for a given language is definable in a meta-language of equivalent or lesser order than the language itself then the Liar paradox will be recreatable in the meta-language. More specifically, if ' Tr ' is definable in a meta-language of equivalent or lesser order than the language for which it is defined then it can be proven that a contradictory

pair of sentences in meta-P^0 will be among the consequences of this definition. Since the class of all provable formulae in the metalanguage in which 'Tr' is defined is assumed to be consistent, it follows that 'Tr' cannot be defined in a metalanguage of equivalent or lesser order than the language for which it is defined.

Since Tarski's proof is clearly related to the Liar paradox we may legitimately wonder whether a similar proof would be possible in a meta-metatheory employing natural entailment and conditionals (M_N^0). Let us turn our attention to the formalized version of the proof presented above on p. 47. As with the three versions of Gödel's argument we see that a reductio argument forms the core of the proof. Our enquiry, therefore, must be directed towards the status of the premises from which the contradiction serving to refute one of them -- the hypothesis that there exists an adequate definition of truth in meta-P^0 -- is derived.

Let us review the five premises occurring in the formalized version of Tarski's rough proof. The proof is not rigorously deductive which is to say that principles other than those mentioned explicitly in premises 3, 4, and 5 are required. I shall use the fact that Tarski himself must rely, if his reductio proof is to succeed, on the analyticity of these implicit principles as an excuse for not unearthing them further. In fact the same holds for premises 3, 4, and 5. I shall take it as apparent that they are analytic, for to demonstrate this would require, amongst other things, specifying

1) the infinite sequence ϕ which in turn requires specifying and ordering all the expressions in both P^0 and $\text{meta-}P^0$ and
 2) the numbering scheme by which, like Gödel numbering, numbers are assigned to every sign and hence to every sentential function of $\text{meta-}P^0$. This alone is too large a task to undertake here. Furthermore, a rigorous proof of premise 5, which would follow upon these specifications, would be no small chore either.

Assuming premises 3, 4, and 5 to be analytic this leaves premises 1 and 2 to be examined. Premise 1 is the consistency assumption for $\text{meta-}P^0$: if a sentence is provable in $\text{meta-}P^0$ then its negation must not be provable. Premise 2 is the hypothesis that there exists an adequate definition of truth in $\text{meta-}P^0$ (let us call this the hypothesis that D). It is important to note that, as with Gödel's result, Tarski's conclusion is stated in conditional form: if the class of all provable sentences of $\text{meta-}P^0$ is consistent then it is not the case that D. Tarski is aware of the central importance of this assumption:

The assumption of consistency appearing in the part (8) of this theorem [see Theorem 1' p. 40 above] is essential. If the class of all provable sentences of the metatheory contained a contradiction, then every definition in the metatheory would have among its consequences all possible sentences (since they all would be provable in the metatheory), in particular those described in the convention T [p. 37 above].⁵

The whole argument given in §2 concerning the analyticity of the consistency assumption A can be repeated here, though in

⁵Tarski, p. 251.

this case it is the metalanguage and not the object language whose consistency is being assumed. (As Tarski points out, however, given the existence of an interpretation of the metalanguage in the object language, the assumption of the consistency of meta-P^0 is equivalent to the assumption of the consistency of P^0 .) The assumption of meta-P^0 's consistency is either taken to mean that there is an acceptable finite consistency proof or it is taken to mean that the infinite conjunction to which the consistency assumption is equivalent is true. In either case it can be argued, as was done in §2, that the assumption is analytic if it is true. If this is so then, as in Gödel's argument, we see that the assumption that D is a monolithic contradiction. What are the implications of this?

First of all, as we have repeatedly said, since it is not possible to establish anything in languages employing natural entailment by using subordinate proofs whose hypotheses are contradictions, the assumption that D cannot be used to establish anything. Thus, the demonstration Tarski presents does not establish his conclusion -- that it is not the case that D. If we go on and ask about the truth of the proposition 'there exists an adequate definition of truth in meta-P^0 ' we find that the comments made about the truth of 'a says that a is not true' in Chapter III §4 apply here as well. That is, anything which materially entails D will itself be a contradiction or a contradiction generator and hence will naturally entail nothing. But since D can only be true if it is naturally entailed by

true propositions it turns out that D cannot be true. Again, this does not mean that D is false; for this to be demonstrated legitimately a direct proof of not- D , one which did not involve arguing from the fact that D is not true and the Law of the Excluded Middle, would have to be provided. It is not clear how such a proof could be provided. (I see no argument for saying, though, that such a direct demonstration is impossible and that hence we have demonstrated that there are sentences undecidable in the meta-metatheory.)

§5 Conclusions

The following are the major results of this paper:

- 1) Gödel's proof that there exist undecidable formulae in P can only be carried out in a metalanguage M which employs material entailment. In an alternative metalanguage M_N which employs natural entailment his argument cannot be reproduced. Thus we can say that while Gödel's argument shows P to be M -incomplete it does not show it to be M_N -incomplete.
- 2) M_N is to be preferred to M as a metalanguage on the grounds that (i) it employs a concept of entailment which is more closely in accord with our fundamental intuitions about what entailment is, particularly with our intuitions about what contradictions ought to entail and (ii) it offers a new and superior solution (superior at least to Russell's solution) of the Liar paradox. If we agree to call a language 'complete' or 'incomplete' according as to whether

it is complete or incomplete in the preferred metalanguage then, given 1), we can say that Gödel's argument does not show P to be incomplete.

- 3) Once it becomes impossible to argue on the basis of monolithic contradictions it is difficult to see how one could argue that there are undecidable formulae in P: for, only the potentially self-referential formulae of P seem to be linkely candidates for undecidability proofs; and it appears that to hypothesize the provability of such formulae, which is a required step in the sort of proof (a reductio proof) in which such formulae must be used, if they are to establish undecidability, is to hypothesize a monolithic contradiction. Thus, given that Gödel's argument does not show P to be incomplete, and given the unlikelihood of incompleteness through any other sort of argument, we may tentatively assert that there are no good reasons for believing that P is incomplete.
- 4) Tarski's argument that there cannot be constructed an adequate definition of truth in a metalanguage of equal or lesser order than the language for which the definition is constructed also relies crucially upon features of the metalanguage in which it is carried out. Again, by adopting

"It has recently been drawn to my attention that a new proof of Gödel's Theorem has been produced as a corollary to the solution of Hilbert's tenth problem. (See Martin Davis, 'Hilbert's Tenth Problem is Unsolvble', reprinted from the American Mathematical Monthly, Vol. 80, No. 3, March 1973, pp. 233-69.) Whether this new proof is subject to the sort of objections made to the version of Gödel's proof above is a question which I have not yet been able to investigate. It is, of course, deserving of immediate attention.

an alternative and preferred metalanguage (one using natural entailment) we can say that while the claim that there can be constructed an adequate definition of truth in a metalanguage of equal or lesser order than the language for which the definition is constructed is not true there are no good reasons for believing that it is false. That is, there are no good reasons for believing the claim that there cannot be constructed an adequate definition of truth in a metalanguage of equal or lesser order than the language for which the definition is constructed.

APPENDIX I

In this appendix I shall prove:

- I If $P \cdot R$ naturally entails r , where P is non-contradictory and R is analytic, then P naturally entails r .

The proof: We shall A prove that natural entailment is transitive and B prove that P naturally entails $P \cdot R$, where P is non-contradictory and R is analytic. A and B suffice to prove I.

- A Suppose q naturally entails s and s naturally entails t . Now, suppose that natural entailment is not transitive, i.e., that q does not naturally entail t . If this is so then since q materially entails t (material entailment is transitive), q must materially entail t and $\bar{t}$. But then q is a contradiction. Ex hypothesi, however, q naturally entails s and so cannot be a contradiction. Therefore natural entailment must be transitive.

- B We shall assume that we are working in a natural deduction system, i.e., one without axioms. (Any axiomatic deduction system can be converted into such a deduction system by the introduction of rules of inference corresponding to the axioms. For example, in place of the axiom $p \supset p \vee q$ one can introduce the rule of inference which allows one to write the line $p \vee q$ immediately after any line p . Thus, using this rule and conditionalization one can arrive at a formula of the form ' $p \supset p \vee q$ ' for any substitutions for the letters ' p ' and ' q ', except, of course, if p is a contradiction.)

1st case: R is of the form of a conditional whose antecedent naturally entails its consequent (suppose $R = a \supset b$). The following proof is possible:

1	P	hyp (P is a legitimate hypothesis for it is assumed to be non-contradictory)
2	a	hyp
3	b	Some rule of inference of a natural entailment system
4	R	imp int ($R = a \supset b$)
5	P . R	conj int

Steps 1-5 show that P naturally entails $P \cdot R$.

2nd case: R is naturally entailed by an analytic formula or the conjunction of some number of analytic formulae.

Suppose $S_1, S_2, \dots, S_n$ are analytic and when conjoined naturally entail R . Then each of the S_i is itself either of the form of a conditional whose antecedent naturally entails its consequent or is naturally entailed by some analytic formula or the conjunction of some number of analytic formulae. Each S_i for which the first condition holds can be introduced by a one step conditionalization as is R in the 1st case. Those S_i for which the second condition holds must be analyzed in order to discover whether the analytic formulae which naturally entail them are conditionals or whether the analytic formulae which naturally entail them are again naturally entailed by further analytic formulae. This process of analysis must

be carried on until simple conditionals are the only terms in the analysandum. Then each of the S_i which is not itself a conditional can be shown to be naturally entailed through the appropriate number of conditionalization steps. All the S_i obtained, R , can be entered into the proof and $P \cdot R$ obtained by conjunction introduction.

BIBLIOGRAPHY

- Anderson, Alan Ross, and Nuel D. Belnap. 'Entailment'. Logic and Philosophy: Selected Readings. Ed. Gary Iseminger. New York: Appleton-Century-Crofts, 1968, pp. 76-110.
- Fitch, Frederic Brenton. Symbolic Logic: An Introduction. New York: The Ronald Press Company, 1952.
- Gödel, Kurt. 'Some Metamathematical Results on Completeness and Consistency, On Formally Undecidable Propositions of Principia Mathematica and Related Systems I, and On Completeness and Consistency'. In From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931. Ed. Jean van Heijenoort. Cambridge, Mass.: Harvard University Press, 1967, pp. 592-617.
- Lewis, Clarence Irving. 'Strict Implication'. In Logic and Philosophy: Selected Readings. Ed. Gary Iseminger. New York: Appleton-Century-Crofts, 1968, pp. 67-75.
- Mates, Benson. Elementary Logic. 2nd ed. New York: Oxford University Press, 1972.
- Mostowski, Andrzej. Sentences Undecidable in Formalized Arithmetic: An Exposition of the Theory of Kurt Gödel. Amsterdam: North-Holland Publishing Company, 1952.
- Nagel, Ernest and James R. Newman. Gödel's Proof. London: Routledge & Kegan Paul Ltd., 1959.
- Russell, Bertrand. 'Mathematical Logic as Based on the Theory of Types'. In From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931. Ed. Jean van Heijenoort. Cambridge, Mass.: Harvard University Press, 1967, pp. 150-182.

Tarski, Alfred. 'The Concept of Truth in Formalized Languages'.

Trans. J.H. Woodger. In Logic, Semantics, Metamathematics:

Papers from 1923 to 1938. Oxford: Clarendon Press, 1956,

pp. 152-278.

van Heijenoort, Jean. 'Gödel's Theorem'. Encyclopedia of

Philosophy. (1967).

END

17 00 79

FIN