This dissertation has been
microfilmed exactly as received

68-13,319

GUSTASON, William Whitby, 1938-

NEGATION AND ASSERTION IN FREGE AND THE TRACTATUS.

The University of Michigan, Ph.D., 1968
Philosophy

University Microfilms, Inc., Ann Arbor, Michigan

NEGATION AND ASSERTION IN FREGE AND THE TRACTATUS

by
William Whitby Gustason

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in
The University of Michigan
1968

Doctoral Committee:
Professor Irving M. Copi, chairman
Associate Professor Peter G. Hinman
Associate Professor George I. Mavrodes
Associate Professor Julius M.E. Moravcsik

ABSTRACT

NEGATION AND ASSERTION IN FREGE AND THE TRACTATUS

by

William Gustason

Chairman: Irving Copi

In this essay, an attempt is made to explore critically some of Frege's remarks on the topics of assertion and negation in light of certain objections to them brought forward by Wittgenstein in his Tractatus Logico-Philosophicus and other writings. In so doing, some suggestions for a new treatment of these concepts, differing from Frege's in several important respects, are proposed and argued for, and some of the more significant consequences are spelled out.

Although these results are in large part derived from certain remarks made by Wittgenstein in the Tractatus and the Notebooks: 1914-1916, more space is devoted in this essay to Frege than to Wittgenstein; and chapter one is mainly a critical exposition of the salient features of his philosophy of logic as a whole. Even a brief examination of the relevant passages of the Tractatus will suffice to convince anyone that the views expressed there stem almost entirely from a rejection on Wittgenstein's part of certain fundamental points

in Frege's theories. Moreover, it is impossible to understand fully Frege's own remarks on assertion, negation, and truth without first acquiring a firm grasp of the two most fundamental and pervasive elements of his philosophy of logic, viz. the theory of functions and the sense-reference distinction. Hence, a substantial portion of the first chapter will be devoted to these topics.

In chapter two, several lines of criticism are developed. Frege's view of sentences as a kind of complex expression designating the True or the False is examined in the light of Wittgenstein's illuminating comment at 4.431. Also, some of the more dubious consequences of his theory of assertion are discussed at some length, and the classic, truth-functional account of negation, which has gained widespread acceptance since Frege introduced it in his Begriffsschrift, is discussed in detail. In the course of these discussions, some new ideas are introduced which, as I shall try to show in chapter three, may be used to obtain a more satisfactory account of sentences or propositions than Frege's.

In chapter three, the concept of negation is examined further, leading us, hopefully, to a more satisfactory analysis of singular propositions. This analysis derives to a large extent from some of Wittgenstein's remarks on "pictures" in the Tractatus, but I do not accept Wittgenstein's "picture theory" of the proposition. For I do not subscribe to the atomistic onotology of the Tractatus, nor therefore to the existence of

"elementary propositions". Rather, my concern is with logically as opposed to ontologically, simple propositions; that is, with the singular, uncompounded propositions of ordinary discourse. For it is in terms of these propositions that the truth-conditions of (and hence the senses of) general and truth-functional propositions must be expressed. Such propositions are, then, fundamental in logical theory, and I maintain that some of Wittgenstein's remarks about pictures can be utilized to give a more satisfactory account of them than is provided by Frege.

The results obtained in chapter three concerning negation and singular propositions are then used in chapter four to gain an understanding of Wittgenstein's difficult notion of an operation. This notion enables us to refute one of the most important features of Frege's theory of assertion, viz. the distinction between "asserted" and "unasserted" occurrences of a proposition. The other features of his theory are then examined critically and in the end I maintain that Wittgenstein is right in his contention that assertion is a purely "psychological" concept, and that ' $\vdash$ ' is "logically quite without significance". I shall then briefly attempt to expose the errors underlying the analysis of sentences as "names" of truth-values.

ACKNOWLEDGEMENTS

I wish to thank the members of my committee: Professor P. G. Hinman, Associate Professor George I. Mavrodes, Associate Professor J. M. E. Moravscik, and especially Professor Irving M. Copi, for their generous assistance and advice. This essay has been greatly improved because of their help.

TABLE OF CONTENTS

INTRODUCTION 1

CHAPTER ONE: FREGE'S PHILOSOPHY OF LOGIC 6

Frege's Theory of Functions 6
Extension of the Theory of Functions to Ordinary Discourse 13
Frege's Sense-Reference Distinction 15
Sentences and the Sense-Reference Distinction ... 22
Unsaturated Expressions and the Sense-Reference Distinction 43
Truth-Functions and Concepts 46
The Sense of a Sentence 52
Sense and Truth-Conditions 58
Assertion and Judgment 62
Sentences, Assertion, and Truth 69

CHAPTER TWO: EXAMINATION OF FREGE'S THEORIES OF NEGATION AND ASSERTION 76

Frege's Account of Negation 76
Frege on Assertion and Truth 87
Critique of Frege's Theory of Assertion 102
Critique of Frege's Theory of Negation 119
Negation and the Independent Criterion 134
Consequences of the Independent Criterion 144
Final Comments 150

CHAPTER THREE: NEGATION, PICTURES AND SINGULAR SENTENCES 153

A Problem Concerning Uninterpreted Calculi 153
Anscombe on Negation 158
Pictures as Projections of States of Affairs 166
Singular Sentences as "Projections" 176
Negation, Projection, and Anscombe's Problem 197
Extensional and Non-extensional Aspects of 'Not'. 202
Some Consequences of the Foregoing Analysis 221

CHAPTER FOUR: SENTENCES, OPERATIONS AND ASSERTION 234

Extensional and Non-extensional Aspects of Negation in English 234
Operations 242
Frege's Theory of Assertion Again 269
Final Comments on the Theory of Assertion 280
Sentences and Names of Truth-Values 307

SELECTED BIBLIOGRAPHY 315

INTRODUCTION

In this essay, an attempt is made to explore critically some of Frege's remarks on the topics of assertion, truth, and negation in light of certain objections to them brought forward by Wittgenstein in his Tractatus Logico-Philosophicus[1] and other writings. In so doing, some suggestions for a new treatment of these concepts, differing from Frege's in several important respects, are proposed and argued for, and some of the more significant consequences are spelled out.

Although these results are in large part derived from certain remarks made by Wittgenstein in the Tractatus and the Notebooks: 1914-1916,[2] more space is devoted in this essay to Frege than to Wittgenstein; and chapter one is mainly a critical exposition of the salient features of his philosophy of logic as a whole. Even a brief examination of the relevant passages of the Tractatus will suffice to convince anyone that the views expressed there stem almost entirely from a rejection on Wittgenstein's part of certain fundamental points in Frege's theories, and hence the ideas I wish to propound concerning assertion and negation can be properly developed only on the basis of showing that there are important logical features of propositions or sentences

---

[1] Ludwig Wittgenstein, Tractatus Logico-Philosophicus, trans. by D. Pears and B. McGuinness (New York: Humanities Press, 1961).

[2] Notebooks: 1914-1916, trans. by G.E.M. Anscombe (Oxford: Basil Blackwell, 1961).

--many of which are pointed up in the Tractatus--which Frege's theory either misconstrues or ignores. Moreover, it is impossible to understand fully Frege's own remarks on assertion, negation, and truth without first acquiring a firm grasp of the two most fundamental and pervasive elements of his philosophy of logic, viz. the theory of functions and the sense-reference distinction. Hence, a substantial portion of the first chapter will be devoted to a discussion of these topics. Furthermore, looked at historically, Frege's philosophy of logic is a remarkable accomplishment inasmuch as it exposed and corrected some of the confusions which had existed in logical theory since Aristotle's time, and there is much in his philosophy that is accepted by both Wittgenstein and myself--and to a large extent by virtually all modern logicians. Thus, Frege's theory deserves to be presented in its own right as the first truly coherent and comprehensive account of the fundamental issues in logical theory; and whatever counterintuitive elements are contained therein (e.g. that sentences refer to truth-values) are backed up by powerful arguments demanding our acceptance. Finally, like Descartes in the Meditations, whatever there is that is wrong with Frege's views certainly is not wrong in any obvious or superficial way; it still appears quite plausible even after a careful perusal. His mistakes, that is, are of a rather subtle and intricate nature--mistakes that can be very instructive and perhaps guide us to a sounder theory.

In chapter two, several lines of criticism are developed. Frege's view of sentences as a kind of complex expression designating the True or the False are examined in the light of Wittgenstein's illuminating comment at 4.431. Also, some of the more dubious consequences of his theory of assertion are discussed at some length, and the classic, truth-functional account of negation, which has gained widespread acceptance since Frege introduced it in his Begriffsschrift, is discussed in detail. In the course of these discussions, some new ideas are introduced which, as I shall try to show in chapter three, can be used to obtain a more satisfactory account of sentences or propositions than Frege's.

In chapter three, then, the concept of negation is examined further, leading us to a more satisfactory account of singular propositions. We are then in a position to explore more fruitfully the concept of assertion and to discover wherein Frege's theory goes wrong. The analysis of propositions developed here stems to a large extent from some of Wittgenstein's remarks on "pictures" in the Tractatus, but I wish it understood here at the outset that I do not accept Wittgenstein's "picture theory" of the proposition. For I do not subscribe to the atomistic ontology of the Tractatus, nor therefore to the existence of "elementary propositions" of which all other propositions are truth-functions. Rather, my concern in this essay is with logically, as opposed to ontologically, simple propositions; that is, with the singular, uncompounded propositions of ordinary discourse. For

it is in terms of these propositions that the truth-conditions of (and hence the senses of) general and truth-functional propositions must be expressed. Such propositions are, then, fundamental in logical theory, and I shall maintain that some of Wittgenstein's remarks about pictures can be utilized to give a more satisfactory account of them than is provided by Frege. I also wish to make it clear that I am not concerned with Tractatus explication as such. Very possibly I have misinterpreted some passages of that difficult work; however, my main concern is to shed some light on the topics of assertion and negation, not to engage in what Ryle has called "philosophical paleontology", I shall also have occasion to examine the views of some of Wittgenstein's commentators (most notably Miss Anscombe on negation and pictures). Again, my concern is not with whether they have the last word on Wittgenstein, but with whether the views they present are sound.

The results obtained in chapter three concerning negation and singular propositions are then used in chapter four to gain an understanding of Wittgenstein's difficult notion of an operation. This notion, as I try to show, enables us to refute one of the most important features of Frege's theory of assertion, viz. the distinction between "asserted" and "unasserted" occurrences of a proposition. The other features of his theory are then examined critically

and in the end I claim that Wittgenstein is right in his contention that assertion is a purely "psychological" concept, and that '⊢' is "logically quite without significance". I shall then briefly attempt to expose the errors underlying the analysis of sentences as "names" of truth-values.

CHAPTER ONE

FREGE'S PHILOSOPHY OF LOGIC

Frege's Theory of Functions

Frege's remarks on assertion and negation cannot be profitably discussed without first examining his doctrine that a sentence is a complex expression referring to one of two objects, the True or the False. But this itself cannot properly be understood unless we begin by briefly considering two of the most important features of Frege's philosophy of logic. The first of these is his theory of functions, the second, the sense-reference distinction.

The theory of functions grew out of Frege's research into the foundations of mathematics. Consider the following list of expressions:

$$2.1^3+1$$
$$2.2^3+2$$
$$2.4^3+4$$

these expressions respectively stand for or refer to the numbers 3, 18, and 132, and we can abstract from this list a common grammatical form, viz., '$2.()^3+()$'. Frege holds that this abstracted expression represents a function which has the values 3, 18, and 132 for the arguments 1, 2, and 4 respectively. Thus, the expression '2.1^3+1' stands for a value of the function $2.()^3+()$ for argument 1. Complex expressions like '2.1^3+1' can be analyzed into an expression for the function, '$2.()^3+()$' (hereafter, a functional expression), and an expression for the argument, in this case the numeral '1'. By inserting such an argument-expression into the "gaps"

in the functional expression, we thereby obtain a value-expression--a complex expression standing for the value of the function for that argument. An argument-expression can be either simple or complex (e.g. '7' or '5+2'), but if it is complex, it must also be a value-expression derived from a functional expression. For example, '5+2' is a value-expression derived from '()+2'. Though the number 132 is the value of the function $2.()^3+()$ for the argument 4, '132' is not a value expression, but '2.4^3+4' is.

Although a value-expression can be analyzed into an argument-expression and a functional expression, the latter is not a component of the value expression in the same straightforward way as the argument expression. For there are no occurrences as such of the functional expression, and no "places" into which it can be inserted. Rather, it provides the places and provides for the occurrences of the argument-expression by providing the form of the value-expression. Moreover, if we were to make a list of the well-formed expressions of elementary arithmetic, we would not include functional expressions, but only simple and complex expressions like '3' and '2.1^3+1'--expressions which contain no "gaps". Functional "expressions" such as '$2.()^3+()$' essentially have gaps into which legitimate expressions are inserted to produce new, legitimate expressions; they are, in Frege's words, "unsaturated". As such, they should be thought of as logico-linguistic patterns for generating complex, well-formed expressions from simpler well-formed

expressions, but they are not themselves well-formed (nor can they be properly said to be ill-formed). They are not expressions which have a form (satisfy the syntactical rules of the language), but rather they are forms--they express or illustrate just what the rules say. Further, inasmuch as they are patterns or forms for generating complex, well-formed expressions from other well-formed expressions, these functional expressions can themselves be regarded as functions which take value-expressions as values and argument-expressions as arguments.[1] for all these reasons, we are not to suppose that a functional expression refers to or stands for a function in the sense in which a "saturated" expression like '3' or '2+1' does. Functional expressions are not at all like singular terms or "referring expressions", and Frege's ontological commitment to the existence of functions was not based on any belief that functional expressions had to refer to something. Rather, the commitment stems from the necessity of quantifying over functions; that is, because of the need to formulate statements of the form "There is a function, f(), such that . .'. A value-expression like '6!' is not to be analyzed (as Carnap[2] once thought) into an argument expression '6' and a functional expression

---

[1] See G.E.M. Anscombe and Peter Geach, Three Philosophers (Ithaca: Cornell University Press, 1961), pages 143-144.

[2] Rudolf Carnap, The Logical Syntax of Language (London: Routledge and Kegan Paul, 1937), pages 292-304.

'!', for '!' is not a functional expression at all. That a function is represented in '6!' results from the fact that '!' occurs immediately to the right of the numeral, and this is expressed by writing the unsaturated functional expression '()!' .

Frege claimed that there are two mutually exclusive ontological categories: functions (which "correspond to" or "are represented by" or "are associated with" functional expressions) and objects (which are referred to or designated by saturated well-formed expressions, i.e. argument-expressions and value-expressions). So far, the only objects that have been considered are numbers and the only functions considered have been numerical ones, but we will see presently that Frege applied his theory of functions to other kinds of discourse besides arithmetic. One might well ask at this point, however, what reason Frege has for supposing that to the mutually exclusive linguistic categories of unsaturated (functional) expressions on the one hand and saturated expressions on the other, there corresponds the mutually exclusive ontological categories of function and object. Why can't a function also be an object and vice-versa? Inasmuch as the notion of a function was introduced solely as that which is represented by a functional expression, and the notion of an object solely as that which a saturated expression stands for or refers to, it seems reasonable to suppose that there might be entities which are both referred to by saturated expressions and represented by functional expressions. For example, one might well suppose that the same

entity is referred to by the saturated expression 'the square root function' and represented by the unsaturated expression '$\sqrt{()}$'. Such an entity would then be both a function and an object.

But Frege maintains that there can be no such entity because nothing can possibly be the reference of both a saturated and unsaturated expression.[3] His argument rests on a Leibnizian principle of identity of reference: two expressions have the same reference just in case they are intersubstitutable salva veritate. However, we are not yet in a position to state the argument in its most incisive form, and hence it shall be put off until some further relevant topics have been introduced. Suffice to say, for the present, that the categories of function and object are mutually exclusive. Frege now goes on to apply the term 'unsaturated' to functions as well as functional expressions. Presumably this metaphor is supposed to convey

---

[3] 'Reference' here has a very broad meaning whereby we can speak of the function ()-1 as being "the reference" of the functional expression '()-1', just as the number 2 is the reference of the value-expression '3-1'. Earlier, it was said that functional expressions do not "refer to" functions, and here 'refer' was being used in the narrow sense common in many recent works, that is, in the sense in which 'refer' applies paradigmatically to the relation between a singular term (on Frege's extended view, a kind of saturated expression) and an object or particular. Thus, though functions are "references" (in the broad sense), they are not "referred to" (in the narrow sense)--they do not stand to functional expressions in the way objects stand to singular terms or saturated expressions. This point will be discussed further later on.

the idea of an entity which is essentially one that logically relates or hooks together objects, but is not itself an object.

In the Grundgesetze, Frege was careful to distinguish functional expressions like '$2.()^3+()$' from those which "indefinitely indicate" a number, e.g. '$2.\underline{x}^3+\underline{x}$'. We may tend to confuse these, and hence wind up holding, as many mathematicians did in Frege's day, that functions are numbers of a sort, viz. "indefinite numbers". The expression '$2.\underline{x}^3+\underline{x}$', says Frege, is a value-expression, but it differs from ordinary value expressions like '2.1^3+1' in that '$\underline{x}$' does not refer to a particular number. Rather, '$\underline{x}$' is, as we would now say, a "variable"--it stands, roughly speaking, for some number or other. In using '$2.\underline{x}^3+\underline{x}$', the mathematician is asking us to consider any value of the function in question; which value we choose is unimportant.[4]

In Frege's time (and to a lesser extent nowadays), mathematicians, in using expressions like '$2.\underline{x}^3+\underline{x}$' to indefinitely indicate a number, also thought they were designating a function by that expression. They thus tended to think of functions as "indefinite numbers". This confusion is easily engendered by using 'function' as an absolute instead of relative term (which, however, is impossible to

---

[4] The distinction between a functional expression and an expression indefinitely indicating a number is quite similar (though not the same as) Quine's distinction between, respectively, a "predicate" and an "open sentence" in Methods of Logic (New York: Holt and Company, 1950).

avoid in a discussion such as this one). We must clearly recognize the difference between 'the function $2.()^3+()$' and '$2.\underline{x}^3+\underline{x}$ is a function of $\underline{x}$' or '2.1^3+1 is a function of 1'. If, with respect to these latter expressions, we ask: "Which function?", the answer would be: "The function $2.()^3+()$". But $2.()^3+()$ is not a function *of* anything--to say this would be nonsense: only a number or another "definite" object can be a function *of* a number. It is indeed tempting to move from '$2.\underline{x}^3+\underline{x}$ is a function of $\underline{x}$' to '$2.\underline{x}^3+\underline{x}$ is a function': Geach mentions an example in applied mathematics where this sort of confusion often occurs:

> A rod's length is a variable magnitude, and is also a function of the rod's temperature; have we not here, then a function that is a variable magnitude? No; that is a "fallacy of figure of speech". "The length is a function of the temperature" does not imply "The length is a function"; it means rather that there is some function for which the length is *that* function of the temperature; if the rod is r mm. long at t°C., then we always have for a certain function f(), r=f(t).[5]

The length, which will always be a definite quantity, is a function of the rod's temperature, which will also be a definite quantity. But the function consists in the systematic mapping of each denumerable quantity of the first kind with one of the second kind, and this is not itself a quantity (not even an "indefinite" one--whatever that could be).

---

[5] Anscombe and Geach, *Three Philosophers*, page 145.

In the Grundgesetze,[6] Frege uses small Greek letters, 'ξ', 'ζ', etc., to mark the gaps in functional expressions instead of parentheses. Thus, instead of '$2.()^3+()$', we would have '$2.\xi^3+\xi$'. The reason for this is of course that it is easier to express functions of two or more arguments; accordingly, 2.4^3+5 is a value of the function $2.\xi^3+\zeta$ for the ordered pair $\langle 4,5 \rangle$. This notation will be employed from now on.

Extension of the Theory of Functions to Ordinary Discourse

The sense-reference distinction cannot properly be introduced without first noting one of the ways in which Frege extended the theory of functions to expressions of ordinary discourse. The definite descriptions:

the capital of Great Britain
the capital of the U.S.A.

can be regarded as value-expressions derivable from the functional expression 'the capital of ξ' by inserting in the gap the argument-expressions 'Great Britain' and 'the U.S.A.' respectively. Thus we have a function which takes London and Washington as values for the arguments Great Britain and the U.S.A., and clearly the theory of functions has now been extended in such a way that not only numbers but any kind of entity which can be referred to by a singular term

---

[6] Gottlob Frege, The Basic Laws of Arithmetic, trans. and ed. by Montgomery Furth (Berkeley: University of California Press, 1964), page 34.

(in effect, anything which is not itself a function) will be a possible argument of a function.[7]

But now we must recognize that there will be well-formed expressions which fail to refer; that is, there is in fact no value of the function for a certain argument. For example, the value-expression 'the present king of France' has no reference. More importantly, we must also acknowledge certain well-formed expressions which do not simply fail to refer, but are, semantically considered, nonsense; e.g. 'the present king of 13', 'the capital of the U.S.A. +21', 'the author of Tuesday', etc. For on Frege's view, any object whatsoever can be regarded as a possible argument of any function; thus the number 13 is a possible argument of the function the present king of ξ, and Washington D.C. is a possible argument of $\xi+2$. Such expressions, though syntactically allowable, commit "category mistakes". This posed a problem for Frege since he was concerned with constructing an artificial symbolism as part of his researches into the foundations of mathematics. Frege wanted his system to be as rigorous as possible, and consequently felt that a function was not sharply defined or specified unless a definite value could be specified for any and all of its arguments. Hence, for cases like 'the present king of France' as well as 'the present king of 13', Frege laid it down that

---

[7] To be precise, a first-level function.

some reference must be assigned to these expressions--they both, he ruled, referred to the null class. However, we are not interested here in Frege's attempts to construct a rigorous symbolism, but mainly in his analysis of the logical features of ordinary discourse. And in this connection, Frege realized that in applying the theory of functions to natural languages, we must acquiesce in the fact that there will be functions which are left undefined for certain arguments either because as a matter of plain fact there is no value of that function or because of the implicit semantical restrictions contained in natural languages--restrictions which considerations of rigor and uniformity require us to eliminate when constructing an artificial symbolism.

Frege's Sense-Reference Distinction

Not all of the singular terms of ordinary discourse can be regarded as value-expressions. Those expressions commonly described as "proper names", e.g. 'Great Britain', are of course logically simple expressions; they are not formed by inserting an argument-expression in the gap of some functional expression. As such, they are the analogue in natural languages of numerals in elementary arithmetic.

However, there are certain problems concerning the analysis of identity sentences containing singular terms. It is natural to think of identity as a relationship of some sort, but what kinds of things does it relate? Two answers to this question occur immediately: it is a relation between

expressions, and it is a relation between what the expressions refer to. However, the latter alternative is unsatisfactory taken as it stands. For though identity is a relation holding between an object and itself (i.e. a reflexive relation), it should be obvious that this fact will not suffice to explicate a sentence like '$\underline{a}$=$\underline{b}$'. For then there would be no difference between saying '$\underline{a}$=$\underline{b}$' and '$\underline{a}$=$\underline{a}$', inasmuch as both express this reflexive relation. But it is clear that these identity sentences are distinct, for the latter is analytic and quite trivial while the former is synthetic and can be genuinely informative, as in 'The Morning Star is the Evening Star'.

Could we, then, hold identity to be a relation between the terms themselves? If so, then saying '$\underline{a}$=$\underline{b}$' would seem to amount to saying nothing more than that in a certain language, the expressions 'a' and 'b' refer to the same object. But this is not the same as saying that the objects $\underline{a}$ and $\underline{b}$ are identical, for on this view '$\underline{a}$=$\underline{b}$' becomes a concealed metalinguistic statement, and what is worse, a mere statement of linguistic convention; for it is an arbitrary matter as to what names we choose to bestow upon an object. What is left unaccounted for on this approach is the fact that an identity sentence can be empirically true without our already knowing it to be so; i.e. without our having stipulated that the two singular terms refer to the same object.

Frege's response to this puzzle is to draw a distinction for which he has since become quite famous. For any

kind of meaningful expression whatsoever, Frege maintains that it has a sense and (usually) a reference. We are already familiar with the latter notion--the reference of an expression is simply that object or function (as the case may be) which the expression refers to (if an object) or represents (if a function);[8] and an expression like 'the present king of France' has no reference, though it has a sense. The notion of "sense", like the notion of a function seems to be a primitive one and hence can only be explained in a somewhat circular manner, or else by means of examples. The sense of an expression is what we understand by it, or in one of its popular uses, what it "means". And Frege now says that though 'the Morning Star' and 'the Evening Star' have the same reference, they have different senses insofar as one can understand both expressions without thereby knowing that they designate the same object--as would be the case with 'the paternal grandfather of Smith' and 'the father of Smith's father'. In this latter case, the expressions have the same sense, and hence must have the same reference. Thus, we explain synthetic identity sentences like "The Morning Star is the Evening Star' by saying that the component singular terms have as a matter of fact the same reference though they differ in sense. Frege's official description of the sense-reference distinction is that a sign expresses its sense and stands for its reference.

---

[8] Thus, 'reference' is used here in the broad sense noted earlier, while 'refers to' is used in the narrow sense.

Frege applies the sense-reference distinction not only to singular terms (definite descriptions and proper names), but also to sentences and unsaturated expressions as well. Later on, we will examine each of these in some detail; for the present, however, some further remarks on the notion of the sense of a singular term are in order. It was said above that two such expressions which have the same reference differ in sense just in case one can understand both expressions without thereby knowing they have the same reference. But this rough, circular characterization of sense requires modification. For Frege definitely maintains that the expressions '2+2' and '4' have different senses but the same reference. However, this pair is strikingly different from his paradigm case of 'the Morning Star' and 'the Evening Star'. For we feel that if we understand '2', '4', and 'ξ+ζ', this is enough for us to know that '2+2' and '4' have the same reference, whereas, as Frege himself points out, this is not the case with 'the Morning Star' and 'the Evening Star'. On the other hand, there is a sense in which one can be said to understand '1,485' and '(594·85)/34' without knowing they refer to the same number, and perhaps this is what Frege means by 'sense' with respect to mathematical expressions. However, this will not work, at least as it stands; for if this is how we are to conceive the notion of sense, then it would seem that many pairs of expressions which would ordinarily be considered synonymous will no longer be synonymous, e.g. 'Smith's only paternal uncle' and 'the

only brother of the grandson of Smith's great-grandfather of whom Smith is a son.'

Nevertheless, it is only too clear why Frege wanted to say that '2+2' and '4' have different senses. For though the expressions have the same reference, the point of writing '2+2' is to exhibit that number as the value of a certain function, say ξ+2, for a given argument, and this information cannot be supplied by writing the simple numeral '4'. Therefore, these expressions differ in sense. Likewise, we could say on the same grounds that 'the largest railroad center in the U.S.' and 'the largest city in Illinois' differ in sense. Thus, we might use the notion of "exhibiting its reference as the value of a certain function for a certain argument" as a rough, preliminary criterion for distinguishing differences in sense.[9] Actually, though, we should have said, "a certain <u>set</u> of functions", at least for mathematical expressions since '2+2' can be regarded as a value-expression derivable from any of '2+ξ', 'ξ+2', and 'ξ+ζ'. However, 'the largest city in Illinois' can only be derived from 'the largest city in ξ'.

However, this proposed criterion will not work for two <u>ordinary</u> proper names or numerals, for in such cases neither expression exhibits its reference as the value of a function

---

[9] This idea seems to be implicit in Frege's rather obscure remarks concerning the "mode of presentation" of a sign in the opening passages of "On Sense and Reference."

since both are logically simple singular terms. In the case of 'Chicago' and 'the Windy City', we would have to specify the sense of these names in terms of some definite description, e.g. 'the largest city in Illinois', by which we customarily identify the reference. But it must be remembered that the same definite description must be used to give the sense of the two names, for it is an unfortunate consequence of Frege's theory that if one person takes the sense of 'Chicago' to be 'the largest city in Illinois' and another takes it to be 'the second largest city in the U.S.', then they are not using 'Chicago' in the same sense. Frege adds that "so long as the reference remains the same, such variations in sense may be tolerated, although they . . . ought not to occur in a perfect language."[10] However, ordinarily we would not wish to say of two persons who know how to use 'Chicago' correctly, that there is any variation in sense, and not just because of the common reference, but because both rely on roughly the same set of definite descriptions to identify Chicago.

In any case, the proposed criterion of "exhibiting its reference as the value of a function for an argument' is designed to apply primarily to the case of determining whether two singular terms have the same sense when neither term is logically simple.

---

[10] "On Sense and Reference", in Philosophical Works of Gottlob Frege, trans. and ed. by Peter Geach and Max Black (Oxford: Basil Blackwell, 1960), page 58 fn.

Now this criterion is clearly acceptable if we take it to mean:

P. For any two value expressions, E_1 and E_2, which have the same reference, if different functions are expressed by the functional expressions contained in E_1 and E_2, then E_1 and E_2 differ in sense.

Thus, a difference in function expressed entails a difference in sense, but we must not suppose that the converse of (P) also holds. A difference in sense does not entail a difference in function expressed; or in other words, sameness of function expressed does not entail sameness of sense. For Frege's criterion of identity for functions seems to be purely extensional; two functions are said to "coincide"[11] just in case they have the same values for the same arguments. If in fact the capital cities of all nations were also their largest cities, then the functional expressions 'the capital of ξ' and 'the largest city in ξ' would have the same function as their reference even though these expressions, and any pair of value-expressions derived from them by inserting the same argument expression in the gaps, differ in sense. This seems to me a rather objectionable consequence to construing functions extensionally. In any case, the expressions 'the capital of Austria' and 'the largest city in Austria' differ in sense, but if all and

---

[11] For reasons which cannot be discussed here, Frege is committed to holding that the notion of identity, strictly speaking, is senseless when applied to functions rather than objects. Even so, we can talk of the functional analogue of identity for objects, and for this relation, Frege generally uses the term 'coincide' (zusammenfallen).

only capitals are largest cities, we could not infer that they so differ on the ground that different functions are represented in the two expressions, for there would not be two distinct functions at all.

However, from (P) it follows of course that sameness of sense entails sameness of function expressed. Thus, because the definite descriptions 'Smith's only paternal uncle' and 'the only brother of the grandson of Smith's great-grandfather of whom Smith is a son' have the same sense, it follows that the functional expressions 'the only paternal uncle of ξ' and 'the only brother of the grandson of the great-grandfather of ξ of whom ξ is a son' must have the same function as their reference.

In the case of mathematical functions, the matter is more complex. '2+2' and '3+1' differ in sense because they stand for the same value of different functions, viz. $2+\xi$ and $\xi+1$, despite the fact that they can also be regarded as standing for the same value of the same function, viz. $\xi+\zeta$, for different arguments, $\langle 2,2 \rangle$ and $\langle 3,1 \rangle$ respectively. We shall have occasion later to note an instance of this fact in ordinary discourse.

Sentences and the Sense-Reference Distinction

We are now in a position to discuss Frege's application of the sense-reference distinction to sentences. The question that naturally arises first is: what reason is there

for supposing that sentences have references at all? We must now enquire into Frege's reasons for holding this rather counter-intuitive thesis.

Frege is probably most famous for extending the notion of a function so as to apply to complex expressions which admit of truth and falsity. The first such application was to equations in mathematics. Thus, '$\xi^2=1$' may now be regarded as a functional expression, for it represents the function revealed by the following list:

$$\begin{aligned} -1^2&=1 \\ 0^2&=1 \\ 1^2&=1 \\ 2^2&=1 \end{aligned}$$

Here, the function $\xi^2=1$ takes -1, 0, 1, and 2 as arguments. What, then, are the values for these arguments? The answer to this question is by no means obvious; indeed, we seem to be at a loss to say what they could be. With numerical functions, it was perfectly clear that they took numbers as values, for a value-expression like '2.2^3+2' obviously stands for the number 18. But what objects could the above four value-expressions stand for? Frege's reply is that we should note that the first and third equations are true, and the second and fourth are false, and his suggestion is that we regard the values of these functions as truth-values, and there are only two such values: the True and the False. Those value-expressions which are true all refer to the same object, viz. the **True**; and those that are false all refer to the object which Frege calls 'the False'. It should be

noted here that Frege uses 'truth-value' ambiguously: sometimes to mean the property of being true or false (as the case may be) which any proposition will have, and sometimes to mean the objects the True and the False. The former sense is of course its common use nowadays in logic texts, and the latter is a special, theory-laden sense used only by those who agree with Frege that all true expressions refer to the same object, and similarly for all false expressions. It may seem odd to hold that the expression '$2^2=4$', '$2>1$', and '$2^4=4^2$' all name the same object, but we must put off for the present any further discussion of this rather peculiar view.

Equations are not the only expressions which admit of truth and falsity; sentences in ordinary discourse do so too. And Frege correspondingly treated these expressions as standing for truth-values, and in so doing, became the first major philosopher to handle the notion of predication in terms of the argument-function relationship. Consider the sentences:

Brutus killed Caesar.
Cassius killed Caesar.

Both of these can be regarded as value-expressions derived from the functional expression 'ξ killed Caesar' by inserting in the gap signified by 'ξ' the argument-expressions 'Brutus' and 'Cassius' respectively. Since both sentences are true, we may say that the function ξ killed Caesar has the value True for the arguments Brutus and Cassius, and it

has the value False for, say, the argument, Antony--so that 'Antony killed Caesar' is false and therefore designates the False.[12] Thus, singular sentences are broken up into argument-expressions (i.e. singular terms including definite descriptions) and functional expressions which, because they are functional, necessarily contain at least one gap. Thus, the functional expression embodies the form or structure of a singular subject-predicate sentence, and it does this because the corresponding function maps the argument (object referred to by the subject term) onto the values True and False. And for this reason Frege rejected completely the old, Aristotelian analysis of the proposition into subject-and-predicate-separated-by-the-copula, and at one point remarks that the copula is logically superfluous. For the structure of 'Socrates is wise' is not that of '___ is ....', but rather '___ is wise'. The copula is only an inessential part of a larger, necessarily unsaturated, expression--that it is not essential can easily be seen from that fact that we have many such unsaturated expressions where the copula does not and need not occur, as in '___ smokes'. What was traditionally called the "predicate" necessarily carries with it the form of the sentences in which it can occur;

---

[12] For reasons of clarity, expressions designating non-mathematical functions will be underlined, e.g. ξ killed Caesar. We cannot use quotes, for obviously we would then be talking about the functional expression instead of the function.

insofar as it is a functional expression, the form is not something which can be distinguished from the predicate expression. And as we saw earlier, a functional expression is itself a linguistic form, it is not a kind of well-formed expression. The proper way of expressing predicate terms is not simply 'is wise', but 'ξ is wise'--for it is essential to a predicate expression that it be unsaturated. Those functions designated by functional expressions containing only one distinct gap (although there can be more than one occurrence of 'ξ') which, upon placing in the gap a singular term, yields a simple, singular subject-predicate sentence are called concepts by Frege. Thus, concepts take the True and the False as their only values. For example, Socrates and Fido are arguments of the concept ξ is wise, and the value for the first is the True and for the second the False.

But the sentences 'Cassius killed Caesar' and 'Brutus killed Caesar' can also be regarded as value-expressions of another functional expression as well. This function is ξ killed ζ, and here we have a functional expression which contains two gaps in order that different singular terms may be inserted into them. Those functions designated by functional expressions containing two or more distinct gaps which, upon insertion of singular terms into these gaps, yield as a value-expression a simple, singular relational sentence are called relations by Frege. Thus, relations, like concepts, take the True and the False as values, but,

unlike concepts, take on ordered sets of objects as arguments. ξ killed ζ and ξ is between ζ and η are both relations (the first a dyadic, the second a triadic relation). The function ξ killed ζ takes the True as its value for each of the ordered pairs ⟨Cassius, Caesar⟩ and ⟨Brutus, Caesar⟩ as arguments, but takes the False for ⟨Antony, Caesar⟩ and ⟨Caesar, Brutus⟩.

We must be careful to distinguish the function ξ killed ξ from ξ killed ζ: the former does not admit ordered pairs as arguments while the latter does (including those pairs where the paired objects are the same). Thus, ⟨Hemingway, Hemingway⟩ is an argument of the latter function as is ⟨Brutus, Caesar⟩, and the value for both is the True. But only the objects Hemingway, Brutus, etc. are properly speaking arguments of the former.

Philosophers have often balked at Frege's notion of sentences as referring to or designating truth-values--or of referring to or designating anything. There are many reasons for balking at this, but the more sophisticated of these (e.g. those brought forth by Wittgenstein) cannot be profitably discussed at this stage. But it can be said here that to object simply that sentences are not names on intuitive grounds will not do; for we have seen that Frege was led to this view as a consequence of holding that sentences are value-expressions derivable from functional expressions by the insertion of singular terms. Thus, the doctrine of sentences as names is a consequence of the more fundamental

view that language is basically functional in nature together with the claim that logical form is to be explained by reference to those essentially incomplete entities called functions.[13] And presumably Frege would say to such a critic that if he holds the functional view of language, then it is up to him to try and devise a theory which preserves this view and still manages to avoid bringing in the Fregean ontology of objects and functions, and the resultant doctrine of sentences (value-expressions of a certain sort) as names (the best-known attempt to do this was of course Wittgenstein's logical atomism). At any rate, it should at least be clear that we should not object that sentences are not names merely on intuitive grounds and let it go at that, and further, that Frege did not simply assume that sentences are names, nor did he arrive at this conclusion on the basis of a simple-minded "intension-extension" assumption.

In summary, since there is much plausibility in treating the notions of concept and relation as special cases of the notion of a function, it is also quite plausible to regard sentences as names of objects which are the values of

---

[13] There are of course, two separate theses: the first is a logical theory concerning how logical structure is to be analyzed (an alternative would be the Aristotelian subject-predicate doctrine), and the second is an ontological theory about what sort of ontology needs to be presupposed in order to account for this analysis (an alternative here would be the ontology of the Tractatus). Clearly what view is held on this latter level depends on the view held on the former.

those functions. But now we must ask why Frege chose the objects he did; for on purely intuitive grounds, it certainly seems objectionable to have to introduce such unfamiliar and mysterious-sounding entities as "the True" and "the False".

As a first step toward answering this question, it must be pointed out that the expressions 'the True' and 'the False' have so far been understood only as names or labels of two distinct, and in some way (unspecified as yet) identifiable objects. We have not yet said which objects, or what sort of objects they are; and it is still an open question as to whether these objects are numbers, nations, people, etc. or whether they are distinct from all these. That is to say, we could if we wanted to, at this point, take 'the True' and 'the False' as respectively referring to the numbers 1 and 0; in other words, we could if we wished simply take the True to be identical with the number 1 and the False to be identical with 0. Or, if our whims were otherwise, we could take W.C. Fields as the True and Paul Tillich as the False. But Frege in fact regards the True and the False as totally unique objects not to be identified with familiar objects like numbers and persons. He is never quite clear about his reasons for this, and indeed there seems to be no necessity for doing so. Apparently, Frege's only reason for taking the True and the False to be unique objects, not referred to by any other singular term which already has a fixed use (e.g. '1', 'DeGaulle') is that there is no good reason for singling out one pair of numbers,

nations, people, etc. rather than any of the others. And since we must have objects which are the references of sentences, it is intuitively more plausible to select two previously unfamiliar objects. Similarly, it would be natural to suppose that these objects are of a non-empirical sort, for he clearly does not balk at introducing such objects into his ontology (he has already admitted numbers), and remarks that it is "mere prejudice" to suppose that the only objects are those which can in principle be empirically observed. And of course, though we do not customarily regard sentences of ordinary discourse as referring to anything, it is nevertheless true that if we must so regard them, our intuitions would naturally favor unique objects--it is just too much to suppose that 'Jones is stupid' and 'The world is round' both refer to the number 1 or to W.C. Fields.

Nevertheless, it must again be emphasized that it does not really matter which objects the True and the False are--all that matters is that there be two distinguishable objects. The truth-conditions of a given value-expression have nothing whatsoever to do with what objects we take the True and the False to be--we can determine whether a sentence is true or false independently of what these objects are.[14] For example, we know that any value-expression

---

[14] There are trivial exceptions to this. If, for example, we take the True to be 1 and the False to be 0, then '$1=(2^4=4^2)$' is true or designates the True. But this could be determined only by first knowing that '1' also designates the True. If a different object is taken as the True, then the sentence will be false, for '1' and '$2^4=4^2$' will now designate different objects.

derived from '$\xi=\zeta$' will be true (and hence refer to the True, whatever it happens to be) just in case the expressions inserted for the two gaps both stand for the same object. Thus we can say that since '4^2' and '2^4' refer to the same number, the expression '$4^2=2^4$' refers to the True.

So far, then, we have seen that the results of Frege's treating singular sentences of ordinary language in functional terms (as a kind of value-expression) are: first, that sentences must be taken to stand for distinguishable and identifiable objects; second, that there be exactly two such objects--one the reference of all true sentences, the other the reference of all false ones; and third, that these objects are referred to by sentences on the one hand, and by certain special singular terms on the other (e.g. 'the True', 'the False', etc.), but are not referred to by any other singular terms which already have a fixed, definite use (i.e. they are totally unique objects). It should be emphasized that the second and third points are independent of the first, insofar as one can agree with Frege on the first but hold different views with respect to the other two. We have already discussed the first and third points, so it remains for us now to examine the second. This will not be easy, for Frege's remarks in "Function and Concept" and "On Sense and Reference" indicate that he failed to clearly distinguish the second and third points.

One seemingly plausible alternative to Frege's theory that sentences refer to truth-values is that they refer to

propositions. But this view, according to Frege, can easily be shown to be wrong. Consider the sentences, 'Mark Twain wrote Tom Sawyer' and 'Samuel Clemens wrote Tom Sawyer'; it is clear that the references of the component expressions of the first sentence are identical with the references of the component expressions of the second, for 'Mark Twain' and 'Samuel Clemens' stand for the same man, and of course in both cases the concept ξ wrote Tom Sawyer is the reference of the predicate. Now if the references of the component terms are the same in both sentences, the references of the two sentences must themselves be the same. But if the reference of a sentence were the proposition expressed, it would follow that the references of the two sentences in question must not be identical, since the two sentences obviously express different propositions; for one could understand the expressions 'Mark Twain' and 'Samuel Clemens' and still not know that they designate the same man. Thus, as long as we maintain that the reference of a complex expression (like a sentence) is determined by the references of its component expressions, we cannot consistently maintain that propositions (as they are usually understood) are the references of sentences. This principle concerning the reference of a complex expression plays an important role in Frege's philosophy of logic, and we shall have occasion to return to it. Frege apparently assumed it to be self-evident, which it seems to be provided that one is prepared to grant that complex expressions like sentences have references at all.

We have seen that Frege's own view on the matter is that those objects he calls "truth-values" (the True and the False) are the references of sentences. And this position has been spelled out as follows: there are two distinct, identifiable objects, one of which is the reference of all true sentences, the other of all false ones. But we have not as yet:

a. given anything in the way of an argument for this position,

b. given a criterion of identity for the notion of being a truth-value.

Concerning (b), we have not as yet said what it is for an object to be a truth-value insofar as we have not yet specified the conditions under which two sentences refer to the same truth-value. However, it should by now be clear as to what these conditions are:

Two sentences, p and q, refer to the same truth-value just in case the material biconditional ⌜p if and only if q⌝ is true.

Someone might here object that this is circular on the grounds that a biconditional is itself defined as a statement which is true just in case both components have the same truth-value. Thus it would seem that Frege is saying that p and q stand for the same truth-value just in case they have the same truth-value. But this objection rests on a misunderstanding. For as we saw earlier, Frege used 'truth-value' in two different ways: sometimes to mean the circumstance of being true or false and sometimes to mean the objects designated by sentences--objects which he has

merely labeled 'the True', 'the False', and, taken together, 'truth-values'. Thus there is nothing wrong in saying p and q refer to the same truth-value just in case they have the same truth-value, for in the first occurrence, 'truth-value' means the objects which sentences refer to, and in the second, the circumstance of being true or false. Frege, then, is just explaining one sense of 'truth-value' in terms of another, just as one would explain one sense of 'baseball' (viz. a certain kind of game) to someone by using a different sense (viz. a certain spherical object covered by horsehide). So there is no circularity here (though we can certainly take Frege to task for using the same expression in such entirely different ways), and Frege's criterion of identity is satisfactory after all.

Thus, if we grant that truth-values, as defined above, are what sentences refer to, then since sentences can be only true or false (have only two truth-values in the "circumstance" sense), there can be only two truth-values (in the "object" sense)--one the reference of all true sentences, the other the reference of all false ones. But what reason is there to suppose that truth-values as thus defined are what sentences designate? In other words, we must now return to (a) and give Frege's reasons for regarding the True and the False as the references of sentences.

We saw earlier that Frege's reason for rejecting propositions as the references of sentences was that two sentences such as 'Mark Twain wrote Tom Sawyer' and 'Samuel Clemens

wrote Tom Sawyer' express different propositions, but insofar as the references of the component expressions of each sentence are the same, the references of the sentences must themselves be identical. What, then, could be the object which is the reference of both? Presumably, to answer this question we would have to point to some characteristic or feature that both sentences have in common. Now we know that when a component expression of a sentence is replaced by another term having the same reference, the reference of the new sentence must be the same as the reference of the original one. Now we can ask: what, if anything, remains the same when such replacement takes place? Obviously, the most striking feature that remains unchanged is the truth-value--if the original sentence is true, the new one will likewise be true, and clearly any further sentence we form in this manner will also be true. Thus, Frege is led to seek a common object which is the reference of all true sentences --at least, this is the most natural way to interpret the following rather obscure passages from "On Sense and Reference" (it should be noticed in passing how Frege has confused the senses of 'truth-value' in these selections):

> We have seen that the reference of a sentence may always be sought, whenever the reference of its components is involved; and that this is the case when and only when we are inquiring after the truth-value.
>
> We are therefore driven into accepting the truth-value of a sentence as constituting its reference. By the truth-value of a sentence I understand the circumstance that it is true or false....Every declarative sentence concerned with the reference of its words is therefore

> to be regarded as a proper name, and its reference, if it has one, is either the True or the False. These two objects are recognized, if only implicitly, by everybody who judges something to be true--and so even by a sceptic.
>
> ...If our supposition that the reference of a sentence is its truth-value is correct, the latter must remain unchanged when a part of the sentence is replaced by an expression having the same reference. And this is in fact the case. Leibniz gives the definition: Eadem sunt, quae sibi mutuo substitui possunt, salva veritate. What else but the truth-value could be found, that belongs quite generally to every sentence if the reference of its components is relevant, and remains unchanged by substitutions of the kind in question? 15

These are the only passages where I can find any explicit argument for the claim that the True and the False are the references of sentences. And it should take little reflection to realize that if Frege's argument comes to nothing more than this, then he is guilty of a gross non sequitur. For Frege, using the principle that the reference of a complex expression is determined by the reference of its components, has correctly inferred that propositions are not the references of sentences. But now, going on the assumption that the reference of a sentence must be the common reference of sentences whose component references are the same, he infers that since the truth or falsity remains unchanged when we replace 'Mark Twain' in 'Mark Twain wrote Tom Sawyer' by 'Samuel Clemens', then truth and falsity are

---

15 "On Sense and Reference" in Philosophical Works, pages 63-64.

necessarily the common, unchanging features he is looking for; and so he is led to postulate an object which is the reference of all true sentences and one which is the reference of all false ones. But this step is clearly unwarranted; it is correct that 'Mark Twain wrote Tom Sawyer' and 'Samuel Clemens wrote Tom Sawyer' have the same reference, and it may be inferred that, say, 'The author of Huckleberry Finn wrote Tom Sawyer' designates that same object, and it can also be concluded that this latter sentence, like the previous ones, is true. But what does not follow is that the object which is the reference of these sentences is the reference of all true sentences whatsoever. For there is nothing in the foregoing argument which in any way requires us to say that an object which is the reference of one true sentence must also be the reference of another true sentence when the references of the component expressions of the one sentence are different from the references of the component expressions of the other; there is no reason why, for example, we are forced to conclude that 'Mark Twain wrote Tom Sawyer' and 'Chicago is west of New York' should refer to the same object simply because both are true.

Again, it is clear that when 'Samuel Clemens' is substituted for 'Mark Twain' in 'Mark Twain wrote Tom Sawyer', the property of being true remains; and it is also clear that the reference of the resulting sentence should be the same as that of the original one. But what reason is there to suppose that these two facts are necessarily connected

with one another? Why, specifically, must we then infer that the reference we are seeking must be the reference of all true sentences? There is, it appears, no reason for this inference, and indeed, there seem to be grounds for inferring just the reverse. For if the references of the component terms of two sentences are different, is it not more reasonable to assume that the references of the two sentences themselves are also distinct? This at least suggests that the fact of the property of being true remaining when the sort of substitutions Frege has in mind occur, is not relevant to the matter of determining what sort of object must be the common reference of the original sentence obtained by the substitution. What is relevant seems intuitively to be the fact that there is a common predicate occurring in each of the sentences obtained through such substitutions; and since the subject expressions of these sentences designate the same object, it seems more plausible to suppose that there is a common reference for these sentences and only these. This object might be called a "fact", since the ordinary use of 'fact' resembles the use proposed here; thus the common reference of:

Mark Twain wrote Tom Sawyer
Samuel Clemens wrote Tom Sawyer
The author of Huckleberry Finn wrote Tom Sawyer
etc.

is a fact, and this fact or object is distinct from the fact or the object which is the reference of:

Chicago is west of New York
The Windy City is west of New York
The largest city in Illinois is west of New York
etc.

In spite of all this, we must find some way of establishing the following claim:

1. Two complex expressions may have the same reference even though the references of their component expressions are different.

For it can easily be seen that we cannot remain content merely to hold that:

2. If the references of the component expressions of two complex expressions are the same, then the references of the complex expressions must themselves be the same.[16]

Now (2) is a weaker claim than (1), and it is clear that Frege wants to hold both. However, unless we can establish (1), there is nothing to prevent us from holding (2) and its converse instead of (2) and (1), and to make things worse, there seemed to be intuitive grounds favoring the converse of (2) over (1). For holding (2) and its converse is to hold that "facts" are the references of sentences; holding (2) and (1), while it does not entail that the True and the False are the references of sentences, certainly makes it by far the most plausible view (since its only serious competitor, the "facts" theory, is ruled out). Now a moment's reflection will convince us that there are grave

---

[16] This is not to say that 'The cat eats the rat' and 'The rat eats the cat' must have the same truth-value. Rather, the former has the value, true, for the ordered pair ⟨the cat, the rat⟩, whereas the latter has the value, false, for ⟨the rat, the cat⟩.

difficulties in holding both (2) and its converse. For the converse of (2) is clearly false with respect to mathematical functions, and if it is false for them, it seems to follow that it is false for concepts and relations as well. Consider, for example, the function $\xi+\zeta$; this function has the value 4 for the argument $\langle 2,2 \rangle$ and for the argument $\langle 1,3 \rangle$. Thus, the value-expressions '2+2' and '1+3' have the same reference, viz. 4, even though the references of their component expressions ('2' in the first case, '1' and '3' in the second) are different. This clearly shows us that (1) is true and the converse of (2) false for mathematical functions, and the question to which we must now address ourselves is: can we find similar examples with concepts and relations?

The answer to this question can be clearly seen by examining an argument employed by Church.[17] From the functional expression 'ξ is ζ' (or '$\xi=\zeta$'), we may derive two value expressions which are synonymous, yet the component singular terms of one of them have a reference different from that of the component terms of the other. These sentences are:

A. Scott is the man who wrote 29 Waverly novels.

B. The number of Waverly novels written by Scott is 29.

---

[17] Alonzo Church, Introduction to Mathematical Logic (Princeton: Princeton University Press, 1956), pages 24-25.

(A) and (B) clearly have the same sense, but if so, they must have the same reference as well. Moreover, it is obvious that the component references of (A) and (B) are different, for in (A), the two singular terms refer to Scott, whereas in (B), they refer to the number 29. Hence, (A) and (B) stand for the same value of the function ξ is ζ for the distinct arguments ⟨Scott, the man who wrote 29 Waverly novels⟩ and ⟨the number of Waverly novels written by Scott, 29⟩. Therefore, it follows that two complex expressions can have the same reference even though the references of their component expressions differ; we must, then, reject the "facts" theory.

But must we then hold Frege's "truth-value" theory? We must if we accept (as Frege does) Leibniz's identity principle, that is:

1. a=b if the expressions 'a' and 'b' are intersubstitutable salva veritate in any complex expression.[18]

Now since a is the reference of 'a' and b the reference of 'b', it is clear that (1) gives a principle of identity of reference. Further, we also know that:

2. Sentences have references.

And since truth-functional sentences are ipso facto complex expressions containing other sentences as components, we may derive the following principle of identity of reference for sentences:

---

[18] Excluding, of course, intensional and quoted contexts.

3. Two sentences, S and S', have the same reference if S and S' are intersubstitutable *salva veritate* in any truth-functional sentence.

But now let us ask whether it follows from (3) that there is exactly one object which is designated by all true sentences, and another which is designated by all false ones. Now we know from our earlier discussion that the reference of a truth-functional sentence like 'Either Mark Twain wrote *Tom Sawyer* or O. Henry did' is determined by the references of 'Mark Twain wrote *Tom Sawyer*' and 'O. Henry wrote *Tom Sawyer*'. Further, we know according to (3) that a given sentence, *S*, will have the same reference as 'Mark Twain wrote *Tom Sawyer*' if it can be substituted for it in complex sentences like the above disjunction and the resulting sentence has the same truth-value as the original. Now the original disjunction is true owing to the truth of 'Mark Twain wrote *Tom Sawyer*', and it is clear that if we substitute *any* other true sentence, the resultant disjunction will be likewise true; and it is also clear that this holds for any truth-functional sentence we pick. Since truth-value is preserved through such substitution, we must then admit that this true sentence which was substituted--suppose it was 'Eisenhower is bald'--has the same reference as 'Mark Twain wrote *Tom Sawyer*'. The question now is: what could be the common reference of 'Eisenhower is bald' and 'Mark Twain wrote *Tom Sawyer*'? The most striking feature of such otherwise disparate sentences is that they are both *true*, and thus we are led to the conclusion that there is just one object

which is the reference of all true sentences; and of course a similar argument can be given to show that there is just one object which is the reference of all false ones.[19]

Unsaturated Expressions and the Sense-Reference Distinction

Before passing on to the sense of a sentence, some brief remarks on the sense-reference distinction as applied to unsaturated expressions are called for. The reference of such an expression is of course a function; thus, 'ξ is wise' (hereafter, a concept expression) has as its reference the function ξ is wise, and 'ξ killed ζ' (a relation expression) has as its reference ξ killed ζ. So concepts, not "classes" or "extensions" or any other kind of object, are the references of concept expressions. Frege says very little about the sense of such an expression, and offers no criterion for determining when two such expressions have the same sense. He is mainly interested in their references and in certain objects correlated with them (viz. " "value-ranges", or more specifically in the case of concepts,

---

[19] The basic outline of this argument has been presented by Geach in Three Philosophers, pages 152-153. Geach attributes the argument to Frege though I can find no explicit statement of it in his works. It can, however, easily be read into certain passages of "On Sense and Reference" and "Function and Concept" (see Philosophical Works).

"extensions").[20]

Since Frege's criterion of identity for functions is extensional, it is clear that two concept expressions can differ in sense yet have the same reference. Yet even if functions were conceived non-extensionally, there would still be a point to distinguishing a sense and a reference for these expressions. For though 'ξ live in Alaska' and 'ξ lives in the largest state of the Union' have the same reference (according to the principle of identity of reference for complex expressions--of which this would be a rather unique sort), they would still differ in sense inasmuch as the component singular terms differ in sense.

We may now characterize a concept more precisely as a function of one argument which takes only truth-values as values. It must again be emphasized that though Frege regards concepts and relations as "references" of unsaturated expressions, they are not "referred to" by these expressions in the sense in which singular terms and sentences refer to objects. Frege was committed to the existence of functions not because he thought functional expressions had to "refer" à la singular terms, but because it is necessary to formulate

---

[20] Concept expressions must not be identified with what are often called "propositional functions" in elementary logic texts. The notion of a propositional function, as it was expounded for example by Russell, tends to obscure the distinction between concept expressions (such as 'ξ is a man') and what Quine calls an "open sentence" or what Frege might call an expression indefinitely indicating a truth-value (such as 'x is a man').

higher order statements of the form 'There is a function, f(), such that ...'. To say that a function or concept is "the reference" of a concept expression is only to say that there is an entity which corresponds to a set of concept expressions which become value-expressions having the same truth-value upon the insertion of the same argument-expression.

We can now briefly state Frege's argument that the categories of function and object are mutually exclusive. The concept expression 'ξ is a dragon' and the singular term 'the concept dragon' cannot have the same reference if the Leibniz principle is accepted. For the former expression cannot be substituted salva veritate (or even salva meaning) in the sentence 'The concept dragon is uninstantiated.' The concept expression, in Frege's words, is "essentially predicative", and hence cannot meaningfully occupy the subject position.[21] By the same token, 'the concept dragon' cannot meaningfully replace the concept expression as it occurs in 'Faffner is a dragon.' There are several important problems arising from this conclusion which obviously have a great bearing on the notion of predication. However, space does not permit any further elaboration here.[22]

---

[21] Accordingly, Frege is forced to say that singular terms like 'the concept dragon' refer to objects which "go proxy" for functions. See his "Concept and Object" in Philosophical Works.

[22] See, for example, G.E.M. Anscombe, An Introduction to Wittgenstein's Tractatus (London: Hutchinson University Library, 1959), chapter 6; and Peter F. Strawson, Individuals (London: Methuen and Co., 1959), chapter 5.

Truth-Functions and Concepts

One of Frege's main reasons for drawing the sense-reference distinction with respect to sentences was to give a satisfactory account of *oratio obliqua* constructions. However, we must pass over this important topic and concentrate instead on certain other aspects of his theory. It is necessary first to discuss a rather special concept introduced by Frege in "Function and Concept".

We saw that concepts take only truth-values as values, but in Frege's formal system they can take *any* object as an argument; for he would have nothing to do with the somewhat dubious procedure of stipulating "ranges of significance" for functions (although he did hold a simple theory of types for functions). The concept to be introduced now is represented by the expression '—— ξ'. It should be noticed immediately that this expression, unlike the concept expressions discussed previously, is not a linguistic form or pattern of English or any other natural language, and consequently, Frege gives us an explicit definition of it. '——ξ' represents that function whose value is the True when the True is taken as argument, and the False for all other arguments. Thus, if we insert in the gap an expression referring to the True, the result will be a value expression which likewise refers to the True. Accordingly, '——1+3=4' refers to the True, whereas '——1+3=5' and '——4' refer to the False. Although '1+3=5' and '——1+3=5' both refer to the False, only '——4', and not '4', refers to the False. This

should serve to emphasize that the horizontal line (Frege calls it simply "the horizontal") together with the numeral '4' comprise a value expression which exhibits the False as the value of a certain function for the argument 4.

The natural question to be raised here is why Frege bothers to introduce this function at all. One important reason is that it is helpful for the introduction of those functions which have since become known as truth-functions. Now a truth-function is usually thought of as one whose arguments are only truth-values--no other objects are to be considered as possible arguments. In order to preclude these other objects from occurring as arguments, it must be guaranteed that the argument expressions inserted into the gaps of expressions representing truth-functions refer only to the True and the False. And we have just seen that no matter what an expression, S, stands for, ——S will stand for a truth-value. Thus Frege tries to develop his theory of truth-functions in such a way that for any functional expression representing a truth-function, expressions of the form '——x' are primarily to be regarded as the argument expressions insertable in the gaps.

The qualification "primarily" in the preceding remark is essential. For Frege actually defines truth-functions in such a way that they, like ——ξ, are concepts, i.e. they can take any object as a possible argument. But of course the only arguments that are of any interest are truth-values, and the argument-expressions with which Frege is concerned

are those of the form '——x', since these expressions always refer to a truth-value. For example, Frege introduces negation as follows:

> The next simplest function...is the one whose value is the False for just those arguments for which the value of ——ξ is the True, and conversely, is the True for the arguments for which the value of ——ξ is the False. I symbolize it thus:
>
> —⊢—ξ
>
> and here I call the little vertical stroke the stroke of negation. I conceive of this as a function with the argument ——x:
>
> (—⊢— x) = (—⊢— (——— x))
>
> where I imagine the two horizontal strokes to be fused together. But we also have:
>
> (———(—⊢— x))=(—⊢— x)
>
> since the value of —⊢— ξ is always a truth-value. I thus regard the bits of the stroke in '—⊢— x' to the right and to the left of the stroke of negation as horizontals... 23

From the definition of '—⊢— ξ', it follows that —⊢— ξ is a concept which all objects except the True fall under. Thus, it is not a truth-function in the strict sense of a mapping of truth-values (only) onto truth-values. However,

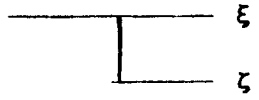
---

23 "Function and Concept" in Philosophical Works, pages 34-35. In the actual passage, Frege uses the letter 'x' throughout instead of 'ξ' where I have used it. In the Grundgesetze, Frege distinguished between variables and letters indicating the gaps in functional expressions. For the former, he uses the small letters 'x', 'y', etc., and for the latter, the small Greek letters 'ξ', 'ζ', etc. Although he was aware of the distinction when he wrote "Function and Concept", Frege uses 'x' and 'y' to perform both tasks, and in the interest of clarity, I have restricted them to serving as variables and adopted 'ξ', 'ζ', etc. from the Grundgesetze for the purpose of indicating gaps.

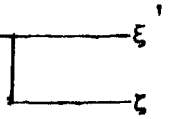
he also speaks of ———x as an argument instead of merely x, and goes on to affirm the general proposition:

¬ x=(¬(——— x))

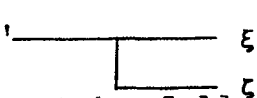
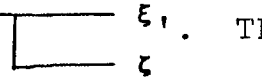
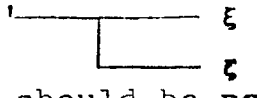
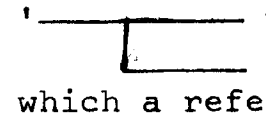
which says that for any argument, x, the functions ¬ ξ and ¬(——— ξ) will have the same value for that argument. This means that it makes no difference whether we regard the value-expression '¬(——— S)' as derived from '¬ ξ' or '¬(——— ξ)', since the functions are in fact the same. Hence, the expression '¬ S'--which refers to the same truth-value as '¬(———S)'--can be regarded either as derived from '¬———ξ' with 'S' being inserted in the gap, or as derived from '¬ ξ' with '———S' inserted in the gap, and its horizontal then "fused" with the '¬———'. Thus, ¬ ξ, though a concept by definition, can in practice be worked with as a genuine truth-function.

Frege introduces and treats the other truth-functions in the same way. The function known nowadays as "material implication" is symbolized: -- 

Its value is the False if the argument expression inserted for 'ζ' refers to the True and argument expression inserted for 'ξ' refers to some object other than the True. In all other cases, its value is the True. As with negation, we can always regard truth-values as the arguments of this function.

Functional expressions like '¬ ξ' and ' 

'

exist of course only in the formal language of the Grundgesetze. Unlike the concept and relation expressions discussed earlier, they are not linguistic forms of English or some other natural language. However, English does have the form 'if ζ then ξ' which can to some extent be explicated in terms of the function  . This is, certain uses of the unsaturated expression 'if ζ thenξ' are analogous to '  ξ' in Frege's system. But the fact that it cannot be fully explicated along these lines shows that it cannot be freely interchanged with '  ξ'. The objections to explicating 'if ζ then ξ' solely in terms of '  ξ ' are too well known to be repeated here, but it should be pointed out that there is a further disparity between these expressions. For Frege's language, unlike natural ones, contains no stipulations concerning what kinds of expressions may properly be inserted in the gaps; so that '  the Taj Mahal, is a well-formed expression to which a reference must be assigned, whereas 'if 4 then the Taj Mahal' is nonsense. But we can safely ignore this disparity if we regard Frege's work in "Function and Concept" not just as offering us an artificial symbolism which from a logical point of view is free of the deficiencies of natural languages, but rather as an analysis of the logical form of these languages. And we can do this regardless of the fact that natural languages do not require that every expression have a reference, and also that they, unlike Frege's system, implicitly contain certain semantical prohibitions.

Waiving this disparity, then, it would certainly seem that the function —⊢— ξ provides a satisfactory account of negation in English. This issue will be more extensively discussed later, but nonetheless there is no reason why we should not regard '—⊢— ξ' and 'it is not the case that ξ' as representing the same function. From this, it would follow, however, that the grammatical form of negative sentences in English is misleading. For in 'Socrates is not wise' it looks as if negation attaches to the predicate expression instead of the sentence as a whole. On Frege's view, the sentence is not to be analyzed into 'Socrates' and 'ξ is not wise', because this latter expression does not represent a concept--there are no "negative concepts". Negation is a truth-function, and as such 'Socrates is not wise' must be a value expression derived from the argument expression 'Socrates is wise'.[24] Thus, as far as logical form is concerned, the sentence is more properly written 'It is not the case that Socrates is wise' or in Frege's notation '—⊢— Socrates is wise' is an argument expression

---

[24] Strictly speaking, 'ξ is not wise' is not a legitimate, logical unit of analysis like 'ξ is wise'--it does not represent a concept. We may derive 'Socrates is not wise' as a value-expression from 'Socrates is wise', and in the same way 'x̄ is not wise' can be derived from 'x̄ is wise'. But we cannot get by this procedure 'ξ is not wise' from 'ξ is wise' for the latter expression cannot sensibly fill the gap of 'it is not the case that ξ'. Only a well-formed saturated expression can do this, and 'ξ is not wise' does not satisfy this requirement--it cannot be used as an argument-expression to fill the gap of a first-level functional expression like 'it is not the case that ξ'.

which has been inserted in the gap of 'it is not the case that ξ' or '$\neg\xi$'.

The Sense of a Sentence

Although Frege is not much concerned with the notion of sense as it applies to singular terms and functional expressions, he devotes a great deal of space to the sense of a sentence. He calls the sense of a declarative sentence a thought, and as in the case of reference, the sense of a sentence is determined by the senses of its component expressions. But there are also cases where two sentences express the same thought but whose component expressions differ in sense; for example:

Scott is the man who wrote 29 Waverly novels, and

The number of Waverly novels written by Scott is 29.

Many of Frege's commentators prefer the term 'proposition' to 'thought' even though the German expression is 'Gedanke'. One reason for this is the fact that Frege conceives of thoughts not as subjective, mental entities like ideas, but as public ones capable of being grasped or apprehended by any number of people. Frege spends a good deal of time arguing this view--more so than would be required today--but we must remember that in Frege's time, most logicians were idealists of one sort or another. However, I will continue to use 'thought', for it has certain advantages for my purpost that 'proposition' does not share.

We have seen that two sentences may express the same thought even though the senses of their component expressions may differ. It is also true that the same thought can be expressed by sentences of logically different kinds, and that a thought can have more than one correct logical analysis. The thought expressed by the singular sentence 'I was stopped by the door of a tomb' is also expressed by the particular, affirmative sentence 'Some tomb is one of whose door I was stopped.'

The thought expressed by a sentence parallels to some extent what was later called the "cognitive meaning" or "descriptive content" of a sentence, but thoughts are also regarded by Frege as the bearers of truth and falsity. Thoughts are true or false, sentences refer to the True and the False. In his essay, "The Thought", Frege says that words like 'still' and 'already' have no bearing on the thought expressed by any sentence in which they occur--hence the analogy with cognitive meaning.

> With the sentence 'Alfred has still not come', one really says 'Alfred has not come' and at the same time hints that his arrival is expected, but only hinted. It cannot be said that since Alfred's arrival is not expected, the sense of the sentence is false. The word 'but' differs from 'and' in that with it one intimates that what follows is in contrast with what would be expected from what preceded it. Such suggestions in speech make no difference to the thought. 25

And of course the emotive qualities of discourse have no bearing on the thought either. In 'My God! Dear old Mother

---

25 "The Thought", translated by Anthony Quinton in Mind, vol. 65, 1956, pages 295-296.

is drunk again!', neither the interjection nor the suggestion that mother is deserving of our sympathy are parts of the thought that the speaker's mother is drunk again.

Frege does not deny that imperatives, interrogatives, etc, have sense; but with the exception of some interrogative sentences, they do not express thoughts inasmuch as they cannot be appraised as true or false. Thus, these sentences have no reference--they do not refer to truth-values or any other kind of object. Frege recognizes, however, that declarative sentences are often used to express senses other than thoughts, as in 'All members will pay their dues by June 30.'

It is common nowadays to distinguish not only between a sentence and its meaning or sense, but also the statement (or "assertion" or "proposition") made (asserted) by a speaker in using that sentence on a particular occasion. Sentences containing pronouns or demonstratives provide the best illustration of this trifold classification. The sentence 'It is raining' has the same meaning for any two speakers who utter it at different times and in different places, but they do not assert the same statement--as can be seen from the fact that what one of them says may be true whereas what the other says is false. "What is said" or asserted, then, cannot be identified with the meaning of the sentence uttered. By the same token, 'President Johnson was in room 19 of the Pentagon in Washington D.C. on August 9, 1965' clearly does not mean the same as 'Johnson was here today'. But

if the latter sentence was uttered assertively by a janitor in room 19 on August 9, he will certainly have asserted the same statement as that asserted by a Johnson aide who uttered the former sentence at a news conference the next day. Indeed, if on August 10 the janitor were to utter assertively 'Johnson was here yesterday', he would be asserting the same statement as the one he made the previous day though obviously the meanings of the two sentences differ.

Although the distinction between meaning and statement, once it is made, appears rather obvious, it may nevertheless easily pass unnoticed. Most philosophers, until fairly recently, have in fact failed to notice it, and Frege seems to be one of them. For he describes a thought as the sense of a sentence (and seems to mean by this, "meaning"), but also makes it the bearer of truth and falsity which, it now appears, is a characteristic only statements can have. However, he also uses 'thought' synonymously with 'statement'. In "The Thought", he says:

> If someone wants to say the same today as he expressed yesterday using the word 'today', he must replace this word with 'yesterday'. Although the thought is the same, its verbal expression must be different so that the sense, which would otherwise be affected by the different times of utterance, is readjusted....In all such cases, the mere wording, as it is given in writing, is not the complete expression of the thought, but the knowledge of certain accompanying conditions of utterance, which are used as means of specifying the thought, are needed for its correct apprehension. 26

The second sentence of this passage shows the distinction

---

26 Ibid., page 296.

between sense (or meaning) and statement to be at Frege's fingertips, and yet it seems he failed to recognize it explicitly. When the janitor uttered the sentences 'Johnson was here today' and 'Johnson was here yesterday' on August 9 and 10 respectively, he was clearly making the same statement--what Frege here calls the same "thought". Because different "verbal expressions" are used, he speaks of the sense being "readjusted"; and if this is to mean that the two sentences <u>differ</u> in sense or meaning, he is correct. However, there is the clear suggestion that the <u>sense</u> of 'Johnson was here' (minus the demonstratives) is affected by "the different times of utterance", and this is <u>not</u> correct if by 'sense' is meant "meaning". It is the statement (what Frege here calls the "thought") that is so affected. What is here called "the knowledge of certain accompanying conditions of utterance" help determine not the sense or meaning of one's remarks but the statement asserted, for the sense of 'Johnson was here' remains unaffected by the fact that it was uttered by the janitor at a certain time in a certain place. If someone had uttered that sentence ten years earlier in an entirely different place, he would still have expressed the same sense or meaning, viz. that Johnson had been in the immediate vicinity, but would have made a different statement.

It seems, then, that Frege uses 'thought' and 'sense' of a sentence' to cover both <u>meaning</u> and <u>statement</u>. From

here on, I will use these terms to mean only what has here been called a statement. We will for the most part ignore the notion of meaning since Frege is obviously not concerned with it.[27] However, it is also likely that Frege would not have wished to sharply distinguish between meaning and statement. Like many philosophers in recent years, he probably would have construed senses or thoughts as the meanings of what Quine has called "eternal sentences". That is, the sense or thought expressed by uttering 'Johnson was here today' in the circumstances mentioned above may be identified with the meaning of a sentence containing explicit specifications of time, place, and person talked about; in this case, something like:

Lyndon Baines Johnson, thirty-sixth president of the United States, was in room 19 of the Pentagon in Washington D.C. on August 9, 1965.

In the case of such eternal sentences, meaning and statement coincide, and Frege's functional analysis of sentences and the doctrine of truth-value as their reference must be

---

[27] It may seem odd to use 'sense of a sentence' to a mean "statement" but my reason for doing so is that Wittgenstein uses it in a similar way in the Tractatus, and we shall have to examine some of his comments on Frege later on. The sense of a proposition, for Wittgenstein, is the particular claim that is made on an occasion of the proposition's use. And a proposition is a sentence in its "projective relation to the world". This is why I do not wish to translate 'Gedanke" as 'proposition', for it would not allow us to state the issues of disagreement between Frege and Wittgenstein in a neutral way--for Wittgenstein, propositions have senses, for Frege, thoughts are senses.

thought of as applying primarily to sentences of this sort.[28]

Sense and Truth-Conditions

Frege was also concerned with how the sense of a complex sentence is determined, i.e. how we can specify which thought it expresses (by a complex sentence is here meant one involving truth-functional connectives or quantifier expressions). In section 32 of the Grundgesetze, he tells us that the sense of a sentence (the thought it expresses) is the thought that the conditions under which the sentence would stand for the True are fulfilled. That is, in specifying under what conditions the thought expressed by a complex sentence is true, we have thereby determined which

---

[28] Many philosophers have acquiesced in the identification of statements (thoughts) with the meanings of eternal sentences even though there are objections to it. For example, statements can be described as true or false, plausible, unfounded, exaggerated, etc. whereas it would indeed be odd to ascribe such characteristics to meanings. Suspicious though this view is, it seems the best way of handling the problem for Frege. For if meaning and statement are distinguished in all cases, then many of Frege's earlier remarks would have to be drastically revised. It would, for example, be no longer intelligible to speak of "the reference of a sentence" per se, since an ordinary sentence can be used to make different statements, some true and some false. We would instead have to speak of sentence-tokens as referring to truth-values, and if a sentence is used to express n different statements, then we would have n equivalence-classes of tokens of that sentence formed according to the equivalence-relation "expresses the same statement as". All the tokens of a given class could then be regarded as referring to the same truth-value. Such a modification of Frege's view seems much too drastic to be feasible, and hence we must accept the rather dubious identification of thoughts with the meanings of eternal sentences.

thought it is. Specifying the truth-conditions amounts to specifying a sense, and knowing the sense of a sentence consists in knowing what would be the case if it were **true**. If S and S' are sentences, then the sense of ⌜S and S'⌝ can be given by saying that it will stand for the True (or its sense will be true) just in case S stands for the True and S' stands for the True. Knowing the sense of 'Everything has mass' consists in knowing that it will stand for the True if and only if for every singular term inserted in the gap of 'ξ has mass', the resulting sentence stands for the **True**.

On the other hand, ordinary singular, subject-predicate sentences like 'Socrates **is** wise' cannot be said to have truth-conditions in this sense, since they are the building blocks out of which complex sentences are constructed--there are no further <u>thoughts</u> which must be **true** as there was for example with ⌜S and S'⌝. We need only specify the senses of the component expressions.[29] We cannot even say that the thought expressed by 'Socrates is wise' is true just in case the object referred to by 'Socrates' falls under the concept <u>wise</u>, for the notion of "falling under" must be explicated in terms of "true"; and moreover, since that object is also the reference of 'the man who taught Plato', we would wind up with the sentences 'Socrates is wise' and 'The man who

---

[29] However, if, as in the atomistic ontologies of Russell and Wittgenstein, it is maintained that such sentences are covertly complex, then their senses will have to be specified in terms of further thoughts which are true, i.e. they will have truth-conditions.

taught Plato is wise' expressing the same thought--which is clearly wrong.

Frege's view that the sense of a sentence consists in its truth-conditions being fulfilled has become a sort of dogma in recent philosophy--rarely questioned and even more rarely acknowledged by those philosophers whose views presuppose it. Though it is a dogma, it should not of course be thought that there are no disputes as to what the truth-conditions of a given kind of sentence are. A striking example is the well-known conflict between Frege and Russell over sentences containing definite descriptions. On Russell's view, the truth-conditions of 'The author of Waverly is a Scot' are (in Fregean terminology): first, that there exists at least one singular sentence derived from 'ξ wrote Waverly' which expresses a true thought; second, there is no other such sentence containing a singular term with a different reference; and third, the singular sentence derived from 'ξ is a Scot' by inserting the same singular term as was inserted in 'ξ wrote Waverly' expresses a true thought.[30] Thus, for Russell, included in the sense of this sentence is a claim of the existence of one and only one person who wrote Waverly, and on this account, if there is no such person or more than one, the thought expressed will be false --just as it would be if there were such a person but he was

---

[30] The formal mode of speech is awkward here; the force of the first two conditions is simply: one and only one person wrote Waverly.

not a Scot. Frege, however, treats sentences containing definite descriptions just as he would sentences containing proper names, and such sentences as we have seen involve no truth-conditions as such. But what if there is no author of Waverly, or more than one? Frege's reply is that we must distinguish between the truth-conditions which specify the thought expressed by a sentence, and a presupposition for the thought being either true or false; and he assigns 'There is one and only one person who wrote Waverly' to the latter category. Thus, if the presupposition is false, then the thought expressed by 'The author of Waverly is a Scot' is neither true nor false, and has no reference. So if 'the author of Waverly' has no reference, the entire sentence --in accordance with the principle of identity of reference --likewise has no reference. Thus, there are, in natural languages, sentences which have a sense but no reference.

Frege's original statement of the truth-condition principle was: the sense of a sentence--the thought it expresses --is the thought that its truth-conditions are fulfilled. And the phrase 'are fulfilled' may seem puzzling. For he seems to be saying that the sense of a sentence does not consist simply in its truth-conditions, but in those conditions actually obtaining or being the case. To explain this, we must first note that Frege held the "eliminability" theory of truth; in "On Sense and Reference", he says that the sentences '5 is a prime number' and 'The thought that 5 is a

prime number is true' have the same sense--they express identically the same thought. Contrary to what might first be supposed, the latter sentence does not express the thought that the thought expressed by the former sentence is true; it is not itself an assertion about the former thought, rather it is simply the thought that 5 is a prime number--the very same thought expressed by the first sentence. As we will presently see, it is quite wrong to suppose that there are thoughts which are, in and of their own accord, assertions. Thus, when Frege states the truth-condition principle in terms of those conditions being "fulfilled", he is only saying that the thought expressed by '5 is a prime number' is the thought that 5 is a prime number; that is, the thought which is also expressed by 'The thought that 5 is a prime number is true' where 'is true' has the same force as 'is fulfilled'. However, it must not be supposed that knowing, the sense of a sentence consists in knowing that these conditions are fulfilled', for if this were true, we could never grasp the sense of a sentence designating the False. Rather, knowing the sense consists in knowing its truth-conditions, but such a sense is not merely something of the form: "x's-being-F", rather it is of the form: "that-x-is-F".

Assertion and Judgment

A declarative sentence, if its component terms have reference, refers to a truth-value, and in uttering it, the speaker is thereby expressing a thought which is either

true or false. However, in many contexts, he may not be concerned with whether the thought is true or not. For the sentence may be uttered or written in the course of telling a joke or as part of a play or perhaps simply to put forth a thought for consideration, and in such cases, says Frege, the thought is not being judged to be true. When a speaker assertively utters or writes down a sentence, he is committing himself to the truth of the thought expressed, or in other words, warranting the sentence as a designation of the True. Frege is here delineating a certain broad kind of linguistic performance which will include such "illocutionary acts" as reporting, saying, stating, etc. For in all such acts, the speaker expresses a commitment to the truth of the thought expressed. He uses the term 'judging' to cover all such cases, and when he speaks of "judgment", he intends this noun in the "process", not the "product", sense. In "The Thought", he says:

> We are probably best in accord with ordinary usage if we take judgment to be an act of judging as a leap is an act of leaping...Judging...is acknowledging the truth of something; what is acknowledged to be true can only be a thought...If the judgment is an act, it happens at a certain time and thereafter belongs to the past. With an act there also belongs an agent, and we do not know the act completely if we do not know the agent. 31

What we judge, then, are thoughts, not sentences; but we may also speak of the sentence being"uttered assertively" in such cases, or for short, being "asserted".

---

31 Mind, vol. 65, page 292.

In ordinary speech and written language, we must rely on the context of utterance in order to determine whether a speaker or writer has assertively uttered or written a sentence. But in Frege's symbolism, there is a special kind of sign such that when attached to a sentence forms a new expression which expresses the judging of the thought expressed by the component sentence. The whole expression does not itself express a thought, but rather a certain act with the thought of the component sentence as its object. This expression is called the assertion-sign, and the new expression thus formed an assertion. Assertions are formed in the following way: the sentence 'Jones is bald' is inserted in the gap of '—— ξ' to get the value-expression '——Jones is bald'. At the left tip of the horizontal we now attach a vertical stroke (the "judgment stroke") thereby getting the assertion '⊢——Jones is bald'. The complex sign '⊢—— ' is the assertion sign, and Frege insists that ' ⊢——Jones is bald', unlike '——Jones is bald', does not refer to anything. Rather, he says it "asserts something"; in effect, that the thought expressed by 'Jones is bald' is in fact true.[32] It is obvious, then, that ' ⊢——' does not

---

[32] This account of the assertion sign is based on Frege's later works, such as "Function and Concept" and the Grundgesetze. His exposition of it in the Begriffsschrift is faulty, and was corrected later on.
The assertion-sign would not be of much use when flagging sentences expressing empirical thoughts unless it is made clear from the context just which thought a (non-eternal) sentence-token expresses on a particular occasion of its use. Just seeing ' ⊢——Jones is bald' on a page does not tell us which Jones is being referred to or to whether this judgment about him is made at age 5 or age 55.

mean for Frege what it has come to mean in mathematical logic nowadays; it is not a metalinguistic sign which, when attached to a sentence of the object-language, says of it that it is a theorem. Rather, an expression consisting of a sentence with the assertion sign attached expresses the author's judgment that the thought in question is true--it indicates the asserting of a "content" and hence does not itself express a content.

The assertion-sign enables us to expose many confusions present in the traditional logic. In many older logic texts, one finds the doctrine that in a "proposition", the predicate is "asserted of" the subject--as if in forming a proposition, it is thereby being asserted or judged to be true. If this were the case, then all truth-functionally valid arguments would become circular, for since Q appears in one of the premisses of:

If P then Q
P
therefore Q

it would follow that Q has already been asserted, and the argument would be a petitio. But the first premiss is to be written, not ⌜if ⊢ P, then ⊢ Q⌝, but ⌜⊢ (if P, then Q)⌝, and this makes clear the fact that Q is not asserted in the premisses. The assertion sign resolves the old confusion between predication and assertion. As Frege says: "By combining subject and predicate, one reaches only a thought, never passes from a sense to a reference, from a

thought to a truth-value."[33] Each step in an inference is to be preceded by '⊢——', and the rules of valid inference are to be formulated in terms of assertions, not sentences *per se*. Moreover, the occurrence of a sentence in a truth-functional context is made clear by the assertion sign. For in ⌜Either P or Q⌝, P is not being asserted--it merely stands for a truth-value.[34] The assertion sign can properly attach only to the truth-functional sentence as a whole.

In Frege's symbolism, the assertion that Jones is *not* bald would be rendered '⊢¬— Jones is bald', and the assertion that if the accused was in Rome, he did not commit the murder would be:



The sign '⊢¬—' is a compound of '|' and '—¬—', and only the former carries the assertive force. This is important, for otherwise it might be thought that '⊢——' and '⊢¬— ' are polar opposites--as in the traditional distinction between "affirmation" and "denial". That is, in ⌜⊢—— S⌝ and ⌜⊢¬— S⌝ there are two distinct acts, affirming (or

---

[33] "On Sense and Reference" in Philosophical Works, page 64.

[34] I use the corner signs as "quasi-quotation" devices, and hence the letters 'P' and 'Q' in such contexts are to be understood in the manner explained by Quine. Later, I shall use small letters like 'p', 'q', etc. in the same way. See W.V.O. Quine, Mathematical Logic (Cambridge, Mass: Harvard University Press, 1940), section 6.

asserting) and denying, which are signified by '⊢——' and '⊢¬— ' respectively, and of which S is the common component. Frege does not hold such a view, and devotes a good deal of space in "Negation" to refuting it (he does not there use the signs ' ⊢——' and ' ⊢¬—', but his view can be more easily stated in terms of them). Only the vertical stroke carries assertoric force, and what is asserted is in the one case ⌜—— S⌝ and in the other ⌜—¬— S⌝. Negation is part of the content of what is asserted, it is not something which is the polar opposite of asserting. It is true that if one says ⌜not-S⌝, he can be described as "denying that S", but it hardly follows from this that 'not' somehow serves to indicate an act diametrically opposed to assertion. If this were so, then in formulating the negative of S, I would ipso facto be denying S, i.e., be committing myself to the falsehood of the thought it expresses. But clearly I can consider the negative of S without in any way committing myself as to its truth-value. Moreover, on this view, we can make no sense out of truth-functional contexts; in saying ⌜If not-P then Q⌝, I certainly am not denying that P. Thus, though S is a common component of ⌜——S⌝ and ⌜—¬—S⌝, this only means that it is the common argument-expression of two distinct value expressions, not the common object of two different acts. Therefore, in ⌜⊢——S⌝ and ⌜⊢¬—S⌝, we should not contrast '⊢——' with ' ⊢¬—', but ⌜——S⌝ with ⌜—¬— S⌝; the contrast is not between two "acts", but between two value expressions. We may speak of negative

sentences (and derivatively, of negative thoughts), and further of asserting a negative sentence (judging a negative thought), but the phrase 'denies that S' is to be construed 'asserts 'not-S'' or perhaps 'asserts 'S is false''.

Such a position derives naturally from Frege's view, developed earlier, that negation is a truth-function, and applies to a sentence as a whole, not to "predicates". And once it is realized that 'not' represents a truth-function rather than indicating an "act of denying", then such an act becomes a superfluous notion in logical theory. There is no point in postulating an act of denying, for we do not need three primitive notions: "affirmative" assertion, "negative" assertion, and the truth-function of negation. We need only assertion and the truth-function of negation.

However, this result must not be taken to mean that Frege held the absurd view that there is no illocutionary act of denying. Rather, it only means that denying, like reporting, stating, etc., can be subsumed under the general heading of assertion or assertive utterance--that there is no polar opposite of this. Frege uses 'assert' in a technical way--it does not signify just one illocutionary act among many, but rather signifies that broad range of linquistic performance in which our concern is for truth. That is, it covers all those illocutionary acts in which the speaker commits himself to the truth of the thought expressed by the sentence uttered, and this will include denying as well as reporting and stating. There is no point in

introducing a second "logical act" in which our concern is (presumably) for what is false. It should also be noted here that though one can be said to "assert something to be false" in the ordinary sense of 'assert', this does not hold for Frege's technical sense, since it is based on the fact that our aim in using descriptive language is to express true thoughts. In Frege's sense, we must say that one asserts or assertively utters a sentence of the form 'S is false'. A similar distinction between ordinary and technical senses holds for 'judge'.

Sentences, Assertions, and Truth

The notions of asserting and judging, together with the assertion-sign, are indispensable to Frege's philosophy of logic, but we have not yet mentioned the most important role they play in his theory. It is, I think, what Frege has in mind in a passage from "On Sense and Reference" where he says that in judging, we "advance from a thought to a truth-value." This rather difficult metaphor must now be spelled out.

We have not as yet said much about the notion of truth. We know that the True and the False are what sentences refer to, but this is not of much explanatory value since the notion of a "truth-value" (in the sense of an object sentences refer to) is defined in terms of the notion of truth--two expressions stand for the same truth-value just in case the thought expressed by the material biconditional containing

those sentences is true. We cannot, then, explain the predicate 'is true' by reference to the True, and Frege does not try to do this. In "The Thought" and "On Sense and Reference", he expounds the "eliminability theory", for he maintains that when we say of a thought that it is true we are not attributing to it a particular quality or property even though it may look that way. 'The thought that 5 is a prime number is true' says nothing different from '5 is a prime number'--we are not adding anything to the thought when we say that it is true. It might be objected that we are adding something, viz. judging or asserting it. But this is a mistake; both of the above sentences are assertable contents--they are expressions to which the assertion sign may be attached. To suppose otherwise, would be to suppose (as Frege himself did in the Begriffsschrift) that '⊢——' and 'is true' mean the same; that is, by attaching 'is true' to an expression of the form 'The thought that....' we thereby form not just a sentence, but an assertion. Such a view embodies the traditional confusion between assertion and predication mentioned earlier.

Nevertheless, there are difficulties with the eliminability theory--not the least of them being that if '5 is a prime number' and 'The thought that 5 is a prime number is true' express the same thought, this thought would also seem to be the reference of the subject expression of the latter sentence (what its sense would be is anyone's guess).

However, even if we pass over such difficulties, a problem still remains. For the eliminability theory and the doctrine of truth-values as the references of sentences still do not suffice to explicate fully the notion of truth. Such an explanation must account for the fact that truth is what we (normally) aim for in uttering declarative sentences--that just as the point of playing a game is to win, so the point in uttering declarative sentences is to communicate thoughts which are (held to be) true. Frege's theory is incomplete in the sense in which a description of a game would be incomplete if, in addition to specifying the rules, the description told us only that after the final move is taken, a player is in one of two possible circumstances, "W" and "L". Unless we are further told that "W" means what is usually meant by 'win', we do not know that the object of the game is to wind up in the "W" circumstance. Likewise, Frege has said that sentences stand for those objects he has labeled "the True" and "the False", but these expressions are only labels--just as "W" and "L" are in the game. Frege's theory has not accounted for the fact that it is the object of using descriptive language to utter sentences which express true thoughts--sentences which refer to the True. This lacuna in Frege's theory is filled by the assertion-sign and the related notions of asserting and judging. When we assertively utter or write down a sentence, and when we judge its thought ("advance from a thought to a truth-value"), we are giving concrete expression to this

interest or concern we have in the truth.

Nevertheless, some nagging doubts concerning Frege's theory remain--doubts in addition to whatever shortcomings his eliminability theory may have. For the introduction of the assertion-sign, and the activities of asserting and judging, seem to take on an ad hoc character; as if they were designed specifically to plug up a gap left open by the other elements in Frege's theory; and this makes it look as though there should have been no gap in the first place. Against this accusation, it may be urged that the assertion sign is not just an ad hoc device; it would be necessary in any theory. For were it not introduced, we would not be able to mark the distinction between assertion and predication. But as we have seen, Frege's theory of assertion does more (and must do more) than merely expose this ancient confusion. It is of course absurd to suppose that a sentence, insofar as its predicate is attached ("asserted" in the old jargon) to the subject, asserts of itself that it is true, and no one nowadays would wish to dispute this. But for all that, there does seem to be a sense in which one can legitimately speak of what a sentence says or asserts. At 4.022 in the Tractatus, Wittgenstein says: "A proposition shows how things stand if it is true, and it says that they do so stand." (His italics) In the first half of this quote, Wittgenstein is clearly talking about the sense of a proposition; he essentially agrees with Frege that the truth-conditions of a proposition determine

its sense, but he does not commit the old error of assuming that propositions are ipso facto assertions. Nevertheless, he does speak of a proposition as "saying" something, and we do not seem to be going against ordinary usage in speaking of what a sentence or proposition "says" or "asserts". This is something which we would ascribe to sentences or propositions but not to any other kind of complex expression, and the reason clearly is that propositions are the only kind of expression which can be appraised as true or false (in Fregean terms, only sentences are expressions whose senses can be so appraised). Being appraisable as true or false is that feature of the senses of sentences which makes it natural for us to speak of what a sentence says or asserts. It would be absurd to speak of what a definite description "says", and it would be equally absurd to place an assertion-sign in front of it. For this use of 'says' seems to be derived from another use wherein we speak of what a person "says" in the course of uttering a sentence--but only a sentence. In reporting something I have heard from someone else, I would utter a sentence of the form: 'He said that so-and-so' (or "informed me that" or "told me that" etc.); and obviously only a sentence can sensibly occupy the "so-and-so" position. Only an expression whose sense is appraisable as true or false can be a genuine candidate for assertive utterance, and what makes it natural for us to describe a sentence as "saying something" is that it is the only kind of expression a person can use to make an assertion.

Sentences, then, are complex expressions which express senses of a quite unique kind, and Frege's doctrine of sentences as a kind of referring expression leaves this fact unaccounted for. For on his theory, sentences are just one kind of value-expression among many, and they stand for objects in exactly the same sense in which any other complex, well-formed expression does. Frege's theory of sentences as value-expressions naming the True or the False does not, then, explain why the senses of these expressions are appraisable as true or false whereas those, say, of definite descriptions are not. Since in Frege's theory, sentences are introduced merely as value-expressions designating one of two objects, his theory does not come to grips with the fact that the sense of a sentence is uniquely appraisable as true or false--the theory seems to preclude any attempt to explain this feature by pointing to some special characteristic of sentences as such. To make up for this, the assertion-sign and the acts of asserting and judging are introduced. Now, at least, this feature of sentences has been recognized, but it still has not been explained, and the assertion-sign and the acts of asserting and judging appear again to be ad hoc devices designed to take care of a rather glaring omission in Frege's theory. It is true that they also serve the constructive purposes of exposing an ancient confusion and characterizing that use of language in which our concern is for truth, but their primary purpose (even though Frege did not explicitly realize it as such)

is to make allowance for the fact that sentences, though they are value-expressions, are certainly a very special kind of value-expression. We shall follow up this observation in the next chapter.

CHAPTER TWO

EXAMINATION OF FREGE'S THEORIES OF NEGATION AND ASSERTION

Frege's Account of Negation

We have expressed some rather general misgivings with respect to Frege's remarks on sentences, assertion, and truth, and we must now try to spell out these misgivings in a clearer, and more definite way. As a means to this, let us return to his discussion of negation, for his views on that very important topic lie at the heart of his theory of assertion.

Since Parmenides' time, philosophers have found the notion of negation very puzzling, and it will be to our advantage in the long run to try first to make clear some of the problems it raises. Suppose the thought expressed by 'a is F' is true. Then it is natural to suppose that there is something "in the world" ("reality") which renders it true--a state of affairs in virtue of which it is true. And it is also natural to suppose that this situation or state of affairs in some way involves the individual, a, and the property or concept, F, as constituents. But what, then, about the thought expressed by 'a is not F'? Although it is false, we can suppose it to have been true, and now we may ask: what state of affairs would render it true? Certainly it cannot be the same state of affairs that rendered the first thought true, but nevertheless it would,

presumably, have to be one that also involves a and F as constituents. But what, then, would distinguish these two different states of affairs? Surely it would be absurd to suggest that the second one contains an additional element signified by 'not'-- as if 'not' referred to some sort of strange property or individual. Could we, then, say that in the second state of affairs a and F are related in a way different from the way they are related in the first, and that 'not' signifies this relation? But what could such a relation be? Could we say anything more about it than that it is that relation which a and F bear to each other when they do not stand in the first relation (that they are "negatively related")? And does it make any sense to talk about it as a relation at all? If we think of the relation which is presumably expressed in 'a is F' in terms of the traditional universal-particular relation, then in saying 'a is not F' are we really saying that a and F stand in some relation other than this one? Rather, are we not merely denying that they stand in this relation? But if, in saying 'a is not F', we are expressing no relation between a and F, then it would seem that there is no state of affairs of which they are constituents--contrary to our original hypothesis.

Considerations of this general nature (though they have often been expressed in a different form) have led some philosophers to say that there are no negative thoughts, for if there is nothing which could render it true, how much sense

is there in speaking of a thought at all? And from here it was inferred that a negative sentence is either senseless (as Parmenides had said) or at least its sense is of a radically different nature from what we might first expect. With respect to this second alternative, if no relation between a and F is expressed in 'a is not F', then if we regard (as Frege did) a thought as a complex sense determined by the senses of the component expressions of the sentence expressing it, then it is difficult to hold that the sense of the negative sentence really is a thought unless we maintain that the word 'not' introduces some sort of relation which ties together the component senses in a way quite different from the thought expressed by 'a is F'. And it now remains utterly mysterious as to what sort of relation this could be; one might again object that 'a is not F' merely denies the relation expressed in 'a is F'. And many philosophers in Frege's time held, consequently, that inserting 'not' into 'a is F' merely "dissolves" or "breaks up" the thought expressed by 'a is F' instead of forming a new and different thought.[1]

---

[1] Similar arguments to the above can be employed to show that there is nothing in reality corresponding to a false thought, and hence that there are no false thoughts (they could exist only if something corresponded to them, and in that case they would be true). In the opening passages of "Negation", Frege argues against such a view by invoking the distinction between grasping the sense of a sentence and judging what is grasped to be true. The argument against false thoughts obviously confuses these notions, but it is not quite so obvious in the case of negative thoughts.

However, in "Negation"[2] Frege shows conclusively that the "dissolving" view cannot be right, for if it were true that merely by inserting 'not' in 'a is F' we deny that a and F stand in the relation expressed in 'a is F', then we could not come to understand the sense of the negative sentence without also denying the truth of the thought that a is F, and this is clearly absurd. We can form the negative of 'a is F' and grasp its sense without in any way committing ourselves to the truth-value of the thought expressed by 'a is F'. It is of course correct in a sense to say that 'a is not F' denies what 'a is F' affirms, but all this can mean is that the sense of the former is true whenever the latter is false and vice-versa, and hence when a speaker asserts the former, he is denying the thought expressed by 'a is F'. The sense of 'a is not F' cannot properly be said to "deny" or "dissolve" anything, but that sense does consist in being true when the sense of 'a is F' is false and vice-versa. The "dissolving" view, taken in this way, confuses grasping a sense with judging that sense. Thus, the sense of 'a is not F' must be something which can be understood independently of any judgment we make as to its truth-value, and if an advocate of the "dissolving" view now admits this, but still persists in holding that the sense of 'a is not F' simply consists in the senses of 'a' and 'F' not being related so as to form a complex, then he would

---

[2] Philosophical Works, pages 122-124.

seem to be totally at a loss to account for the Law of Double Negation. For if, by inserting 'not' into 'a is F', the thought originally expressed is simply dissolved, how is it that it miraculously reappears when we negate it again? As Frege says: "Negation would thus be like a sword that could heal on again the limbs it had cut off."[3]

Viewed against a background such as this, Frege's explanation of negation and the ensuing theory of assertion can be seen as the major accomplishments they are. If the "dissolving" view will not work, we must then accept that there are negative thoughts, and return to the original question of just how 'not' when inserted in 'a is F' helps determine a new thought which always has the opposite truth-value from that expressed by 'a is F'. Does 'not' introduce a new component sense which together with the others determines the complex sense expressed by 'a is not F'? Or does it just somehow relate the component senses of the thought expressed by 'a is F' in a different way so as to produce a new complex sense? In what way, exactly, does 'not' serve to introduce a new element which then helps determine the sense of the whole? This is what originally seemed so mysterious about negation. And Frege's account of it as a kind of function is designed to put to rest such questions once and for all. By taking 'not' to apply to the sentence as a whole rather than the predicate, and construing

---

[3] "Negation", in Philosophical Works, page 124.

'a is not F' as standing for the value of a certain function for the argument referred to by 'a is F', Frege resolves the puzzle of how 'not' is supposed to introduce a further element which in part determines the sense of 'a is not F'. We can then if we wish speak of 'not' as introducing such an element, but it is now clear that it is an element of a logically different order from the others, but which can be explained in a very precise way. A function is defined in terms of the values it takes for its various arguments, and by construing 'not' as representing a function from truth-values to truth-values and sentences as value-and argument-expressions, Frege has given us an account of the thought expressed by 'a is not F' which avoids the puzzles raised by earlier accounts. And it seems we need no longer worry about the ancient question of what, if anything, could correspond to a negative thought and render it true, for this problem trades on the assumption that 'not' must either signify some element in addition to a and F and which is a part of the "negative" state of affairs, or else signify some sort of mysterious, inexplicable relation holding between a and F. But now it can be seen that it does neither --rather, it serves to represent a particular truth-function such that the negative sentence stands for the value of that function for the argument referred to by the positive

sentence.[4]

Frege's account of negation, together with the theory of assertion, thus seems to dispose of many traditional puzzles pertaining to logical theory. I say "seems" because while Frege is clearly on the right track in treating negation truth-functionally, the account he gives does not fit in very well with some of the other principles in his theory, and the fact that it does not points up some rather serious shortcomings in his philosophy of logic.

At 4.431 of the Tractatus, Wittgenstein says:

> ...A proposition is the expression of its truth-conditions.
> (Thus, Frege was quite right to use them as a starting point when he explained the signs of his conceptual notation. But the explanation of the concept of truth that Frege gives is mistaken: if "the True" and "the False" were really objects, and were the arguments in $\sim p$, etc., then Frege's method of determining the sense of '$\sim$' would leave it absolutely undetermined.)

I think Wittgenstein is getting at something very important here. He agrees with Frege that the truth-conditions determine the sense of a sentence (the thought expressed), and hence for any negative sentence like '$\underline{a}$ is not F', its sense must be specified by specifying the conditions under which it would be true. But Frege's account of negation as

---

[4] Again it must be noted that Frege introduced $\neg\xi$ as a concept, but with the emendations suggested earlier, it can be regarded as a genuine truth-function. It should also be noted that just as '$\underline{a}$ is not F' stands for an object which is a truth-function of the reference of '$\underline{a}$ is F', so the thought expressed by '$\underline{a}$ is not F' can be regarded as a function of the thought expressed by '$\underline{a}$ is F'. Frege calls such functions "sense-functions"; see his essay, "Compound Thoughts", translated by R. Stoothoff in Mind, vol. 65, 1963.

a function of truth-values to truth-values leaves the sense of the negative sentence unspecified--only its reference is determined. For negation has been explained by introducing the function $\neg\xi$ as that function whose value is the True for just those arguments for which the value of $\neg\xi$ is the False and vice-versa; and so Frege is saying that 'a is not F' is that sentence whose reference is the True when the reference of 'a is F' is the False etc. But this does not enable us to specify the sense of 'a is not F', for as Frege himself would admit, the sense of an expression cannot be specified merely by specifying its reference.

Thus, if truth-values are the references of sentences, and specifying a reference does not entail specifying a sense, then Frege's account of negation--stated in terms of the conditions under which a negative sentence would refer to the True--is unacceptable inasmuch as the sense of the negative sentence has not been determined. This should not be taken as a critique of the truth-condition principle--we specify the sense of 'a is not F' by stating the conditions under which it would be true. Wittgenstein's remarks are clearly directed to Frege's statement of that principle; that is, in terms of sentences referring to the True, instead of merely thoughts being true. There would thus seem to be a fundamental incompatibility between the principle that truth-values are the references of sentences and the principle that the truth-conditions determine the sense

of a sentence. As long as it is possible to formulate the truth-condition principle in terms of sentences referring to the True or the False--that is, as long as truth-values are references of sentences--we cannot specify the sense of a sentence by stating its truth-conditions. And it is quite clear that for Frege the primary statement of the truth-condition principle would be in terms of sentences referring to the **True**, not in terms of thoughts being true. For the combined import of his eliminability theory of truth and the principle that the True and the False are the references of sentences is that when we say of a thought that it is true, we are not ascribing a property to it (as we do when we say of a table that it is round). Rather, the relation between a thought and its truth-value should be analogized to that which the sense of a definite description bears to the object to which the description refers. And I think we can regard Wittgenstein here as saying that this picture of Frege's is a bad one since it conflicts with the truth-condition principle.

Wittgenstein's criticism applies obviously not only to sentences containing 'not', but to sentences containing any logical connective at all. However, of all the truth-functions, negation is of special interest, and we shall return to it later on.

Now I think we must agree that if one of these incompatible principles must go, it must surely be the principle that truth-values are the references of sentences. For it

is far less "self-evident" than the truth-condition principle (stated, of course, without reference to "the True" and "the False"); indeed, it required a good deal of argument in the last chapter. It is, moreover, a commonplace in present-day logic texts to introduce negation by saying that the thought or proposition expressed by '_a_ is not F' is that thought which is true just in case the thought expressed by '_a_ is F' is false and _vice-versa_. Now showing that the True and the False are not the references of sentences does not by itself entail that sentences have _no_ references at all. However, the arguments presented in the last chapter for saying that truth-values are what sentences stand for seemed air-tight once we grant the underlying premiss that the sentences of ordinary discourse, like equations in mathematical discourse, can be analyzed functionally. That is, once we have granted the very plausible thesis that sentences, like other complex expressions, are value-expressions, and hence stand for objects which are the values of certain kinds of functions for certain arguments, it readily follows that truth-values (the True and the False), as defined in the previous chapter, were those values, and thus, the references of sentences--all other plausible candidates were effectively ruled out. In other words, if we say that sentences have references at all, then truth-values must be those references; and now if we are forced to reject this latter claim (owing to its incompatibility with the fundamental principle that the truth-conditions

give us the sense of a sentence), then we seem forced to deny the former claim as well--the arguments in the earlier chapter notwithstanding. It is by no means clear as to just what is wrong with Frege's contention that sentences, like other complex expressions, refer to objects (an answer to this can only come much later); Wittgenstein's argument, as with any reductio, only indicates that it is wrong. But if we deny that sentences have references, we must clearly have to find some other means by which to give a functional analysis of sentences. We certainly do not wish to reject every form of functional analysis, for it is just this kind of analysis which gives modern logic its superiority over the traditional, Aristotelian account. We could perhaps treat the sentence as itself the value of a function (a "linguistic function"), or perhaps take its sense as the value of a function (the "sense-functions" Frege speaks of in his later writings). These suggestions will not be pursued any further here, though we shall have occasion to return to them. Instead, we must note that the main thrust of Wittgenstein's remark was that the sense-reference distinction ought not be applied to sentences. That they have sense cannot be denied, but if we reject the view that they stand for values of certain functions (in effect, that they are just another kind of value-expression), then we no longer have any reason to suppose them to have references--and every intuitive reason to think otherwise. By this I do not mean that they may merely fail to refer owing to lack of reference on the part of a component expression,

for Frege, as we saw, would have allowed for this. Rather, the point is that sentences are complex expressions of a logically different order, and that the question of their referring or failing to do so does not arise. And this takes us back to one of the misgivings mentioned toward the end of last chapter, viz. that sentences, since their senses are uniquely the kind which admit of truth and falsehood, are a radically different kind of value-expression from, say, definite descriptions. However, we must dig deeper to see wherein this difference consists, and as a first step we shall now return to Frege's theory of assertion.

Frege on Assertion and Truth

Another misgiving noted in the last chapter concerned Frege's theory of assertion. On the one hand, it seemed eminently reasonable inasmuch as it exposed and corrected some of the mistakes of earlier logicians, and yet on the other, it seemed to be curiously ad hoc--as if it were designed to plug up a rather glaring omission in Frege's theory. Perhaps we are now in a position to explore this matter further.

Frege's theory of assertion had three distinct purposes: first, to distinguish between the assertive and non-assertive uses of sentences (and these uses from the sentence itself); second, to dissolve the many logical puzzles which trade on the confusion between assertion and predication; and third, to account for a feature of the notion of truth which had previously been ignored by the eliminability theory and the

doctrine of truth-values as references, viz. that truth is what we aim for in assertion. And he accounts for this by saying that in asserting a sentence or judging the thought expressed, we "advance from a thought to a truth-value."

Now certainly there can be no quarrel with regard to the first two purposes--at least as they have been stated thus far. These features of the theory are something which anyone would accept nowadays. No one would want to deny that there is, in an obvious sense, a distinction to be drawn between merely grasping or understanding a thought and judging it to be true, and correspondingly, on the linguistic level, between the mere uttering of a sentence and the various illocutionary acts falling under Frege's general heading of "assertion" that may be performed in uttering it. Thus, this feature of Frege's theory would be neutral with respect to any particular philosophy of logic; indeed it is not only compatible with, but a necessary prerequisite to, any properly stated philosophy of logic--even one which denied, for example, that sentences have a reference.

Moreover, there seems to be a philosophically neutral element implicit in the third purpose as well. For certainly no one would deny that in asserting a sentence we are expressing or manifesting our interest in the truth. What is not neutral of course is the interpretation Frege puts on this; that we are "advancing from a thought to a truth-value"--from a sense to a reference. Since we have already found reason to question the sense-reference distinction as applied

to sentences, this interpretation now becomes suspect as well. Moreover, it should also be noted that there have been many philosophers (most notably Wittgenstein and Russell) who would object to Frege's sharp distinction between the thought and its truth-value. They would maintain that being-true-or-false is an essential ingredient in the sense of a declarative sentence, and deny the possibility of thoughts which are neither true nor false owing to lack of reference of a component term in the sentence expressing it--such thoughts would simply be false. Thus they would object to drawing the sense-reference distinction with respect to sentences and therefore to construing assertion or judgment as an "advance" from a thought (sense) to a truth-value (reference). However, these difficult issues can be fruitfully discussed only after some further points have been developed. For the present, we must turn to another suspicious feature of Frege's theory of assertion, viz. that the notion of truth occurs in it at two different and seemingly unrelated points: first, merely as something ascribable to a thought whenever the sentence expressing it stands for the True[5], and second, as that which

---

[5] This is a rather tortured phrase. My only reason for using it is to avoid using the word 'property'. I do not wish to give the impression that I am at this point rejecting Frege's eliminability theory. The phrase is supposed to convey only the fact that thoughts or propositions are the things to which the predicate 'is true' is applied; I do not want here to raise the question of whether in applying this predicate we are attributing a property to a thought or merely, as Frege would hold, re-expressing this thought by means of a different sentence.

is attained when the act of asserting has, so to speak, been "executed successfully". There is nothing wrong in saying that we express our concern for the truth when we engage in assertive utterance, but it is quite mysterious as to just how this feature of truth--that it is what we wish to attain when we utter sentences assertively--is related to Frege's previous characterization of it as what is ascribed to a thought when the sentence expressing it has a certain object for its reference. For there is nothing peculiar or unique about this object nor (in Frege's theory) about the value-expressions which refer to it, i.e. nothing which might explain why truth and falsehood are not just two mutually exclusive categories for classifying thoughts. Truth and falsehood, according to Frege's theory, are on an equal footing--more like heads and tails than winning and losing. Just as a sentence will refer to either the True or the False, so a coin, when tossed, will either come up heads or tails. And now the theory of assertion, belatedly it seems, takes note of the fact that they are also like winning and losing--the point of playing a game is to win, and the point of assertive utterance is to communicate a truth.[6] But Frege has not shown us any connection between the "heads-tails" and "win-lose" aspects of truth and falsehood--the latter aspect on Frege's theory seems to be unique

---

[6] We do of course sometimes lie, but lying would not make sense except on the presupposition that we commonly intend to say something we hold to be true.

and sui generis. The theory of assertion exhibits a recognition that it exists, but being a belated recognition, it is left entirely unaccounted for. Nothing in the eliminability theory or in the doctrine of truth-values as references--nothing, that is, in his remarks that pertain directly to truth--sheds any light whatever on this feature of the concept.

Now of course one can take a coin and make a bet with it, and in this way we can see how the concepts of "win-lose" and "heads-tails" are related. We could let "heads" mean I win, tails I lose. Can we now use this analogy to explicate the relation between the two features of truth mentioned above?

Frege's theory of truth certainly seems to resemble the coin-bet example. Just as one uses a coin in betting, so one uses a sentence in assertive utterance, and just as we aim to win when betting, so we aim for truth when asserting a sentence. The bet is considered "successful" (a "winning bet") if and only if the coin lands with the heads side facing up, and likewise an assertive utterance may be considered successful just in case the sentence used refers to the True. Success in betting is correlated with the coin's coming up heads, and success in assertive utterance is correlated with the sentence's referring to the True--the only difference is that we have a separate word for success in betting, viz. 'winning'. Compare "If the thought expressed is true, then

my utterance is successful' with 'If the thought expressed is the case, then that thought is true'. Herein lies the different aspects of truth: being the case, and the fact that success in assertive utterance consists in expressing something which is the case; likewise we would have the coin's coming up heads, and the fact that in this game success or winning consists in its coming up heads. Finally, to complete the analogy, just as it is purely a matter of fact as to whether the "heads" side comes up or not, so it is a matter of fact (if, as has been assumed, we are dealing with ordinary, contingent thoughts) as to whether the sentence in question refers to the True or to the False.

Clearly the chief difficulty with the analogy is this: our decision to let "heads" mean "win" was a purely arbitrary one--we could just as easily have chosen "tails". But is it plausible to suppose that this can be carried over to the case of sentences? If Frege's theory makes it look this way, then it surely becomes suspect. In the coin case, there are two possible distinct games: one in which our interest is in having the "heads" side facing up after a toss, and the other in having the "tails" side showing. That is, in the first case, our bet is a successful one if the heads side is showing; in the second, if the tails side is showing. Now in ordinary speech, according to Frege, an act of assertive utterance is successful if the sentence asserted has a fact the True as its reference. Could we, then, form a new "language-game" in

which the act of assertive utterance is regarded as successfull if the sentence asserted has the False as its reference? Could we aim for falsehood just as we now aim for truth? And would it, then, be a purely arbitrary "decision" on our part that we had rejected the latter game in favor of the former?

Well certainly we could imagine two people playing a game (we won't ask why for the present) in which the speaker's assertively uttering a sentence, S, is to be taken by the hearer as indicating that the speaker holds the thought expressed by S to be false. That is, we can at least conceive of a game which involves the use of a natural language such as English, and which contains a rule to the effect that one should assertively utter a sentence only when the speaker (in Fregean terms) holds that sentence to refer to the False; the act of asserting is executed "successfully" just in case S does in fact have the False as its reference. So falsehood in this game would be the analogue of winning in the coin example--not truth as in our actual "game". Thus we have two possible "language-games"; in the first (hereafter, "language-game$_1$"), an assertion is successful if the sentence uttered refers to the True--this is our actual game--and second (hereafter, 'language-game$_2$'), a successful assertion is one in which the sentence uttered refers to the False. Clearly these two games parallel the games spoken of in connection with the coin example.

Now our language-games consist of two elements: that which is used (the sentences of the language), and a rule

specifying what constitutes a correct use of the sentence, and the two language-games differed only with respect to the latter. However, we could form yet a third language-game which contains the same rule of correct assertion as language-game$_1$, but this rule is applied to a different "language", viz. one which contains the same sentences as English but the senses of these sentences have been "reversed". That is, the sense of S in English is in this new language the sense of not-S and vice-versa. We shall refer to this language as "English-N". Then we could formulate a language-game$_3$ which contains the same rule of successful assertion as language-game$_1$, but which is applied to English-N instead of English. And obviously there would be a language-game$_4$ which contains the same assertion rule as language-game$_2$, but which, like language-game$_3$, is applied to sentences of English-N.[7]

---

[7] Our interest in this essay is directed mainly toward ordinary simple, subject-redicate sentences and the thoughts they express. For it is in terms of these sentences that the truth-conditions of sentences containing quantifiers, logical connectives, etc. must be expressed. Thus they are the simplest sentential unit of logical analysis (we are not concerned here with whatever convert complexity "logical atomists" purport to find in them). Hence, though the notion of a language like English-N takes on a suspicious air when we turn to sentences with oratio obliqua clauses, and sentences containing words like 'all', 'some', 'and', 'or', etc., we should not be put off by this fact. We can perhaps learn something by restricting our attention to simple subject-predicate sentences and their negative counterparts.

It should be apparent now that the linguistic behavior exhibited as a result of using language-game$_1$ is identical with the behavior exhibited in language-game$_4$. For suppose a speaker in the latter game were to assertively utter a sentence, S. Since the sense of S in English-N is the same as the sense of not-S in English, and since the speaker in question holds the sense to be false, then clearly a speaker in the first language-game would assertively utter that sentence which expresses the opposite sense, and this would be to assertively utter S since its sense is held to be true. And it is equally obvious that the same linguistic behavior is exhibited by two speakers one of whem used language-game$_2$ and the other language-game$_3$. This means that were we to stumble across a community whose members said "It is raining" whenever the sun was shining and "It is not raining" whenever it <u>was</u> raining, their behavior could seemingly be described in two alternative ways; either by means of language-game$_2$ or by means of language-game$_3$. Yet it seems quite obvious that we would describe it in terms of the latter game rather than the former. I think there is a good reason for this, for there is something very wrong in supposing that a community of language users could seriously go about aiming for falsehood, and I shall eventually try and show this. However, it will first be necessary to lay some groundwork.

Whenever it is correct to assert a negative sentence in language-game$_2$, it is also correct to do so in language-game$_3$ and vice-versa. What, then, would be the difference between the two? A natural answer to this question would be to say that in the latter game, the negative sentence expresses a positive thought since the language used is English-N. What is meant here by 'positive thought' is one expressed by a positive sentence in English (it being the standard accepted language). In other words, the remark that negative sentences are used to express positive thoughts in language-game$_3$ merely means that in that game, employing as it does English-N, a negative sentence is used to express a thought which is expressed in a language-game employing standard English by the corresponding positive sentence. And our presupposition has been that any sentence that counted as negative in English would equally count as negative in English-N, and likewise for positive sentences. English and English-N differed only in that the senses of these sentences were "reversed". Let us follow this out and make our criteria more explicit. On the whole, a sentence of the form 'a is F' was considered positive, and 'it is not the case that a is F', and the synonomous

form: 'a is not F', negative.[8] We have, then, relied mainly on grammatical considerations (the occurrence of 'not') and synonymy of sentences of different grammatical forms. Let us now also stipulate that in order for 'a is F' to be positive, 'F' must not be a "complementary" predicate (except in those cases where its sense does not have exactly the same force as 'not'), e.g. 'incorrect', 'unwise', etc., and likewise for 'a is not F' to be negative. It must be further stipulated that any sentence whose sense is the same as 'a is F' (where 'F' is not complementary) is also positive--thus, 'it is not the case that a is not F' will be positive. Correspondingly, any sentence which expresses the same sense as the negative sentence 'a is not F' (again, where 'F' is

---

[8] Here, as before, only singular sentences are being considered, i.e. those whose grammatical subject is a singular term. For it is characteristic only of these sentences that the same thought is expressed by 'a is not F' as by 'it is not the case that a is F'. A sentence of the form: 'All S are P' is clearly a quite different case; the two negative forms express different thoughts. There are of course singular sentences which are often used to express general thoughts, e.g. "The whale is a mammal" is most often used to express the same thought as is expressed by 'All whales are mammals.' But even here there is no distinction between the thought expressed by 'The whale is not a mammal' and 'It is not the case that the whale is a mammal.' Both are used to express the same thought as is expressed by 'No whales are mammals.' Moreover, our concern here is with distinguishing positive from negative singular sentences of the subject-predicate form (i.e. formed by inserting a singular term in the gap of a concept expression). The following discussion does not apply to identity sentences, nor of course to existential ones--only the "predicative sense" of 'is' is here being considered.

not complementary) will likewise be termed negative. Thus in both English and English-N, the following sentences will all be negative:

> Socrates is not wise.
> Socrates is unwise.
> It is not the case that Socrates is wise.

Moreover, all will express the same thought, but the thought they express in English is of course just the opposite of the thought they express in English-N, i.e. whenever one is true, the other is false. The thought they express in English has been termed "negative", and the opposed thought, expressed in English-N, "positive", The criterion for distinguishing positive from negative sentences could no doubt use some refinement and emendation, but its point should now be clear.

Now I think it is also possible to devise a "non-linguistic" criterion which will distinguish two classes of thoughts corresponding to the classes of "positive" and "negative" thoughts as they have just been defined. That is, there is a method of classifying all thoughts into one of two mutually exclusive and jointly exhaustive classes, and these classes will be the same as the class of positive thoughts and the class of negative thoughts; and this classification can be effected with respect to any given thought without having first to determine whether it is expressed in a given language (English or English-N) by a positive or a negative sentence (as they were defined above). This new criterion can be most easily spelled out in the following way:

consider a class of predicates or concept expressions which represent concepts that are "determinates" of some "determinable"; for example, color predicates. And let us further consider an object, a, which must fall under one of these concepts. Then, if there are n concepts, there will be n possible situations of the following kind:

a's being red
a's being green
a's being blue
etc.

Let us hereafter refer to each such situation as an "alternative", and the thought expressed (in English) by 'a is red' can then be said to "pick out" or "specify" or "describe" the first alternative in the above list. Knowing which thought 'a is red' expresses consists in knowing under what conditions that thought will be true, and it will be true just in case the alternative in question obtains or is the case. Now if the thought expressed by 'a is red' is true, then the thoughts expressed by 'a is green', 'a is blue', etc. will all be false, and the alternatives they specify will not obtain. We shall hereafter refer to this feature by saying that the thought that a is red excludes these alternatives, i.e. a thought excludes an alternative just in case from its being true, it may be inferred that the alternative in question does not obtain. Now let us consider the thought expressed by 'a is not red'. Clearly it excludes only the alternative specified by the thought that a is red; it leaves open, let us say, the remaining alternatives, i.e. one may not infer with respect

to any one of them neither that it is or that it is not the case. Thus, the thought expressed by '_a_ is red' excludes all those alternatives except the one it specified and it leaves no alternative open at all whereas the thought expressed by '_a_ is not red' excludes only the alternative specified by the first thought and leaves all of the remaining alternatives open.[9] Generalizing, we may say that a thought expressed by a sentence of the form '_a_ is F' (where 'F' is one of the color predicates) will leave no alternatives open and exclude _n_-1 of them, but a thought expressed by a sentence of the form 'a is not F' will leave open those very same _n_-1 alternatives excluded by the first thought and will itself exclude only the remaining $\underline{n}^{th}$ alternative. It should be clear, then, from the definitions and from the example itself that we can divide the 2_n_ thoughts which are expressed either by a sentence of the form '_a_ is F' or one like '_a_ is not F' into two mutually exclusive and jointly exhaustive classes; the first class, or α-class, consisting of those that exclude _n_-1 alternatives, and the second or β-class consisting of those that leave open _n_-1 alternatives.

---

[9] It is clear that we will need another term in addition to 'excludes' and 'leaves open', viz. one which describes the relation between the thought that _a_ is red and the alternative it picks out. We might say that it _affirms_ that alternative; however, this notion introduces philosophical problems which will be discussed in great detail in the next chapter but which are not particularly relevent here, e.g. what is the relation between "affirming" and "picking out"? And, what, if anything, does the thought that _a_ is not red "affirm"?

Earlier, we gave a "language-dependent" criterion for distinguishing "positive" and "negative" thoughts--"language-dependent" because the criterion made reference to the sentence which, in English, expresses the thought. A positive thought was one which was expressed in English by a positive sentence (as defined previously), and similarly for a negative thought. This allowed us to describe English-N by saying that a speaker of this language uses a positive sentence to express a negative thought, and a negative sentence to express a positive one. Now if we consider the class of positive thoughts as defined in the color model, it is clear that this class is identical with the α-class, and the class of negative thoughts is identical with the β-class. More generally, the "independent" criterion just given for determining whether a thought belongs to the α-class or the β-class will ipso facto determine whether it is positive or negative. The criterion is "independent" because it does not matter whether we specify the thoughts involved in terms of English or English-N. We did in fact use English, for we relied for the specifications of the thoughts on phrases of the form: "the thought expressed by..." where it was assumed this meant "expressed in English by...". But of course the criterion for distinguishing α-thoughts from β-thoughts could just as well have been expressed in terms of English-N, and the criterion would then have read; a thought expressed by a sentence of the form 'a is not F' leaves no alternative open and

excludes n-1 of them, and a thought expressed by 'a is F' leaves open those same n-1 alternatives and excludes the remaining one. In other words, 'a is F' expresses a α-thought in English-N but a β-thought in English, and our criterion for distinguishing α- and β- thoughts is "independent" of which language is mentioned in explaining the criterion--our object language could have been either one, and the distinction between α- and β- thoughts in no way hinges on whether the sentence involved is positive or negative. Thus, since the "language-dependent" and the "independent" criteria divide singular thoughts into the same two mutually exclusive and jointly exhaustive classes, let us drop the 'α' and 'β' significations and simply speak of the independent criterion as a criterion for distinguishing positive from negative thoughts. We may now describe a positive thought as one which leaves open no alternatives and excludes n-1 of them (all except the one it specifies), and a negative thought as leaving open those same n-1 alternatives and excluding the specified one.[10]

Critique of Frege's Theory of Assertion

Now let us return to the notion of using a language-game in which an assertive utterance is considered successful if and only if the thought expressed is false. It has already

---

[10] It may be wondered here if the distinction between α- and β-thoughts still obtains where n=2. I shall argue later in this chapter that it does, but for the present, there are more pressing issues to be dealt with.

been mentioned that it is conceivable that two persons could play language-game$_2$--perhaps as a diversion or a joke--that is, by persons who would normally play language-game$_1$. Troubles arise only when we try to imagine language-game$_2$ being used as the accepted mode of discourse by a community. The major purpose underlying all assertive discourse is to impart to the hearer certain information which he may wish to have and which may in a broad sense be useful to him. Now suppose I am trying to teach someone to play draw poker; one of the facts I must get across to him is that each player at the beginning of the game is dealt five cards, and if we are playing language-game$_2$, I must get this information across to him by saying something false. Clearly I cannot succeed by saying that each player is dealt six cards, for while the subject now knows that a poker hand does <u>not</u> contain six cards, he still has no idea as to how many it does contain. Obviously I can only succeed by asserting that a poker hand does <u>not</u> contain five cards, and this shows us why the same linguistic behavior results from the use of language-game$_2$ and language-game$_3$. In order for the hearer to do what I wish him to, he must be prepared to act on the negation of what I actually say to him--that is, to act on what is (or is supposed to be) <u>true</u>. And because of this, it is no accident that we, as users of language-game$_1$, would describe the linguistic behavior of a community which said 'It is raining' when it is not raining in terms of language-game$_3$ rather than

language-game$_2$--it is not "mere prejudice" on our part because we aim for truth. Rather, as the poker example shows, there are many contexts where one can act successfully only on the basis of a positive thought which is taken to be true. Only the positive thought leaves no alternative open, and more importantly, excludes all those alternatives except the one specified. It would indeed be impossible for me to play poker in the prescribed manner solely on the basis of positive thoughts I know to be false--that, for example, a hand consists of six cards or seven cards, etc. In order to know what I should do, I would also have to know all the things I should not and that these are all of them--something which is clearly impossible in the example given. Our interest is in true, positive thoughts because on the whole we wish to have specific or definite knowledge of what is the case, and a positive thought, by leaving no alternative open and excluding all except the specified one, does just this. A negative thought, on the other hand, is definite only about what is not the case, the rest being left open. However, definite knowledge of what is the case is also ipso facto definite about what is not since the remaining alternatives are excluded (definite not about how many alternatives do not obtain, but definite in the sense that however many there are, each and every one of them does not obtain).

Thus our interest is in true, positive thoughts because we wish to act successfully in a wide variety of circumstances;

whether it be poker playing or something far more fundamental to our well-being. We could no doubt achieve the end of communicating true, positive thoughts by aiming for falsehood and then assertively uttering a sentence which expresses a negative thought, for we could easily then infer the truth of the corresponding positive one. But to do this would at best be an unnecessary complication and at worse utterly pointless. Success in communicating to someone how to play poker correctly consists in having succeeded in communicating to him certain true, positive thoughts, for only be being positive does it exclude all other alternatives and only by being true (being "the case") does it serve as a basis for the hearer's acting correctly or in the manner the speaker wishes. What point, then, could there possibly be in counting as itself successful an assertive utterance wherein the thought expressed is false?

Moreover, the speaker's knowledge of poker is clearly expressed in the true, positive thoughts he is trying to pass on to the hearer. And indeed what we usually refer to as understanding or knowing how to perform properly some kind of activity presupposes the subject's knowing that certain positive thoughts are true. Further, virtually all conceptual knowledge (e.g. understanding or knowing what, say, a triangle is) equally presupposes knowledge that certain positive thoughts are true; as is evident from the fact that I cannot teach someone what a triangle is by saying what it

is not, i.e. in our language-game, by saying that it does not have four sides or five sides, etc. I can in language-game$_2$ communicate the required true, positive thoughts to the hearer by expressing the corresponding false negative, but I cannot understand how to play poker or conceive of a triangle in terms of positive thoughts which are false. Such cases, especially where an infinite number of alternatives is involved, suggests that though we can use a language-game in which we aim for falsehood, we must think in terms of truth. If so, this will have some interesting consequences for Frege's notion of "judging" as the mental or psychological correlate of assertive utterance.

Though assertive utterance can be aimed at truth or falsehood (i.e. success can be gauged either in terms of a sentence referring to the True or referring to the False), I shall try to show there is no such question of "aiming" with respect to judgment. The rule for successful assertive utterance in our language-game is: assert only those sentences which refer to the True (or express true thoughts); and the rule for language-game$_2$ reads just the same except for 'the False' being inserted where 'the True' occurs.[11] On the other hand, all that could be meant in speaking of judging a thought "successfully" is that one has judged a

---

[11]However, a person playing language-game$_2$ could never correctly describe the game this way precisely because it describes it truly.

thought that is in fact true, i.e. he judges that a is red and in fact a is red. Thus, "success" in judging a thought depends entirely on how the facts are, whereas success in assertive utterance contains an additional and "conventional" element concerning the accepted way in a given community to communicate to another a thought which is the case; in language-game$_2$ one can only do this by uttering assertively a sentence expressing a false thought, and so the true thought can only be gotten across by assertively uttering that sentence which expresses its negative. But a person playing language-game$_2$ will have just as much reason for wanting to "judge successfully" (judge only true thoughts) as a person in our language-game. Inasmuch as the members of a community wish to pass along to one another knowledge of various kinds, and since possession of such knowledge implies having judged various positive thoughts to be true, even a person in language-game$_2$ will strive to judge successfully. The only difference is that for him this knowledge must be communicated to other members via the pointless and unnecessarily circuitous route of assertively uttering a sentence which expresses a false thought. Because we wish to have definite knowledge (and the concepts in which such knowledge can be expressed) our interest is in those positive thoughts which are true.

Frege uses 'judge' in the same way as we generally use 'believe', i.e. in a sense in which to judge a thought is

ipso facto to judge it to be true. However there is a use of 'judge' in which one can sensibly speak of 'judging a thought to be false', just as there is a use of 'believe' wherein one can speak of 'believing a thought to be false'. "Judging S to be false" in this usage would be the same as "judging that not-S" in the first and primary one. Now the foregoing discussion might be objected to by saying that in language-game$_1$ our aim is to judge thoughts to be true, and in language-game$_2$ to judge them to be false, so that in a sense a participant of language-game$_2$ does not think in terms of what is true. But this objection misses the point; whatever thought is in this use of the term judged false by a participant in either language-game, its contradictory will also be judged true and vice-versa. And since the participants of both games desire conceptual knowledge and knowledge of how to do things, they will be concerned with judging a given pair of contradictories the same way. "Success" in judging in this latter use of 'judge' must be described in a different way, but in the end it comes to the same thing: one judges successfully if he judges a thought to have a certain truth-value, and it in fact has that truth-value. And this corresponds to our characterization of the first use: my judgement that a is F is successful just in case a is in fact F. Thus, "successful judging" in either use of 'judge' comes down to judging that which is the case, or judging "correctly", or even as we would say, judging truly.

For this reason, it cannot be objected that since the corresponding false, negative thoughts can equally serve as a basis for gaining conceptual knowledge and learning how to perform various activities, we can simply describe the participants of language-game$_2$ as aiming for these instead of the true, positive thoughts and let it go at that. Perhaps from a psychological point of view this would be correct, but since success in gaining conceptual knowledge and knowledge of how to do things consists in having successfully judged certain thoughts (or alternatively, successfully judged certain thoughts to have a certain truth-value), and success in judging in either of the two uses is simply judging correctly or truly, it remains a purely logical point that for a participant of any of the four language-games, his interest is in true thoughts. It may be correct to say of a participant of language-game$_2$ that he has succeeded in learning to play poker if he has learnt such negative thoughts as that a player is not dealt five cards are false. But this fact must properly be expressed either by saying that:

a. he has judged that five cards are dealt, and in fact five cards are dealt, or

b. he has judged the thought that five cards are not dealt to be false, and in fact it is false.

In either case, he can be said to have judged "successfully" or judged "truly". Success in judging can only consist in that which is judged being in accord with the facts--regardless of whether the "that which" is, as in (a), construed to

be "that five cards are dealt", or as in (b), to be "that the thought that five cards are not dealt is false." Regardless of which locution is used, the fact still remains that success in judging is correlated with truth or "being the case". To say that the judgment was successful is to say in the first instance that it is true (the case) that five cards are dealt, and in the second, that it is true (the case) that the thought that five cards are not dealt is false. Thus, in the primary use of 'judge' (the use expressed in (a)), even a participant of language-game$_2$ will have judged successfully if and only if the thought judged is true, and hence his interest, logically speaking, lies in true thoughts. From here on, we shall stick with the primary use.

Let us now try and tie these remarks together. It has already been pointed out that conceptual knowledge and knowledge of how to perform activities affecting other members of the community entail having successfully judged various thoughts, i.e. having judged thoughts which are in fact true. Moreover, this knowledge must be passed on by one member of the community to another through assertive utterance, and we have seen that for any community employing a given language, such knowledge can conceivably be communicated in two different ways according to which rule of correct assertion is adopted: either by asserting a sentence expressing a true thought, or by one expressing a false thought. In the

first language-game, success in assertive utterance is correlated with truth, in the latter with falsehood. Since a group of language users can, therefore, theoretically "choose" whether to aim for truth or falsehood in assertive utterance (there being no logical contradiction in supposing that they could correlate success with uttering a sentence expressing a false thought), and since in either case the speaker can succeed in communicating to another the thoughts he has judged, it follows that there is a "conventional" and "contingent" element in the notion of success in assertive utterance; "conventional" inasmuch as the members can decide to communicate to one another the thoughts they have judged via either mode of assertive utterance, and "contingent" inasmuch as it is, then, a matter of fact that they correlate success in assertive utterance with truth (again, they could have done otherwise).

Moreover, it would seem that a group of language users can decide to adopt a language-game wherein the speaker aims for falsehood only if they have already learnt and adopted the corresponding language-game in which truth is the aim. No child could naturally learn a game in which falsehood is the aim--rather he instead winds up learning the corresponding "negative language". The child would not, that is, learn to utter assertively a sentence only if the thought expressed is false, he would instead understand as the sense of that sentence the opposite thought, i.e. the one which is

true just in case the other is false and vice-versa. The acquisition of conceptual knowledge or learning the sense of a general term of the language used in the game--and hence thinking in terms of using that language and aiming for falsehood rather than learning and using the "negative language" and aiming for truth--presupposes having participated in a language-game in which truth is the aim. For it is surely wrong to suppose that the children of such a community could be taught the sense of 'red' (so that it means just what we mean by it), and then of their own accord go on to assert a sentence to the effect that something which would be correctly described as red by them and us (since both supposedly use English) is not red. But could they be taught the sense of 'red' that we as participants of language-game$_1$ have given it? To properly teach the child what 'red' means, the instructor, if he is playing language-game$_2$, must point to something red and tell the child it is not red, and such teaching could never be successful; the child would end up learning English-N and playing language-game$_3$. Moreover, children are never explicitly taught a rule for correct assertion as they learn to use language, but the foregoing seems to show that a child in a community which employed language-game$_2$ would have to be taught the assertion rule explicitly. To do so, however, the instructor would have to give him a true description of language-game$_2$ and thus would have to lapse into a language-game which aims for truth.

Hence, though language-game$_2$ can be construed as a possible description of a form of linguistic behavior, and while it could be correctly described as the behavior exhibited (perhaps as a diversion) by a group of persons who normally play a language-game wherein they aim for truth, it is wrong to suppose that there could be a community which employed language-game$_2$ as the natural, accepted mode of assertive utterance. Such a language-game then is derivative from and dependent upon one which aims for truth, and it is not hard to see why. For successful assertive utterance in the latter kind of game is "in accord with" success in judging.

We have further seen that though success in assertive utterance can be correlated with either truth or falsehood, the fact still remains that since knowledge involves having judged thoughts which are in fact the case, our ultimate aim in either of the two language-games is to communicate these thoughts to one another--in one by means of uttering sentences expressing true thoughts, in the other by uttering sentences expressing false ones. Inasmuch as we desire knowledge, our aim is to be successful in judging thoughts, and therefore it can readily be seen that the notion of success in judging does not contain any "conventional" or "contingent" element. To say that one has successfully judged the thought that a is F can only mean that he has judged that a is F and in fact a is F; that is, that the thought judged is true. It is not a contingent matter that success in

judging is "correlated" with truth, for it is simply nonsense to suppose that it could be correlated with false thoughts--just as it is nonsense to suppose that having knowledge could consist in believing something which is false.[12] Indeed, it would seem that the very idea of a community of persons acting in concert and using a common language by means of which knowledge can be acquired and transmitted presupposes an interest on the part of the members in judging thoughts which are the case, and hence "success" in judging can only mean judging a thought which is in fact true. It is not, then, either a conventional or contingent matter that our ultimate aim is for truth, nor, in view of the foregoing remarks on the derivative nature of language-games in which the speaker ostensibly aims for falsehood, that truth is correlated with success in assertive utterance as well.

---

[12] It must again be emphasized that 'judge' is being used here in its "primary" use, and just as 'A judges that S' parallels 'A believes that S', so likewise 'A successfully judges that S' parallels 'A knows that S'. And just as knowing that S has as necessary conditions that A believes that S, and S, so judging successfully that S has as necessary conditions that A judges that S, and S. Just as it is absurd to talk of "false knowledge", so it is equally absurd to talk of successfully judging a thought when in fact that thought is false or not the case. Of course it does make sense to speak of "knowing S to be false" or "knowing that it is false that S", and equally it makes sense to speak of "successfully judging S to be false" or "successfully judging that it is false that S". But here we have slipped back into the other use of 'know' and 'successfully judge' where knowing or successfully judging that S is true is not a redundant remark. And we saw earlier that even in this use, "success" in judging must be correlated with "being the case" or truth.

But Frege's theory makes it look as if it is a matter of fact; his theory makes it look as though there is only a contingent connection between truth qua "being the case" and truth qua what we aim for in assertive utterance. It seems on Frege's view to be a purely contingent matter that we aim to utter assertively only those sentences expressing thoughts which are the case. Just as it is a contingent matter that success in the coin example is correlated with the heads side facing up, so it is also contingent that success in assertive utterance is correlated with expressing a true thought. For a sentence can have one of two objects, the True or the False, as its reference, and just as we could have correlated tails with winning, so we could have correlated success in assertive utterance with sentences referring to the False, and thus aimed to express false thoughts. It is true that in the coin example our decision to let heads mean win was a purely arbitrary one, whereas in the case of assertive utterance, no such "decision", even on Frege's view, could simply be passed off as arbitrary. For Frege does after all recognize that we have an interest in thoughts which are the case, and there is no such prior interest on our part in heads rather than tails. But the difficulty is that Frege's theory makes it seem as if our interest in these thoughts is a contingent matter--as if it would semehow make sense to suppose seriously that our interest could lie in the other direction, that success in judging could consist in judging a thought which was not the case. For the eliminability

theory of truth and the doctrine of truth-values as the references of sentences tell us only that a true thought--one which "is the case"--is a thought expressed by a sentence having a certain object (the True) as its reference rather than a certain other object. And these objects are of logically equal status--there is no intrinsic difference between them any more than there is between heads and tails; nothing at any rate which would account for the fact that it is thoughts expressed by sentences referring to the True that we aim for. The fact that our interest lies in these thoughts--that they constitute success in judgment--appears then on Frege's theory to be, logically speaking, quite independent from and subsequent to the fact that it is just these thoughts that tell us how things are.[13] And again, that our interest is in true thoughts comes out looking like a purely contingent matter--as if it were possible to go in the other direction. Thus, on Frege's theory, it wrongly appears to be logically possible for those thoughts expressed by sentences referring to the False to constitute success in judging--as if the situation were analogous to correlating tails with winning rather than heads.

This result, it seems, shows that there is something seriously wrong with Frege's theory of assertion. Just where

---

[13] The phrase 'how things are' is used here in a broad sense in which even a true, negative thought tells us "how it is". In a narrower sense, the phrase would be restricted only to true, positive thoughts.

the mistake lies is not yet entirely clear, but inasmuch as we have produced a reductio against the theory, we know it cannot be wholly correct. "True" and "false" cannot simply be two mutually exclusive and jointly exhaustive categories for classifying thoughts,[14] for it is not a further and independent thesis that we aim for the thoughts in the first category rather than the second. Frege's theory creates a divorce between the thought ("content of a judgment") and the judgment itself which simply does not exist. We know that there is a distinction to be drawn between what is judged and the judging of it (and correspondingly, between what is assertively uttered and the uttering of it), and as Frege has said, one can surely entertain a thought without in any way having to "acknowledge its truth". But Frege somehow fails to draw the distinction properly; he makes the gap too large. Judging cannot consist in "advancing from a thought to a truth-value", as Frege intends this phrase, because it presupposes that we could conceivably "advance" to the other truth-value instead. The fact that it is not a contingent matter that we aim for truth in judging--and hence that success is essentially connected with being the case whereas there is no such connection between winning and heads--suggests that perhaps there is an assertoric element in the

---

[14] More precisely, "thoughts expressed by sentences whose component terms do not fail of reference."

"content" of a judgment which Frege has wrongly built into the judging relation itself. The same would then hold of sentences and assertive utterance as well. And indeed we find Wittgenstein in the Tractatus at 4.442 describing the assertion sign as "logically meaningless"; it only indicates, he says, that Frege (or whoever wrote it down) holds "the propositions marked with this sign to be true." More importantly, we have seen that at 4.022 he says: "A proposition shows how things stand if it is true, and it says that they do so stand." That is, a proposition or sentence not only expresses its sense or thought, but apparently also "asserts" it to hold; this latter point presumably refers to logical feature of the sentence as such, not something belonging to the "act" of asserting. And even a brief perusal of the Tractatus will convince anyone that Wittgenstein was not, in making these claims, falling into the old error of confusing assertion with predication. It was said earlier that Frege's philosophy of logic made it look like a remarkable and unexplainable fact that sentences have senses which admit of truth and falsity, for sentences are just one kind of value-expression among many, and that they are supposed to have the True and the False as their references in no way explains how it is that the senses of just these value expressions are bearers of truth and falsehood whereas the senses of other such expressions are not. That sentences come out looking this way on Frege's theory is a defect of the theory, and it

may just be, then, that Frege overlooks some sort of assertoric element inherent in the thought qua sense of a sentence, or in the sentence qua sentence--an element which would account for the fact that it is sentences, out of all other value-expressions, that can be used to say something true or false, and that it is thoughts as their senses which are the bearers of truth and falsehood.

A full examination of these questions must wait until later. In chapter four we will try to spell out just what it is that is wrong with Frege's theory of assertion--thus far, we have only speculated. In the meantime, it is necessary to return to the puzzling topic of negation, for in the preceding discussion of language-games wherein one aims for falsehood, we came across some interesting facts about "negative thoughts"--facts which, I think, show the inadequacy of Frege's remarks on negation. Having shown this, we will be in a better position to understand what is wrong with the theory of assertion, and indeed in Frege's philosophy of of logic as a whole. For I hope to make it clear later on that the logical concepts of negation and assertion are more subtly related than most writers on these subjects have heretofore realized.

Critique of Frege's Account of Negation

Earlier, an "independent criterion" for distinguishing positive from negative thoughts was put forth--"independent" inasmuch as they were distinguished independently of the kind

of sentence (positive or negative) of a given language by which they are expressed. And this as we saw means that in English, the thoughts expressed by singular positive sentences differ intrinsically from those expressed by singular negative ones. This intrinsic difference was spelled out as follows: a positive thought excludes all those alternatives except the one it picks out and leaves no alternative open, whereas a negative thought excludes only the alternative picked out and leaves open all the remaining ones. The technical terms 'alternative', 'excludes', etc. were defined on the basis of the "color-model" presented earlier, but it should be apparent that this independent criterion holds generally for singular sentences--not just those containing concept expressions which signify "determinates" of some "determinable".[15] And it would seem that an understanding of any given predicate or concept expression presupposes knowledge of a range of various incompatible predicates as in the color example.

---

[15] Of course the positive thought may not just exclude the remaining alternatives; going back to the color-model, saying that a is red may often seem to "exclude" a's not having any color at all (a might, for example, be an untinted pane of glass). However, for obvious reasons, it would be wrong to construe "a's being uncolored" as an "alternative". Rather, we ought to speak of "a's being colored" as a higher-order alternative, and the thought expressed by 'a is colored' would then exclude such further alternatives as being transparent, etc. Hence, because the thought that a is red entails the thought that a is colored, the former thought can be construed as automatically excluding the higher-order alternatives to being colored. Likewise, the thought that a is not red may be said to leave them open.

Clearly the independent criterion would also hold for species falling under a genus, and for the genera falling under a higher genus, and so forth. The situation is of course complicated by the fact that in ordinary speech the senses of must of our concept expressions are "open" or left indeterminate, and in such cases it will not be entirely clear as to just what is excluded by the positive thought. There may not even be a definite number of excluded alternatives just because of the openness of the corresponding concept expressions. It might just be that an individual, a, is a borderline case with respect to two concepts, F and G, which are not ordinarily taken to be compatible, and we might in this instance be willing to ascribe both F and G to it. In such a case, a's being F and a's being G will not be genuine alternatives. Are we really prepared in advance to say just what is excluded by, e.g. a's being a radio? Or its being on the living room table? Does being a radio exclude, for example, being a chair? But then there just could be a novelty item which we might describe as both a chair and a radio. Moreover, even in the color example, not every color predicate is incompatible with all the others; a's being maroon certainly does not rule out a's being red. But since these cases concern only the relative determinacy of the senses of the concept expressions, they do not affect the distinction between positive and negative thoughts--a positive thought specifies an alternative and excludes other alternatives, but

just what these alternatives are, or how many there are, or just how much ground each of them takes in is another matter.

Let us now explore further some of the consequences of the independent criterion for positive and negative thoughts. We shall suppose a language which is in every way like English--including 'not' as a truth-functional connective--but which, unlike English, contains only two "temperature predicates": 'ξ is hot' and 'ξ is cold'. There are no further ones like 'warm', 'cool', etc., and the two concepts are supposed to be jointly exhaustive of the entire range of hot and cold sensations. Moreover, they are independent, separately learnable concepts in the sense that one of them is not just the disguised complementary of the other.[16] Further, for any object or kind of object, x, which can be sensibly construed as falling under one of these concepts, the sentences '$\underline{x}$ is hot' and '$\underline{x}$ is cold' will express thoughts having different truth-values; if one is true, the other will be false, and vice-versa. There will of course be many "objects" (in Frege's broad sense of this term) which cannot fall under either concept, e.g. numbers, atoms, magnetic fields, etc. In such cases, the derived sentences, such as '14 is hot' and '14 is cold' will not be false, but express no thought at all. With respect to temperature, then, there are in this language only two alternatives for a given, physical

---

[16] This point will be discussed in some detail presently.

object--being hot and being cold--whereas in English there are several.

Now according to the independent criterion, we must say that '_a_ is hot' and '_a_ is not cold' differ in sense or express different thoughts, since they express different _kinds_ of thoughts--one positive, the other negative. The thoughts would indeed be true or false together, and the corresponding biconditional would be true. And generally, for any object, _x_, falling under the concept of temperature, the thoughts expressed by '_x_ is hot' and '_x_ is not cold' would be true or false together. Yet for all that, they would be distinct thoughts, for the former is essentially one which excludes a range of alternatives leaving no alternative open, while the other excludes only one and leaves open the remainder. Of course in this case since there are only two alternatives to begin with, both thoughts effectively exclude the very same alternative. But what is of importance here is that the first thought is of a kind which _can_ (and in English, does, since we divide up the thermal spectrum in a more discriminating manner) exclude _any_ number of alternatives, whereas the second is of a kind which as such can _never_ exclude more than one of a range of mutually exclusive alternatives. The people who would use this language simply do not discriminate very finely with respect to temperature, and hence there are only two alternatives. But there would still be an intrinsic difference between the thought that _a_ is hot and the thought

that a is not cold, though in this case the difference would not be of any real, practical significance.[17]

Yet certain puzzles remain. For it is not as though something's being hot and its not being cold is a purely contingent matter--the corresponding biconditional is not simply true as a matter of fact. For one might claim that in this language, the thought expressed by 'Everything is either hot or cold' is in some sense necessarily true. The relation between being hot and not being cold is not simply that which, say, holds between being a man and being a featherless biped. The thought expressed by 'if this is hot, then it is not cold' would not seem to be just contingently true as would be: 'if this is a man, then it is a featherless biped'. Nor on the other hand, would it be on a par with an explicit logical truth such as 'if this is a bachelor, then it is an unmarried man'. Thus it would be a mistake to say simply that it is a contingent matter that everything is either hot or cold. But even though it is not just contingent that sentences of the form 'x is hot' and 'x is not cold' express thoughts which are true or false together, it is nevertheless the case that

---

[17] Of course, if all of the concepts employed by a community came in exhaustive and exclusive pairs, there would be little point in expressing negative thoughts. Thus, the language employed would very likely not contain 'not' or any synonym, at least as far as singular sentences are concerned. However, this does not affect the point at issue. There is a logical distinction between positive and negative thoughts, though in the present example, one could get by with expressing only positive ones.

these thoughts are distinct. For what is a contingent matter is that the speakers of the language in question do not very finely discriminate with respect to temperature; the existence of English and other natural languages obviously shows that finer discriminations can be and are in fact made. And since in English, 'a is cold' and 'a is not warm' must express different thoughts (the difference being spelled out in terms of the independent criterion on the one hand, and by the senses of 'cold' and 'warm' on the other), the same must equally hold for 'a is hot' and 'a is not cold' in the language under discussion, even though hot and cold are the only temperature concepts admitted by it. The logical distinction between positive and negative thoughts holds in this language just as in English, for it is only a matter of fact that the speakers of this language discern just two temperature sensations.

I do not wish to embark here on a discussion of a priori knowledge; my point in mentioning all this is that it reveals, I think, a defect in Frege's account of negation, and indeed with the standard account given in introductory logic texts. I shall now try and explain.

We saw in the last chapter and again briefly in this one the importance of Frege's truth-functional account of negation, and with it the claim that, grammatical appearances to the contrary, negation attaches to sentences as a whole, not to predicative expressions. By construing negation as a

kind of function, and taking the unsaturated expression 'it is not the case that ξ' (or '¬ ξ') as representing this function, we no longer have to worry about such dubious entities as "negative concepts" of "negative properties", nor the many philosophical puzzles to which they give use. Thus, in the present case, 'not cold' does not signify a property or a concept at all, and we need not worry over whether the "negative concept" of "not being cold" is the same concept as being hot. There are only two concepts, hot and cold, and they are represented by the concept expressions 'ξ is hot' and 'ξ is cold' from which the positive singular sentences, 'a is hot' and 'a is cold' may be derived. And the negative sentence 'it is not the case that a is cold' is generated by inserting the positive sentence, qua argument-expression, in the gap of 'it is not the case that ξ'. As we have seen, 'ξ is not cold' would not, therefore, be a legitimate concept expression. But Frege has also told us that the thought expressed by a sentence is to be specified in terms of its truth-conditions, and in accordance with this, he said (phrasing it now without reference to the True or the False) that the thought expressed by a negative sentence, ⌜not -S-⌝ is that thought which is true when the thought expressed by S is false, and false when the thought expressed by S is true. But it should be clear by now that if this is all there is to specifying the thought expressed by a negative sentence, then 'a is not cold' and 'a is hot' would express the same

thought. For the thought expressed by 'a is hot' is also true whenever that expressed by 'a is cold' is false, and vice-versa. Since these thoughts are true under the same conditions, we seem forced to say that the thoughts are identical--something we have just seen to be false. As long as hot and cold are separate and distinct concepts, each of which can be taught and learnt independently of the other, then 'a is hot' and 'a is not cold' will express different thoughts, though on Frege's own principles we are forced to say that the thoughts are identical.

Of course they may not be distinct concepts. "Hot", for example, might just mean in effect "anything that is not cold", and hence one would be taught that 'cold' applies to a certain range of discernable sensations and then would be told: "All the rest is hot." And if this is all 'hot' means, viz. that part of the range outside the "cold" portion, then 'hot' would obviously function merely as a disguised complementary predicate--it would stand to 'cold' and 'unwise' stands to 'wise', and the positive sentence 'a is hot' would then express the same negative thought as 'a is not cold'. In such a case, the sense of 'hot' could not be learnt separately from 'cold', and 'ξ is hot' would not be a genuine concept expression. If 'hot' were just a disguised complementary, there would clearly be no possiblity of introducing a third temperature predicate incompatible with both 'hot' and 'cold', whereas if 'ξ is hot' is a genuine

concept expression and was taught and learnt in the same way as 'ξ is cold', there would be such a possibility. For example, the speakers of this language might conceivably experience some new sensation and be genuinely in doubt as to whether it is one or the other. In this case, the speakers of the language would have to make one of two choices: either rule that this sensation is hereafter to be included under one of the two concepts, and hence preserve their "range-exhausting" feature, or introduce a new temperature predicate to cover this unforeseen case--thereby creating a disparity between the ranges of application of 'hot' and 'not cold'. But only a difference in application is created, there must have already been a difference in thought expressed between 'a is hot' and 'a is not cold'. On the other hand, if 'hot' is merely the disguised complementary of 'cold', then the "range-exhausting" feature of 'cold' and 'hot' is given a priori, and no decision about the new sensation could in any way destroy this feature--the question of whether to preserve or give up the feature could not arise. The problem would simply be whether to specify with more exactness the sense of 'cold' so as to include the sensation in its range of application, or to specify it with more exactness so as to exclude it. Thus, if 'ξ is hot' and 'ξ is cold' are both genuine concept expressions the senses of which can be grasped independently of one another, then if they are range-exhaustive, it is only a contingent matter that they are. But

if 'hot' is just the complementary of 'cold' so that only 'ξ is cold' is a legitimate concept expression whose sense is grasped in the same way that 'red', 'blue', etc. are, then 'cold' and 'hot' (like 'red' and 'not red') must of necessity be range-exhaustive.

But to say that 'hot' is a disguised complementary is to say that the negative thought expressed by 'a is not cold' is the same as the thought expressed by 'a is hot', and generally, 'a is non-F' (or 'un-F', etc. as long as such prefixes are taken literally) expresses the same thought as 'a is not F'. Since this is so, we are then saying that hot and cold are only in fact range-exhaustive whereas cold and not-cold are of necessity range-exhaustive. However, it has already been said that there are no "negative concepts", and hence "not-cold" cannot be ranked alongside legitimate concepts like cold and hot. Now this is quite true, but the above remarks can be reformulated so as to avoid any reference to the "concept" of "not-being-cold", and we shall do this now.

Consider the following pairs of indefinitely indicating expressions:

A. x is cold x is not cold
B. x is cold x is hot

We will suppose that 'ξ is hot' and 'ξ is cold' are both legitimate concept expressions. Instead of speaking of the concepts hot and cold as range-exhaustive, we shall speak of the classes which are the "extensions" of each expression in

(B) as being jointly exhaustive of all objects falling under the concept of temperature. Since in the language under discussion, hot and cold are supposed to be range-exhaustive, then for any singular term which can meaningfully replace the variable, x, in the expressions comprising (B), the result will be two singular sentences expressing thoughts having opposite truth-value.[18] That is, for every singular sentence derived from one of the expressions in (B) and which expresses a true thought, the sentence derived from the other via the same substitution will express a false thought. Hence, the class of all objects which have temperature can be divided into two mutually exclusive and jointly exhaustive sub-classes--those which are truly described as cold, and those which are truly described as hot. We shall hereafter describe this by saying that the pair of expressions in (B) jointly exhaust the class of objects falling under the concept temperature.

Now it is still possible that one might come across an object, d, which yields a "borderline" sensation--one that

---

[18] The sentences so derived will either express thoughts having different truth-values or else commit category mistakes, and hence express no thought at all. The hot-cold example has been so constructed that no meaningful pair of sentences can be derived both of which express false thoughts. At least, I hope it has. But even if there are counterexamples, the point of it would remain untouched, for it seems possible, at least in principle, to specify two opposing, "exhaustive" properties which can then be used to show the inadequacy of the Fregean account of negation.

cannot flatly be described as hot nor as cold. Here, the speakers of the language would be confronted with the following theoretical decision: either to sharpen the senses of 'ξ is hot' and 'ξ is cold' by including this new sensation within the range of application of one of them--and thus preserve the "jointly exhaustive" feature of the expressions in (B)--or otherwise, introduce a new temperature concept to apply to the new case--thus destroying the jointly exhaustive feature. In the latter case, a third "alternative" with respect to temperature is introduced, and the thoughts expressed by '$\underline{d}$ is hot' and '$\underline{d}$ is cold' will both be false. Of course if hot and cold had been sharply defined in the first place, no such question would have arisen, but it would not have arisen only because they had been so defined.

Now the expressions in (A) are also jointly exhaustive of the class of objects which have temperature, and clearly the sub-class of objects which can truly be said to be not cold is identical with the sub-class which is the extension of '$\underline{x}$ is hot'. But if we now reconsider our borderline case, it is apparent that we are not faced with the same choice as before. There can be no question of giving up the jointly exhaustive feature here as there was with the expressions in (B). For what is essential in negation is that the jointly exhaustive feature is given a priori and no decision the speakers could make would have the effect of removing this feature. The only question would be whether to specify the

sense of 'ξ is cold' more sharply so as to include the new sensation or so as to exclude it. The question of giving up the jointly exhaustive feature does not arise, not because of having sharply specified the sense of 'ξ is cold', but simply because of the occurrence of 'not'. The jointly exhaustive feature belongs to pair (A) necessarily--the question of giving it up can never arise with respect to any pair of expressions of the forms 'x is F' and 'x is not F'. It matters not which concept expression we substitute for 'F' and it is quite irrelevant as to how sharply or fuzzily its sense has been specified. But if the jointly exhaustive feature belongs to (B), it does so only because of the following "factual" considerations: first, 'ξ is cold' and 'ξ is hot' are the only concept expressions in the language which signify determinates of temperature, and secondly, their senses have been so sharply defined as to prevent any borderline cases from arising--cases which, if they did arise, could be resolved by introducing a new temperature concept, thereby undermining the jointly exhaustive feature of pair (B).

These considerations would seem, then, to provide us with a new and perhaps less "theory-laden" basis for saying that the thoughts expressed by 'a is hot' and 'a is not cold' are distinct. Previously, we had relied on the independent criterion to do this job, but now, because of the disparity between pairs (A) and (B), it appears that any pair of sentences derived from 'x is hot' and 'x is not cold' via the

substitution of the same singular term will express different thoughts, despite the fact that for any such substitution, the same truth-value accrues to each member of the pair. For though both pairs are jointly exhaustive, we have noted some significant, "intensional" differences between them. Jointly exhausting the class of objects falling under the concept temperature is built-in, a priori feature of (A), but in (B) that feature could conceivably be removed by introducing a third determinate to cover what would otherwise be borderline cases. In (A), the jointly exhaustive feature stems solely from the occurrence of 'not'--we do not even need to know the sense of 'ξ is cold', much less whether or not it is sharply defined. For we know generally that for any concept, F, the expressions 'x is F' and 'x is not F' will jointly exhaust the class of objects falling under the determinable (or genus, etc.) of which F is a determinate (or species, etc.). Such considerations, together with the related point that the thought expressed by 'For all x, x is hot if and only if x is not cold' cannot be construed simply as a logical truth, surely indicates that 'x is hot' does not stand to 'x is cold' in the same purely logical way as 'x is not cold', and hence, extensional similarities notwithstanding, a pair of sentences like 'a is hot' and 'a is not cold' express different thoughts. Granted they are distinct, the independent criterion may now be brought in as an explanation of this fact; it shows us why they are distinct by

bringing out the inherent, logical differences (in terms of excluding and leaving open) between positive and negative thoughts as such--differences ignored in the standard, extensional account of negation.

Now we can restate our earlier criticism of Frege's account of negation in a more rigorous manner. In a language in which 'ξ is hot' and 'ξ is cold' are both legitimate, separately learnable concept expressions, and in which the pair of expressions '_x_ is hot' and '_x_ is cold' jointly exhaust the class of objects falling under the determinable of temperature (owing to the "factual" considerations discussed earlier), Frege's account of negation, together with the principle that a thought is specified by stating the conditions under which it would be true, entails that the sentences '_a_ is hot' and '_a_ is not cold' express the same thought, since the thought expressed by each will be true whenever the thought expressed by '_a_ is cold' is false and _vice-versa_. But as we have seen, they clearly do not express the same thought. This result has some important consequences for the Fregean or standard truth-functional account of negation, and we shall turn to these now.

Negation and the Independent Criterion

An adequate account of negation should uniquely characterize the relation between the thought expressed by a sentence of the form '_a_ is F' and the (one and only one) thought which is its "negative" and which is expressed by '_a_ is not F'

(and alternate forms like '_a_ is non-F'). If we accept a Fregean functional analysis of negation, then we can restate this as follows: an adequate account of negation should uniquely characterize the relation between the thought expressed by '_a_ is F' and the (one and only one) thought which is its negative and is expressed by that sentence which is a value-expression derived from 'it is not the case that ξ' for the argument-expression '_a_ is F'--in other words, the thought expressed by '_a_ is not F'. Now '_a_ is hot' is first of all not a value-expression derived from 'it is not the case that ξ' and secondly expresses a thought distinct from that expressed by '_a_ is not cold'--a sentence which _is_ so derived. Therefore, the thought that _a_ is hot, though it meets Frege's conditions for being the negative of the thought that _a_ is cold, simply is not its negative, and Frege's account is thus insufficient inasmuch as it fails to characterize _uniquely_ the relation between a thought and its negative. The thought that _a_ is not cold is clearly the negative of the thought that _a_ is cold, but it is not the only thought that is true just in case the thought that _a_ is cold is false and _vice-versa_. And if we agree that every thought has just one negative, then only the thought that _a_ is not cold can properly be said to be the negative of the thought that _a_ is cold, and Frege's account does not succeed in distinguishing it from all others.

This is not to say that Frege's definition is to be rejected outright; rather, it simply does not go far enough. Certainly one cannot deny that any pair of thoughts, P and not-P, will always have opposite truth-value. The trouble is that we can in principle produce pairs of thoughts which are distinct and yet both meet Frege's conditions for being the negative of the thought that a is cold. Therefore, either we must say that Frege's definition is satisfactory and then acquiesce in the counterintuitive claim that both the thought that a is hot and the thought that a is not cold are "negatives" of the thought that a is cold, or if we are willing to admit only the latter thought as the negative, face the fact that Frege's account is lacking. Now it seems to me that the former alternative must be rejected, for an essential ingredient of our notion of negation (and especially in terms of the expression given it in logic texts) is that every thought has just one negative. That is, with the exception of esoteric cases like the hot-cold example, there will be, for any thought, P, just one other thought,not-P, which will be true just in case P is false, etc. The fact that there is more than one in the esoteric case which meets this condition is easily explainable--there are, in the language in question, just two determinates of temperature. And because of this we would all, I think, be willing to call only the thought that a is not cold the negative. Moreover, on a truth-functional account, only 'a is not cold' is a value-

expression derived from 'it is not the case that ξ', and thus we would want to say that only the thought it expresses could properly be the negative of the thought that a is cold. Thus we must find an account of negation that will justify our basic presupposition that every thought has only one negative in light of the fact that merely saying that the negative of P is true if and only if P is false and vice-versa will not do.[19]

The problem, then, with a purely extensional, truth-functional account such as Frege's is that it only specifies a relation of contradictoriness between the thoughts, P and not-P, a relation which could conceivably hold between P and Q where Q is a thought we should not wish to call the negative of P. But merely to state a relation of truth-values between two thoughts is to ignore any logically significant, internal features of the thoughts so related--features which might enable us to provide a satisfactory account of negation. To say that the thoughts P and Q are contradictories is to

---

[19] We are concerned here, in effect, with the sense of the functional expression 'it is not the case that ξ', for we already know how its reference (the concept) is specified. In the Fregean system, it is specified in a purely extensional manner by giving the values it takes for its various arguments. Knowing this much, we know (to revert to Frege's own statement of it) that if a sentence, S, stands for the True, then ⌜¬S⌝ will stand for the False and vice-versa. Correspondingly, if the thought expressed by S is true, then the thought expressed by ⌜¬S⌝ will be false, etc. But we have also seen that knowing this much is not enough to specify the sense of ⌜¬S⌝, i.e., the thought it expresses. Thus, whatever else is needed, it must be contained in the sense of ⌜¬ξ⌝ or 'it is not the case that ξ'.

say nothing about whatever further logical characteristics P and Q may have, and there is no immediate reason why we should interpret Q to be not-P rather than some other thought. And if we now define not-P in terms of the contradictory relation (by stating the usual, truth-functional definition or drawing the appropriate truth-table), we have every right to ask what assurance there is that only that thought can stand in the contradictory relation to P; the hot-cold example shows us we have no such assurance, and hence have not succeeded in uniquely characterizing not-P. Indeed, merely establishing a contradictory relation between P and not-P not only ignores any internal, non-extensional differences between P and not-P, but, since contradictoriness is a symmetrical relation, it invites us to ignore them. For it makes it look as though P and not-P are on an equal, logical footing--that since they stand to one another with respect to truth-value as heads stands to tails, then just as heads and tails are on a logical par with each other, so are P and not-P. Just as P is true just in case not-P is false and vice-versa, so likewise the heads side faces upward just in case the tails side faces downward and vice-versa.

But this is a bad picture. P and not-P are no more analogous to heads and tails than were "true" and "false". For as we will presently see there are significant, internal features of P and of not-P of which there are no analogues in the case of heads and tails. When not-P is defined as

that thought which is true when P is false, etc., we are in effect identifying the relation 'is the negative of' with the contradictory relation. Now in a way, the former relation, like the latter, is symmetrical; not-P is the negative of P, and, also, P can be said to be the negative of not-P inasmuch as not-not-P is the same thought as P, and not-not-P can also be said to be the negative of not-P. In other words, the sentences S and ⌜not-not-S⌝ express the same thought, and ⌜not-not-S⌝ is derived from ⌜not-S⌝ via the same procedure as ⌜not-S⌝ is derived from S. But if what has been said already is correct, we cannot simply identify the negative relation with the contradictory one, for the Fregean definition says in effect: there is one and only one thought which stands in the contradictory relation to P, and if the sentence S expresses P, then this thought is expressed by ⌜not-S⌝. And as the hot-cold example shows, it is possible for there to be another thought which stands in that relation to P, and which we should not wish to term the negative of P. Now what else is there in the negative relation that is not contained in the contradictory relation?

The independent criterion has already been invoked to show that the thoughts expressed by 'a is hot' and 'a is not cold' are distinct, so it should not be surprising that it provides us with the additional element needed to characterize the relation between P and not-P. For that criterion gave us a basis for distinguishing between positive and negative

thoughts as such--not in terms of the kinds of sentences expressing them, but in terms of certain intrinsic, logical characteristics. The thought that not-P is not simply the contradictory of P (for P is equally the contradictory of not-P), rather there is something which makes a negative thought negative--something which makes not-P an intrinsically different kind of thought from P. By the same token, then, the thought that P can be described as positive in virtue of its own logical characteristics. There is no parallel here with heads and tails; as sides of a coin, they are perfectly alike (except for the incidental feature of having certain characteristic markings by virtue of which one side is called "heads" and the other "tails"). Now we have already seen what these characteristics are: the positive thought describes or specifies a definite, possible alternative, and excludes all other alternatives leaving nothing open; and the negative thought, since it must be expressed by a complex sentence containing the sentence expressing the positive thought as a component, specifies that same alternative, but excludes it and leaves open all the remaining ones. The negative thought, then, leaves open just what the positive thought excludes--it does not specify a different alternative from P, for unlike the hot-cold and heads-tails examples, not-P does not tell us what is the case instead of P. Rather, the positive and negative thoughts are concerned with the very same alternative, but they treat of it in

different, though complementary, ways. It is very much like an enclosed figure on a plane; if we shade the indefinite area outside the figure (the shading to correspond to the logical feature of excluding), then we would have a picture of the positive thought, whereas if we shade only the inscribed area, this would be analogous to excluding the specified alternative, and by leaving blank the area outside, we would be leaving open all other alternatives (these alternatives corresponding to possible figures that could be drawn in the exterior area). In this case, then, we would have a picture of the negative thought.[20]

Thus, the thoughts, P and not-P, are not on an equal, logical par with one another as are heads and tails, or even the thoughts that a is hot and a is cold. There are "intensional" differences between them which are not brought out by the standard, Fregean account--a purely extensional account which only specifies a relation of contradictoriness. But the independent criterion enables us to spell out these

---

[20] This analogy is adapted from a remark in Wittgenstein's Notebooks, although he does not use the terms 'excludes' and 'leaves open'. To complete the analogy, we would have to in some way mark the inside area in the picture of the positive thought. For if it is left blank, this would amount to leaving open the specified alternative, and of course the positive thought does not do this. This fact will be discussed at some length in the next chapter, but since we are here concerned only with a critique of Frege, it may be ignored for the present. What is important here are the complementary ways in which the positive and negative thoughts exclude, and that what is excluded by the former is left open by the latter.

differences in terms of the features of excluding and leaving open. Hence, we can now, at least in part, see why not-P cannot be grasped independently of P; that is, why, unlike the thoughts that a is hot and a is cold, the negative of P is characteristically expressed by a sentence containing the sentence expressing P as a component. For not-P must specify the same alternative as P, whereas the thought that a is cold, qua positive, must specify a different alternative from that specified by the thought that a is hot, and these thoughts are grasped independently of one another inasmuch as the concepts hot and cold are separately learnable. Moreover, we also saw that in order to know that sentences of the forms 'x is hot' and 'x is cold' express thoughts which stand in the contradictory relation, we must first know the senses of 'ξ is hot' and 'ξ is cold', whereas negation can be grasped in a purely general manner. We do not need to know the sense of 'F' in order to know that sentences of the forms 'x is F' and 'x is not F'will express thoughts which are contradictories. The independent criterion brings out this feature inasmuch as the excluding and leaving-open features illustrate how the sense of any sentence is altered when 'not' is inserted in it. The criterion provides us with a distinction between two intrinsically different kinds of thoughts; whatever expressions replace 'x' and 'F', the thoughts expressed by the resulting sentences will be of these different kinds. But the thoughts that a is hot and a is cold are both of the same kind, viz. positive.

Thus, while P and not-P stand in the contradictory relation just as heads and tails stand in the "up-down" relation, the comparison between the two cases can be misleading. For there are further, logically significant, and complementary characteristics of P and of not-P which have no analogue in the heads-tails case. Not-P is the contradictory of P (and vice-versa), and it is also the negative of P (and vice-versa). But the characteristics of excluding and leaving open allow us to describe not-P as negative per se--not just as the negative of P--and for describing P as positive, though it may still be legitimately described as the negative of not-P. Moreover, the Law of Double Negation becomes easily intelligible. For what happens when 'not' is inserted into a sentence, S, is to produce a thought which excludes just the opposite of what the thought expressed by S excludes. To return to our illustration, in the positive thought, the area outside the figure is shaded, whereas in the negative thought, the interior is shaded. Thus, when a further 'not' is inserted in the sentence in question, a thought is produced which excludes the opposite of what the thought expressed by ⌜not-S⌝ excludes, and this can only be the same as what the original thought excluded; in other words, the exterior area is once again shaded.

So the independent criterion, which enables us to distinguish between the thoughts that a is hot and a is not cold, and thus to show the insufficiency of Frege's account

of negation, also provides us with the additional element needed to give an adequate account of what it is to be the negative of P--it uniquely characterizes the relation between P and not-P. Indeed, it shows us why the negative of P ought to be construed as a function of P--why the sentence expressing the negative thought is a value-expression with the sentence expressing P as the component argument expression--in other words, why negation is truth-functional at all. The usual, extensional account of negation is not to be rejected outright, for it is an essential feature in negation, and it is surely sufficient for working with the propositional calculus. Thus, from a pragmatic standpoint, there is nothing wrong with the explanation we are given in introductory logic texts. But as a definition, it fails to characterize uniquely the negative of P, and even if we could not produce any cases like the hot-cold example where there are two distinct thoughts which stand in the contradictory relation to P, it would still be reasonable to ask what grounds there are for speaking of the negative of P--of why, in the normal case, there is just one thought that is true when P is false, etc. And only the independent criterion will give us an answer.

Consequences of the Independent Criterion

Earlier in this chapter, Frege's account of negation was critieized on the basis of Wittgenstein's remark at 4.431 of the Tractatus. If truth-values are references, and specifying

a reference is not sufficient for specifying a sense, then the sense of a negative sentence can never be specified in the way Frege tried to do it, viz. by stating under what conditions the negative sentence would stand for the True. The truth-condition principle conflicts with the doctrine of truth-values as references, and it was argued that since we do in fact specify the sense of a negative sentence by stating its truth-conditions, we must reject the principle that truth-values are the references of sentences--and, in the end, that sentences have references at all.

It might now seem that this result is imperiled insofar as the foregoing discussion of negation appears to cast doubt on the truth-condition principle. For if the sentences '<u>a</u> is hot' and '<u>a</u> is not cold' express different thoughts, but are both true when the thought that <u>a</u> is cold is false and <u>vice-versa</u>, doesn't this show that the two thoughts can have the same truth-conditions without being identical? The answer is no. It must be remembered that the truth-condition principle was introduced to explicate the senses of <u>compound</u> sentences and sentences containing quantifier terms like 'all' and 'some'. The thought expressed by the compound sentence $\ulcorner S_1$ and $S_2 \urcorner$ is that thought which is true just in case both the thought expressed by S_1 and that expressed by S_2 are true. The thought expressed by 'Everything is red' is that thought which is true if and only if for every singular term inserted in the gap of 'ξ is red',

the resulting sentence expresses a true thought. But an ordinary, positive, subject-predicate sentence involving no logical connectives or quantifier terms is a different matter, as we saw in the last chapter. When we state the truth-conditions for a compound sentence or one containing a quantifier term, we are in effect giving the general procedure by which the sense of the logical connective or quantifier term is specified--in the two examples just cited, of 'and' and 'every' respectively. And a positive, uncompounded sentence like '_a_ is hot' does not contain any such expressions; rather, such sentences serve as the building blocks in terms of which the truth-conditions of compound sentences (or better, the thoughts they express) are stated. Thus, a simple sentence like '_a_ is hot' does not have any truth-conditions as such; the best we can do is give the rather trivial and circular description of it as the thought which is true just in case the object referred to by the subject falls under the concept _hot_--in effect, that the specified alternative obtains. But the thought expressed by '_a_ is not cold' does have truth-conditions which are spelled out in the customary definition of 'not'. Now the independent criterion distinguished between two different kinds of thoughts--positive and negative--and the latter is expressed by a sentence containing the sentence expressing P as argument-expression; that is, by a compound sentence containing a logical connective. Now granted that the truth-condition principle applies only to

thoughts expressed by compound sentences--and hence not to those expressed by simple positive sentences--then stating the truth-conditions for a negative thought will suffice to distinguish that thought from all others and to identify it as the negative of the corresponding positive one, for the independent criterion, by making the preliminary distinction between positive and negative thoughts as such, rules out any application of the principle to the thought expressed by a sentence like 'a is hot'. This thought will in fact be true just in case the thought that a is cold is false and vice-versa, but it is so because of the "non-logical" considerations mentioned earlier; to wit, because in this language, there are only two determinates of temperature. So the independent criterion, which enabled us to distinguish between the thoughts that a is hot and a is not cold in the first place, also preserves the truth-condition principle by setting the limits of its applicability--limiting it only to a specification of the latter thought. And it does this by first distinguishing between two logically different, though complementary, kinds of thoughts, and secondly illustrating why the negative variety is expressed by a compound sentence containing a logical connective together with the sentence expressing the corresponding positive thought as argument-expression. Thus, the reductio argument developed earlier against Frege's account of negation still holds good. For as long as we rely on just the truth-condition principle,

we will not, on Frege's definition, be able to distinguish the thoughts that a is hot and a is not cold; thoughts which we have seen on quite separate gounds to be distinct. But once the truth-condition principle is supplemented by the independent criterion, then not only can it be seen why these thoughts are distinct, but it can also be seen why the principle can properly apply only to those thoughts in the negative category--thus vindicating the principle from the charge that it does not do the job it was intended for.

But now another difficulty arises. Wittgenstein's critique of Frege said in affect that if truth-values are the references of sentences, and if specifying a reference is not sufficient for specifying a sense, and if specifying the sense of a sentence consists in stating its truth-conditions, then Frege's account of negation, in terms of stating the truth-conditions of not-P, does not succeed in specifying not-P. And it was then argued that since we do succeed in specifying not-P by giving its truth-conditions, we must reject the premiss that the True and the False are references of sentences. But now it may be objected that since an account of negation solely in terms of truth-conditions is insufficient to characterize not-P, then it is the unsupplemented truth-condition principle that must be at fault, not the premiss that truth-values are the references of sentences. All Frege need do in order to render his account of negation adequate is insert the independent criterion--there

is no longer any reason why he must also reject the principle that truth-values are what sentences refer to. In other words, by bolstering the truth-condition principle with the independent criterion, the conclusion of Wittgenstein's reductio no longer follows, and hence there is no reason to reject any of the premisses.

But this does not follow. Although we have now (hopefully) succeeded in uniquely characterizing the negative of P, we must still reject the principle that sentences are a kind of referring expression naming either the True or the False. For in speaking of the "truth-condition" principle as applied to negation, we spoke of the thought that not-P is true, etc., and 'thought' and 'sense of a sentence' have been used throughout to mean what is often meant by other writers as a proposition or statement--an intensional entity distinct from the sentence expressing it and which is the bearer of truth-value. Thus, when the truth-condition principle for negation is stated in terms of thoughts being true or false, we are stating it in a relatively neutral manner. However, as we have seen, Frege's primary characterization of the truth-condition principle (and the one Wittengenstein obviously had in mind) is in terms of sentences referring to certain objects, the True and the False. And this characterization of the truth-condition principle is still ruled out by Wittgenstein's critique. The "neutral" statement of the truth-condition principle for negation does not by itself

suffice to characterize the negative of P, but insofar as it sets up the contradictory relation, it is nevertheless an essential ingredient. But Frege's version of it does not even go this far, for it still remains true that a sense cannot be specified by specifying a reference, and if we continue to think of truth-values as objects designated by sentences, and not simply as that which can be uniquely ascribed only to those senses which are expressed by sentences, then to attempt to specify a thought as the negative of another by means of stating its truth-conditions, where this involves specifying what the reference of the sentence expressing it will be, must end up a failure. I can no more specify the sense expressed by the sentence 'not-S' by specifying what its reference is given the reference of S than I can identify the sense of 'the capital of England' merely by being told that it refers to London since 'England' refers to England. The principle that truth-values are the references of sentences must be rejected on the grounds that when the truth-condition principle is formulated in those terms, it is rendered impotent; and inasmuch as it provides us with the means for characterizing the sense of a compound sentence, it certainly is not impotent.

Final Comments

The conclusions drawn in this chapter concerning negation have been developed not simply to point up the insufficiency of the customary, truth-functional account of negation, but

because the independent criterion, as I hope to show in the next chapter, will enable us to see more clearly just where Frege's theory of assertion has gone awry. In doing so, a better theory will be suggested, one derived mainly from some of Wittgenstein's remarks in the Tractatus and the Investigations, but which does not entail the atomistic ontology of his earlier thought. Before delving into this, however, one further remark on negation is in order.

It ought to be quite apparent by now that if the description provided by the independent criterion of a negative thought is included in the definition of 'the negative of P', then that definition strictly speaking becomes circular. For the technical terms 'excludes', 'leaves open', etc. as they were introduced obviously require the use of 'not' in their explication. Moreover, the various statements made in connection with the color-model clearly presuppose the principles of Excluded Middle and Non-contradiction. However, I think it would be wrong for us to expect otherwise; once we move beyond the contradictory relation and try to spell out the intrinsic, logical features of positive and negative thoughts as such, a circular account becomes unavoidable. This ought to become clearer in the next chapter when we will return to Frege's remarks on assertion and truth. But even at this stage, it begins to look as though negation is a fundamental, primitive concept incapable of strict definition. Like the notion of a function, it can be fully

explained only by means of examples, analogies, and metaphors. So while negation cannot strictly be defined, the independent criterion can nevertheless be employed to elucidate it, much in the same way as the remarks at the beginning of the last chapter can be said to "elucidate" the notion of a function.

It is the mystery of negation: This is not how things are, and yet we can say how things are not.

Wittgenstein, Notebooks, page 30.

CHAPTER THREE

NEGATION, PICTURES AND SINGULAR SENTENCES

In this chapter we shall continue our discussion of negation and attempt to get a more complete understanding of that difficult concept. This will of necessity require us to undertake a more general examination of the nature of singular, uncompounded sentences and the senses they commonly express. The results obtained here were suggested by certain observations made by Wittgenstein in his Tractatus Logico-Philosophicus and Notebooks; 1914-1916. These results will then be utilized in the next chapter to gain a better understanding of the concept of assertion, and more specifically to show what is wrong with Frege's remarks on sentences, thoughts, and assertions.

A Problem Concerning Uninterpreted Calculi

In Symbolic Logic,[1] Lewis and Langford claim that there is no "logical significance" to the matrix method (i.e. the uninterpreted calculus which, when interpreted "logically",

---

[1] C. I. Lewis and C. H. Langford, Symbolic Logic (New York: Appleton-Century-Crofts, 1932), page 227.

becomes the truth-functional calculus). They consider a matrix relation "I" defined (partially) by the following table:

p	q	pIq
1	1	1
1	0	0

They then say: "any inference, q, based on the fact that 'p is assertable and pIq holds' will be a true proposition." And of course this is to say that the proposition: 'If p is assertable and pIq holds, then q is assertable' is analytic. They go on to say:

> This does not depend upon any peculiar "logical" significance of such a relation, pIq. Whatever more it may be, the matrix method at least is a kind of game which we play with recognizable marks, according to certain rules. If in any game which deals with any kind of things, p's and q's, there is an operation or move, pIq, which according to the rules can be taken, when p has the property A, only if q also has the property A, then ipso facto if p has A and pIq is allowable, then q has A.

In saying that the matrix method has no "logical" significance, Lewis and Langford are claiming that the uninterpreted calculus can be interpreted "non-logically"; letters like 'p', 'q', etc. can stand for "any kinds of things"--they do not have to stand for propositions or sentences, and '1' and '0' as they occur in the above table can stand for the presence or absence, respectively, of any property--they do not have to stand for the presence or absence of the property of being true. However, their view also requires that there be a game

in which pIq is a move. Moreover, this move can be taken (i. e. is an allowable move) only when both p and q have the property A. Now it follows from all this that the following claim will be analytic:

S: If p has A, then q has A if and only if pIq is an allowable move.

By showing in this manner that the matrix method need not have explicated the notion of logical truth, i.e. they want to claim that this notion can be explained solely in terms of rules for the manipulation of certain symbols. By doing this, they presumably hope to dispel any rationalistic notion that knowledge of logical truths is some special, transcendent kind of knowledge, possessing greater philosophic interest than knowledge of contingent truths.

However, this view concerning the nature of the uninterpreted matrix system has been subjected to a telling criticism by Geach,[2] and which has been reiterated by Anscombe.[3] It is, of course, true that the above table does not have to be interpreted logically, but Lewis and Langford nevertheless go on to say that there must be a game such that (S) holds analytically. But now we have an inconsistency; for the 1's in the 'p' and 'q' columns indicate the presence of any given property A, while the '1' in the pIq columns indicates the

---

[2] "Ifs and Ands", in Analysis, vol. 9, 1948.

[3] Introduction, pages 54-55.

presence of a particular property, viz. that of being an allowable move in some game or other. Lewis and Langford have, by their own description of the matrix in terms of a game, tacitly interpreted the figures in the 'pIq' column while leaving uninterpreted the figures in the other columns. Of course we could remove the inconsistency by taking A to be the property of being an allowable move, but first, p and q would no longer be "any kinds of things" (i.e. we would no longer have an uninterpreted calculus), and second, there is obviously no necessity in so interpreting the system. Geach constructs an example of considerable interest in its own right:

> Let the figures '1' and 'o' connote the presence or absence of an hereditary property A, and let 'pIq' mean 'offspring by p out of q'. Then the table will have no reference to moves in any game. It will state that the trait A is present in the offspring, when it's present in the sire, if and only if it is present in the dam...

The important point to be drawn here is that on this interpretation, the following will not be analytic:

S_1: If p has A, then q has A if and only if pIq has A.[4]

Moreover, it would seem that if we were to interpret the matrix as talking about moves in a game, (S_1) again would not belong to the terminology of any particular game, rather as

---

[4] It is, of course, being assumed here both by Geach and myself that contingently hereditary properties are being considered as interpretations of 'A'.

Geach says: "It expresses what we may fittingly call a logical relation of the move pIq to the moves p and q, so that the significance of the notation 'pIq' is after all "logical". 'I' expresses no such relation in the biological example; the relation it expresses clearly does belong to the terminology of a particular "game".

So Lewis and Langford were apparently correct in saying that (S) is analytic (assuming, of course, we read each occurrence of 'A' as 'is an allowable move'), but for the wrong reasons. They thought of an uninterpreted calculus as a scheme or formula for all kinds of games, whereas we have seen that a calculus can be such a formula only if we interpret it that way. An uninterpreted calculus is simply a set of marks with rules for combining them, but these rules do not suffice to give us a formula for games, and (S_1), as it stands in the uninterpreted calculus, will be neither analytic nor synthetic. For it is just a skeleton which does not have a truth-value until the marks 'p', 'l', etc. are interpreted. The rules in an uninterpreted system are just "formation" rules--they tell us that combinations such as 'pIq' will count as (syntactically) meaningful, and others such as 'pqI'[5]

---

[5] This is not to say that there will not be transformation rules, but rather that such "rules"--specified in a purely syntactical fashion--only become rules of inference when the calculus is given a "logical" interpretation. The rule that given a two-item list of wffs of the form 'p' and 'pIq', we may write a third item 'q', surely cannot be a rule of inference if the calculus is given the biological interpretation suggested by Geach, for none of these items are propositions--they are animals.

will not. They do not give us any information as to whether or not 'I' expresses a logical relation.

Anscombe on Negation

In chapter three of her book on the Tractatus, Miss Anscombe expounds this critique of Lewis and Langford in connection with the question of justifying the claim that every thought (or "proposition" as she puts it) has one and only one negative. Now it should be clear from the preceding chapter that the independent criterion may be brought in here, but we shall for the present ignore this fact. Suppose a skeptical student in an introductory logic course were to ask what grounds there are for supposing that any given thought will have just one other thought as its contradictory.[6] Suppose further that he is not to be put off by our appealing to his intuitions on the matter, or by claiming it is obvious or "self-evident"--he simply says that it is not self-evident to him and that he wants some "proof". One thing we might do to

---

[6]This is the form in which Miss Anscombe raises the question, and she seems to feel that asking for the negative of a proposition and asking why it has just one contradictory are one and the same thing. She would argue that every proposition does in fact have only one contradictory, but that this fact requires explanation. But in the last chapter, it was argued that it is possible to construct cases where a given proposition has two contradictories although only one of them can properly be said to be its negative. However, even though I do not feel that Miss Anscombe has raised the problem correctly, I shall not discuss this issue any further here, for my purpose is to examine some of her subsequent remarks on the problem at hand.

alleviate his doubts is point out that every proposition has exactly two truth-values just as every coin has exactly two sides. Then we could show him that we can have only two possible significant bets for a given coin:

coin	bet 1	bet 2
H	W	L
T	L	W

Next, we can compare this table with the following truth-table:

p	p	~p
T	T	F
F	F	T

and then show our student that there are only two possible significant combinations of T's and F's; and this shows that the proposition corresponding to the one combination meets the conditions laid down for being the negation of the other (and vice-versa), and since there are no other possibilities, it is the only proposition meeting those conditions. However, our student is not easily satisfied; he quite correctly states that we would not dream of using the same pair of symbols in all three columns--we certainly would not use 'H' and 'T' in the bet columns--why then do we use the same symbols, 'T' and 'F', throughout in the truth-table columns? It will not, of course, be satisfactory merely to say that in the latter case we have a mapping from thoughts to thoughts (and hence from truth-values to truth-values), whereas in the former we have a mapping from coins to bets, for this just raises the problem

all over again. Why do we in the one case wind up with the same kind of thing we started out with, while in the other we wind up with something different (a bet rather than a coin)?[7]

Miss Anscombe claims that Geach's critique of the Lewis-Langford view helps us to see why we use the same signs in all the columns of the truth-table. I find this claim rather puzzling, for it seems to me that Geach's objection--though certainly well-taken--makes the situation even more troublesome than it was to begin with. And this fact will, I think, have some interesting consequences concerning negation.

We saw that the mistake Lewis and Langford made was one of tacitly assigning different properties to the column of the matrix. Now in the biological example Geach mentions, there was a consistent assignment to the 1's and 0's in all

---

[7] We should keep in mind that 'bet' can mean: first, the possible distinct bets (or in another usage, kinds of bets) that can be made using any coin; second, the possible, distinct bets (or kinds of bets) that can be made with a particular coin; and third, the bets I actually make at various times using a particular coin. Bets in the third category are tokens of the types in the second, and these types are what would be signified in the above table if we inserted a name of a particular coin where 'coin' appears. Bets in the first category are what would be signified in the table if we inserted a "coin-variable" in that place (in effect, 'coin' as it occurs in the table, actually plays this role). Corresponding distinctions can be made in the truth-table case. The letter 'p' can be construed as a sentential variable or as the name in the symbolism of a particular English sentence. In the former case, we would have the analogue of the first category, and in the latter, of the second one. The analogue of the second category, then, would be sentences or thoughts, and the analogue of the third category would thus be actual assertions of these sentences or thoughts made by individuals on various occasions. In this context, however, we are interested mainly in the first two categories.

the columns, but the statement derived from the schema (S_1) by replacing 'A' with an expression signifying a certain hereditary property was not analytic, and the 'I' in 'pIq' carried no logical significance. Now Anscombe gives us another interpretation of that table which, she maintains, is such that the 1's and 0's in <u>all</u> the columns indicate that moves in a game are respectively allowed and forbidden, and she then claims that 'I' on this interpretation will have logical significance and that the following statement will be analytic:

S_2: If p is an allowable move, then q is an allowable move if and only if pIq also is an allowable move.

She then goes on to say:

> Now [S_2] at least sounds like a substantial piece of information about a game; at any rate if one can specify p, q, and pIq independently as moves: p might be a diagonal move of a certain piece on a square board, q a move of another piece parallel with an edge of the board, and 'pIq' the name given to a move by yet another piece along one of the edges of the board. Then the information 'If you can move this piece diagonally, you can move this piece along the edge if and only if you can move this piece parallel with one of the edges' might either be a rule of the game, or inferable from the rules of the game.
> If, however, we now ask why moving this piece along the edge of the board should be symbolized by a sign mentioning those two other moves, then we can answer by saying that
>
> | p | q | |
> |---|---|---|
> | 1 | 1 | 1 |
> | 1 | 0 | 0 |
>
> defines a possible allowability in terms of given allowabilities, and 'pIq' is a notation in which this allowability is set forth: That, and that alone, is the meaning of the notation.

> Then the substantial information as far as this game is concerned is that there do exist some moves whose allowability is so conditioned and that the move along the edge is in fact one of them; but that move describable as pIq is not a "substantial" piece of information [i.e. is analytic].[8]

Miss Anscombe is clearly imagining a case where the pIq move is such that it will be allowable only if certain conditions are satisfied, and these conditions are laid down in the table. However, let us assume (as Miss Anscombe apparently does) that the moves p and q are not themselves such that their allowability is conditional upon anything else, i.e. let us assume that they are "basic" or unconditional moves. Then what sense is there in saying that pIq in an allowable move in some cases and not allowable in others? If p and q are basic moves, then it is not as though the two rows of the above table determine different possible situations within the same game, rather, they determine different games altogether. For insofar as a move is basic, it is ipso facto allowable, and a basic move such as q cannot be both allowable and not allowable in the same game. Thus, the allowability of the pIq move is not conditional upon the allowability of the p move and the q move, rather if the allowable moves p and q are actually taken, then one can if he wishes make the pIq move; and if only p is taken and not q, one cannot make the pIq move (in that case, it is not allowable). So Miss Anscombe is guilty

---

[8] Introduction, pages 55-56.

of the same mistake she accused Lewis and Langford of making--on her interpretation of the table, the 1's and 0's in the p and q columns indicate whether an allowable move has in fact been taken or not, but the 1's and 0's in the pIq column indicate, not that pIq has or has not in fact been performed, but that it can in the first case be taken (i.e. is allowable) and cannot be taken in the second. It is only via this inconsistent assignment that one can say that pIq is a move whose allowability is (logically) conditioned; on the consistent assignment, each distinct row of the table determines a distinct game, and we will have an "S-statement" for each one. Moreover, these statements will not be analytic. For example, the statement for the first row will read:

S_3: If p and q are allowable moves, then pIq is an allowable move.

(S_3) simply states how, as a matter of fact, we have decided to set up a certain game--we just aren't going to let pIq be allowable unless we decide to let p and q be allowable. That is, for whatever reason, we have decided that if we should set up a game in which p and q are allowable, we will also make pIq allowable. Obviously, if our whims are otherwise, we can do just the opposite. Further, it should also be clear that in this case 'I' has no logical significance whatsoever; we could just as well have referred to the move as 'r' instead of 'pIq'--logically, there is no point of using the complex symbol.

On the other hand, the S-statement for the inconsistent assignment is analytic, and 'I' does indeed carry logical significance; the S-statement will be:

S_4: If the allowable move p is taken, then pIq is allowable if and only if the allowable move q is also taken.

Here, of course, there is a point in symbolizing the conditional move as 'pIq'--the nature of its conditionality is explicitly given since the definition of 'I' is simply the table itself. Thus, it would not have been satisfactory merely to have labelled it 'r'.[9]

In summary, the allowability of the pIq move is not conditional upon the allowability of the p and q moves, but upon those (ipso facto) allowable moves actually having been made. And we have seen that if we let the 1's and 0's indicate that moves are allowed or forbidden in all the columns, then we do not have a game at all, but rather four distinct ones, and pIq's "conditional" character is due solely to our whims--it bears no logical relation to p or q--the only "relation" involved is a psychological one.

Let us now return to the betting example; consider the following table ('α' and 'β' are variables ranging over coins, and '$' is a logical symbol defined by the table):

---

[9] It might be objected that since (S_2) is analytic, and since it logically implies (S_3), that (S_3) must be analytic as well. However, the above discussion shows that (S_2) must be replaced by (S_4). It is (S_4), then, that is analytic, and it clearly does not logically imply (S_3).

α	β	α$β
H	H	W
H	T	L

Here, we have a case where it is explicitly recognized that the property indicated in the first two columns is distinct from the one indicated in the bet column. Further, it should be clear that '$' is a sign having logical significance, and it certainly appears to perform the same sort of job the logical connectives have in truth-functional logic. Moreover, the following statement is analytic:

S_5: If α comes up H, then α$β is W if β comes up H.

Thus, the coin example is exactly like Anscombe's example concerning games (where the latter is regarded as involving an inconsistent assignment). And this fact, together with the fact that the property in the biological example remained the same in all the columns but that the connective did not have logical significance, seems to indicate that in order for the connectives of an interpreted matrix system to have logical significance, the property designated in the "component" columns of the matrices must be distinct from the property designated in the "compound" column. Now it would surely seem that this general thesis is false; but since the connectives of the truth-functional calculus clearly do have logical significance, it might well be in order to ask whether there are different properties in the "component" and "compound" columns of an ordinary truth-table. That is, do the

T's and F's mean the same in all the columns, and more particularly, does 'p' in the table:

p	p	~p
T	T	F
F	F	T

stand for the same "kind of thing" in both its occurrences? In discussing this question, we will, I think, be able to resolve the problem surrounding the "general thesis" stated above.

Pictures as Projections of States of Affairs

Oddly enough, it is Anscombe who indicates an answer to this question. She begins chapter four of her book by saying:

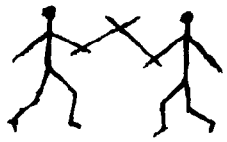
> We have been troubled..by placing the same signs, T and F, under the signs of the elementary proposition in the truth-table and in the final column, and by the (usual) justifications of this procedure... Let us now consider remedying this procedure. I adopt two new signs, P and N, which mean "Positive" and "Negative", which I put under the signs of the elementary propositions:

p	p	~p
P	T	F
N	F	T

> The proposition which is an element in a truth-function is thus introduced as having two senses, the positive and the negative, rather than two truth-values, T and F.

Now what does it mean to say that a proposition or sentence has both a positive and a negative sense? I think we can see what it means by regarding a simple singular sentence as

functioning in part like a picture--not that it is a picture, but only that it has certain logical features in common with pictures. Let us consider the following drawing taken from Wittgenstein's Notebooks.



If the figures in the drawing are correlated with persons, then I can assert that these persons are fencing by holding up the picture and saying: "This is how things are." But of course I can also hold it up and say: "This is how things are not." It is important to note there that the "this" is the same in both cases, viz. a certain, possible situation or state of affairs. If the figures represent Mr. A and Mr. B, then the state of affairs would be "A's fencing with B" or perhaps "that A and B are fencing". Thus, the (labelled) picture presents us with, or specifies, a possible state of affairs; however, it does not just do this. The state of affairs in question will either exist or not exist, and as we have just seen, I can use the picture either to assert its existence or to assert its non-existence--to say that A is fencing with B or that he is not.

Now this means that the picture, independently of whether or not it is actually used to make an assertion, can be "regarded" or, in Wittgenstein's phrase, "thought out" in two opposed, though complementary, ways--as picturing something existent or something non-esistent, the "something" being the

same in both cases. At 3.11 in the Tractatus, he says: "The method of projection is to think out the sense of the picture." And at 4.464, he makes it clear that the sense (the possible state of affairs) can be "thought out" in two different ways:

> ...a proposition, a picture, or a model is, in the negative sense, like a solid body that restricts the freedom of movement of others, and, in the positive sense, like a space bounded by a solid substance in which there is room for a body.

And in a corresponding passage in the Notebooks,[10] he adds the following diagram:



Thus, the analogy, just as the rectangular figure marks our a certain definite area on an indefinite plane, so the picture, by expressing a sense, picks out (or specifies) a particular state of affairs from an indefinite range of others. And just as the same figure can be "thought out" or "projected" either as a solid body surrounded by an open area or an open area within a solid body, so likewise the same picture can be "thought out" or "projected" in one of two ways--either as representing an existent state of affairs or a non-existent one, but the same state of affairs in the two cases. Now though the notion of"thinking out" or "projecting" the sense

---

[10] Notebooks, page 30.

of a picture--either as existing or as not existing--may seem rather obscure, it will, I hope, become clearer upon the following considerations.

First, a picture does not stand to the state of affairs it represents as a definite description stands to its reference. The pictured state of affairs cannot be the reference of a picture because since every picture, qua picture, is a picture of something (the state of affairs), it would follow that if this state of affairs were construed as its reference, then no picture could ever fail to refer. And from this it would follow that no picture could ever be a picture of a state of affairs that in fact does not exist--which is clearly false. Rather, it is one thing for a picture to represent something, and it is then a separate and further issue as to whether this "something" exists. Thus, the represented state of affairs is not the reference of a picture, and indeed it would in virtue of the foregoing seem to be a mistake to suppose that it could properly be said to have a reference at all. But since knowing what a picture represents consists in knowing which state of affairs is represented (and which may then either exist or not), it would follow that the state of affairs ought more accurately be construed as its sense. A picture, then, has a sense independently of what the facts are--we can know which state of affairs a picture represents without in any way having to know whether that state of affairs does or does not obtain. For this reason, it would be wrong

to suppose that there are two "kinds" of pictures qua pictures--those representing existent states of affairs and those representing non-existent ones--just as it would be wrong to suppose that there are two "kinds" of rectangular figures--those which are open spaces in otherwise solid bodies and those which are solids in open spaces. We can of course classify pictures and rectangles in this way (e.g. "true" pictures and "false" ones), but such a classification would be of a purely ad hoc nature and not pertain to the pictures or rectangles as such (qua representations or geometric figures). The state of affairs represented by a picture will either exist or not, but whether it does or not has no bearing on the picture's expressing a particular sense (representing a particular state of affairs).

Thus, the picture drawn above represents or specifies the state of affairs: A's fencing with B, and it does this independently of whether in fact A is fencing with B. But it surely would be wrong to suppose that there is another picture which specifies the "negative" state of affairs: A's not fencing with B. For one cannot make a picture of A's not fencing with B, but only of their doing something other than fencing--again, we must not confuse what is not the case with what is the case instead of it. As Wittgenstein says in the Notebooks: "If a picture presents what-is-not-the-case... this only happens through its presenting that which is not the

case."[11] I can make a picture of A and B shaking hands or playing tennis, etc., but I cannot make a picture of their not fencing. Now this should not be construed so as to deny the obviously true claim that one can make a picture of a state of affairs that in fact does not exist. Indeed, this is the very point; a picture has a sense which is independent of the facts and thus it can represent a non-existent state of affairs. Nor should these remarks be construed as denying the equally true claim that a picture can often inform us that what is pictured does not have a certain characteristic. To use an example of Anscombe's,[12] a map of Birmingham tells us that Birmingham does not have a large river passing through it by not showing a large river passing through it.

Thus, if in fact A and B are not fencing, it would be absurd to suppose that I could express this by holding up a "negative" picture and saying: "This is how it is", for there are no such pictures. Rather, I must hold up the single picture of their fencing and say: "This is not how it is." And herein lies the ideal behind the notion of logically conceiving or "thinking out" (or "projecting") the same thing in two different, mutually opposed ways. Just as the rectangular figure:

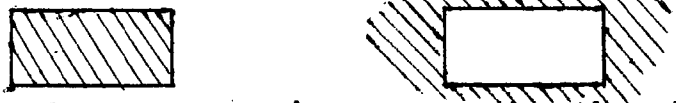
---

[11] Notebooks, page 25. The italics are his.

[12] Introduction, page 70. Many of these same points are developed by Anscombe in chapter four, though, as will become apparent, there are several areas of disagreement between us.



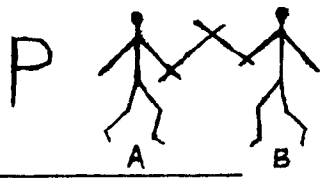
can be thought out or projected either as a solid in an open space or an open space in a solid, and the result of such projections being expressed in the shaded illustrations:



so the sense of the same picture can be thought out or projected in two different, mutually opposed ways--either as being the case or not being the case. It is on this kind of analysis that the relation between the picture on the one hand and the existence (or non-existence) of what it represents on the other becomes intelligible. When I hold up the picture and say: "This is how it is", the sense I thereby express is mutually opposed to the sense that is expressed when I hold up that same picture and say: "This is not how it is". These senses, as opposed to the sense of the picture per se, are clearly negatives of one another, and stand to each other analogously to the way in which the above pair of shaded figures stand. Indeed, if we now employ the shading technique to signify the notion of "excluding" as it was developed in the last chapter, the above pair of figures would be a picture in "logical space" of the two thoughts--one positive and one negative--which mutually specify the same "alternative". Thus, in holding up the picture and saying: "This is how it is", I express a positive sense. Hence the notion of logically conceiving or projecting the sense of the same

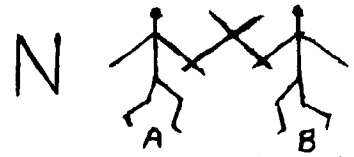
picture as either being the case or not being the case can be explicated in terms of using the same picture to express either a positive sense or a negative sense respectively. For this reason, we may speak of projecting the sense of a picture "positively" as well as projecting it as "being the case" or "existing", and similarly for the negative case.[13] We can, then, distinguish two aspects or dimensions of a picture: first, that it represents and thereby singles out a particular state of affairs (just as the rectangle marks out a definite area on a plane), and second, that with respect to this state of affairs, it may be projected in two different, complementary ways: positively (being the case) or negatively (not being the case), where this means that its representation (the picture) can be used to express both a positive and a negative sense as these notions have been explicated according to the independent criterion.

Now, instead of uttering the words, "This is how it is" or "This is not how it is", we can introduce two signs, 'P' and 'N' respectively, into the picture to do the same job. Thus, instead of holding up the picture and saying, "This is how it is", I can hold up the drawing:



[13] It should be clear that if we had two pictures, we could then project or think out their senses in 14 different ways, corresponding to the 14 significant truth-functions of two or more arguments, i.e., all truth-functions of two or more arguments except tautology and contradiction.

where the 'P' to the left of the picture carries the same force as saying "This is how it is."[14] That is, this drawing both specifies the state of affairs and projects it positively. Likewise, in the drawing:



the represented state of affairs is projected negatively. Now this does not mean that there are "positive" and "negative" pictures in the sense discussed earlier; rather the same picture is an element of the two drawings. Instead, these drawings function in a much different way; for the picture (i.e. without either the 'P' or the 'N') merely represents or specifies a state of affairs, but the above drawings express a positive and a negative sense respectively, and thus may be characterized as true or false. The state of affairs will, after all, either exist or not exist, and the first drawing in effect says that it does exist, while the second says that it does not. However, the fact that they can be

---

[14] It may be objected that we cannot replace the utterances, "this is how it is" and "this is not how it is" with purely linguistic devices like 'P' and 'N', on the ground that the former, in addition to indicating positive and negative projections, also carries assertoric force. That is, I assert the state of affairs to obtain when I say "This is how it is." However, this is not correct; merely saying the words "this is (not) how it is" does not suffice to automatically produce an assertion. For I can hold up a picture of A and B fencing, and then one of A and B shaking hands, and say: "If this is how it is, then (second picture) this is not how it is." And it is clear that I am here neither asserting that the first state of affairs obtains nor that the latter one does not.

described as "saying" the one or the other does not mean that they are, in and of their own accord, "assertions". For inasmuch as they can occur in truth-functional and intensional contexts--as well as fictional ones--it follows that 'P' and 'N' bear no logical resemblence to Frege's assertion sign. Nevertheless, these drawings do more than merely project the sense of the picture in one way or the other, though they do not go so far as to carry assertoric force. For when I insert 'P' into the picture or say the words, "This is how it is", I produce a sense which is the bearer of truth-value; and this means that the drawing containing a 'P' or an 'N' is like a sentence in the indicative mood rather than in the imperative or[15]

---

[15] I could, thus, hold up the picture and say, for example: "Is this how it is?", and I would then produce a sense which would be a question, not a thought. This makes it clear that some sort of "neustic-phrastic" distinction could be worked in here; or that we could further distinguish between the "semantic mood" of an utterance wherein the picture was employed and the way the picture was projected in that utterance. Thus we could perhaps contrast:

This is how it is: P ( ). with
Is this how it is: P ( )?

where the '( )' is the picture, and 'P( )' functions analogously to what is nowadays termed the "phrastic". Thus, 'P' here serves only to project the sense of the picture. The "newstic" or "model" element is then supplied by phrases like "this is how it is" (but not "this is not how it is" for the positive-negative dichotomy is logically antecedent to the neustic). However, the points I wish eventually to make concerning assertions and indicative sentences do not, as far as I can see, turn on any particular version of the neustic-phrastic distinction. And since that distinction is a rather controversial and "theory-laden" one, and since my interest in this essay lies entirely within assertoric logic, to introduce the distinction here would only complicate matters unnecessarily.

interrogative ones. But inasmuch as the drawings are bearers of truth-value, there is, as we saw back in chapter one, a sense in which they can be described as "saying" something--they depict or show how things do (or do not) stand *if* they are true (inasmuch they contain the picture as an element), and further, because of their "projective" dimension, they *say* that things do (or do not) stand in the depicted way. And this parallels Wittgenstein's comments about propositions or sentences at 4.022. From here on, the drawings containing 'P' or 'N' will be termed *illustrations*, and the difference between an illustration and the picture it contains should now be clear.[16]

Singular Sentences as "Projections"

Now I think that the two "dimensions" as they were called--the "descriptive" and "projective" ones--can be found in the singular sentences of ordinary discourse. We already have ample reason for saying that a sentence has no reference--and that in any case the specified state of affairs cannot be that reference (the "facts" theory of chapter one). Now we have seen that it would be a mistake to distinguish two "kinds" of pictures *per se*--those representing existent states of affairs and those representing non-existent ones--for such a distinction

---

[16] However, though they are distinct notions, there still remains a sense in which every illustration is *also* a picture, just as every rectangle which is a solid in an open place *is* a rectangular figure. Likewise, every picture which is projected in one way or the other *is* a picture.

would be of a purely ad hoc nature and not pertain to the pictures qua pictures--existence, that is, is not an "intrinsic" property of a picture. And even though it is the illustrations that are properly the bearers of truth-value, it would be wrong for the same reason to suppose that "true" and "false" classify illustrations as such--though of course "positive" and "negative" would do so. Now to say that an indicative sentence has a sense is at least to say that it specifies or describes a possible state of affairs[17]--one which may or may not be the case. And just as in the case of pictures or illustrations, it would be wrong to say that there are two kinds of sentences as such--those where the specified state of affairs exists and those where it does not. Such a distinction would not, logically speaking, be on a par with, say, distinguishing between singular and general sentences. Nor would a classification of indicative sentences into those expressing true thoughts and those expressing false ones be a genuine classification of intrinsic "kinds" of sentences. "True" and "false" do not distinguish among sentences qua sentences (or thoughts qua thoughts) any more than they do in the case of illustrations. A sentence has a sense indepently of how the facts are; the

---

[17] It must of course be kept in mind that a (non-eternal) sentence can be used to express many different senses, just as the labelled picture discussed above, or better, the illustration, can be used to express different senses. Again, 'sense' or 'thought' here mean roughly what is nowadays called a "statement", i.e. the particular claim that is made on an occasion of the sentence's (or illustration's) use.

facts can have no bearing on the logical characteristics of sentences as such (or their senses).[18] Thus on page 23 of the Notebooks, we find Wittgenstein saying: "For a proposition to be true does not consist in its having a particular relation to reality, but in its really having a particular relation." However, a more accurate statement would read" "For a proposition to be true-or-false consists in having a particular relation." The relation in question would be the analogue of how apicture represents something else. And at 4.061 in the Tractatus, we read:

---

[18] This raises two very difficult and important questions which can only be mentioned here because they are not strictly germane to the main subject of this essay. First, what is it that makes a set of marks into a picture of a possible situation (or variety of similar situations)? That is, what is it that enables us to use a set of marks to represent something else? This question is of course discussed at great length in both the Notebooks and the Tractatus. Secondly, what is it that makes a string of words into a certain kind of complex expression, a sentence? This question is of interest both in the philosophy of logic (since it can be construed as asking for a logical analysis of the concept of predication) and more generally in "philosophical linguistics". However, my concern here is not with predication nor with the principles underlying sentence formation in natural languages, but rather with exploring the various logical concepts pertaining to sentences as a whole. As is well-known, Wittgenstein gave essentially the same answer to both of the above questions, for he held that the most fundamental sentences ("elementary propositions") really were pictures. I do not believe there are such sentences, and it should be obvious that the singular sentences of natural languages, tho they may resemble pictures in various logical ways, are certainly not themselves pictures.

> It must not be overlooked that a proposition has a sense that is independent of the facts; otherwise one can easily suppose that true or false are relations of equal status between signs and what they signify. In that case, one could say, for example, that 'p' signifies in the true way what '~p' signifies in the false way, etc.

That is, in such a case, "true" and "false" would classify "kinds" of sentences or propositions, the state of affairs would be specified or signified either "truly" or "falsely" (not, positively or negatively). And if that were the case, we could gain knowledge merely by examining "how" a sentence signifies, instead of undertaking an actual investigation to see whether the state of affairs obtains.[19]

We saw in the last chapter that Frege's truth-functional analysis of negation requires that 'not' attach to the sentence as a whole rather than the predicate. Thus, for logical purposes, the sentence 'a is not F' might be more perspicuously

---

[19]The remarks at 4.061 should not of course be taken as a direct criticism of Frege, for he obviously did not hold this simple-minded view. Though Frege distinguished between sentences standing for the True and those standing for the False, he clearly did not intend this distinction as a means of classifying kinds of sentences or propositions. Such a classification would be made according to the logico-grammatical characteristics they have or to the senses they commonly express, but certainly not to their references. However, though Frege did not intend his distinction between sentences standing for the True and for the False in this way, he nevertheless falls into the trap. For as we saw in chapter two, he attempts to specify the senses of various sentences by specifying what references they will have under certain conditions.

rendered 'Not: a is F'. It was also argued in that chapter that the senses or thoughts expressed by 'a is F' and 'Not: a is F' can be termed "positive" and "negative" respectively on purely logical grounds--in other words, independent of the fact that they are expressed by positive and negative sentences. This was the "independent criterion", and part of this criterion went to the effect that the two sentences must specify or describe the very same alternative, or as we are now calling it, the same "state of affairs".[20] Now, that the indicative sentences, 'a is F' and 'Not: a is F' specify the same state of affairs and then, in virtue of the former expressing a positive sense and the latter a negative one, go on to treat of it in the complementary ways outlined in the last chapter, at least suggests that there is a common abstrctable element in the two sentences whose sense can then be logically "thought out" in the same two opposed ways as a picture specifies (represents) a possible state of affairs, so the abstracted element--which for the present we might (somewhat misleadingly) characterize as "a's being F"--specifies (describes) a possible state of affairs. And this element would

---

[20] Actually, it was said that the thoughts specified the same alternative, not the sentences themselves. However, it will be more convenient in terms of the points developed in the last few pages to reexpress this in terms of the sentences, by virtue of their expressing the particular senses or thoughts they do on some occasion of utterance, that specify the alternative. Nothing is really altered by this shift in terminology.

therefore stand to the two sentences as the picture stands to the two "illustrations"; that is, just as an illustration projects the state of affairs represented by the picture in a certain way, so likewise a sentence projects the state of affairs described by the abstractable element in one of the two possible ways. We might, then, imagine the indicative sentences, '_a_ is F' and 'Not: _a_ is F' as being logically "decomposed" as follows:

It is the case: _a_'s being F.
It is not the case: _a_'s being F.

Thus the expressions, 'it is the case' and 'it is not the case', function like the 'P' and 'N' introduced earlier in connection with pictures--they "project" the sense of '_a_'s being F' in the two possible ways and also supply the indicative mood. But it would be wrong to analyze 'Not: _a_ is F' as:

It is the case: _a_'s not being F.

For just as there are no "negative pictures", so there are no "negative" abstractable elements; rather, the positive-negative dichotomy is logically subsequent to the abstractable element. In order to say that something is not the case, I must say _what_ it is that is not the case, and the "what" here is the state of affairs of _a_'s being F. Hence, on the picture analogy, we do not have the state of affairs: _a_'s not being F, we have only the state of affairs, _a_'s being F, which may then be projected in the negative way. However, this does

point up the fact that the verbal noun, '_a_'s being F', since it admits of a negative, is not an entirely accurate linguistic rendering of the descriptive dimension of a sentence. And indeed, it does not seem as though there is any existing expression in natural languages that is completely adequate. But I do not see any reason for supposing that this fact undermines the line of analysis suggested here. For the picture analogy only requires us to acknowledge two conceptually distinct aspects or dimensions of a sentence--not that there must exist in natural languages two classes of well-formed expressions corresponding to each, and which, when combined, produce a sentence in the manner of building blocks being put together to make a building. To suppose this would be like criticizing Frege's theory of functions on the ground that '$3\underline{x}^2+\underline{x}$' cannot be broken down into '$\underline{x}$' and '$3()^2+()$' because the latter expression is not well-formed.

Now I wish to maintain that the 'N' as it occurs in an illustration operates much like 'not' in English (though this claim will have to be qualified eventually). The common abstractable element in '_a_ is F' and '_a_ is not F' is like the picture considered independently of the ways in which it can be thought out. To return to the plane analogy, this element inscribes the figure on the plane, and it will then be a _further_ question as to how the interior or exterior areas formed as a result are to be shaded. And so introducing 'not'

into a sentence is like telling us that we should shade the interior area (again, the shading to signify "excluding") and leave the exterior blank, i.e. the insertion of 'not' indicates that the sense of the abstracted element is to be projected "negatively" as this term was introduced earlier. We shall return to this topic presently.

For the moment, however, we must note that there is a difference between 'it is the case' and 'P' as it occurs in the illustration, for in the former expression there is no explicit indicator of a positive sense--nothing, that is, which does the same kind of job 'not' performs in 'it is not the case'. We shall discuss this point in greater detail later on, but for now it can be said that in lieu of the occurrence of 'not', it is presupposed that the sense of the abstracted element is projected positively. We do not need to (though of course we could) introduce a sign for this purpose. If the sign were 'pos', we would then have:

It is pos the case: a's being F,

and the sentence would read 'a is pos F' instead of merely 'a is F'. It should not be surprising that there is no positive indicator; both the picture and abstracted element specify a possible state of affairs and pick it out from an indefinite range of others. To put it crudely, they pick out a definite element from an indefinite range of other such elements. Hence it is only natural that, barring an indication to the contrary, our attention be directed to what is picked out rather than

that from which it was picked, and it is the positive sense which projects the "definite" element as <u>being</u> the case. As we saw in chapter two, it is the positive thought which, if true, gives us definite knowledge of what is the case. Thus, the positive projection of the sense of a picture or abstracted element acquires a certain fundamental primacy over the negative, making unnecessary an explicit indicator. If I know that A and B are not fencing, then upon being confronted with the labeled picture, I would likely say, "That's <u>wrong</u>, they're not doing that." Our natural tendency, that is, is to understand the picture as the positive illustration.

We have said that just as the illustration can be distinguished from the picture as its abstractable element, so we may distinguish the sentences: '<u>a</u> is F' and 'Not: <u>a</u> is F' from their common abstractable element. The latter, inasmuch as it has a sense, specifies a state of affairs which may then be projected as either being the case or not being the case. Now this fact, together with the fact that each of the above sentences expresses a sense which <u>is</u> a projection (in the "product" sense) of the state of affairs (analogous to the way in which an illustration "is a projection of" the state of affairs represented by the component picture) <u>and</u> the claim, argued in chapter two, that the senses of these sentences may, according to the independent criterion, be legitimately described as "positive" and "negative" on purely logical grounds, means that we may justifiably describe the

state of affairs as being projected "positively" (in place of "being the case" or "existing") and "negatively" (in place of "not being the case" or "not existing"). It is, then, because of their common abstractable element (hereafter, for short, the "CAE"), that '_a_ is F' and 'Not: _a_ is F' specify or describe the same "alternative" as it was called in chapter two, and then go on to treat of it in the two different, complementary ways outlined in connection with the independent criterion. In that criterion, the notions of "excluding" and "leaving open" were introduced on the basis of what would or would not be the case if the thought or sense in question were true. Thus, the sense of '_a_ is F' excludes all alternatives except the specified one and leaves none open, whereas the sense of 'Not: _a_ is F' excludes the specified alternative and leaves open all the rest. And it should by now be clear that we can introduce a third notion coordinate with these which we shall call "affirming". We shall say that a thought or sense of an indicative sentence _affirms_ an alternative just in case one can infer from the truth of that thought or sense that the alternative in question obtains or is the case. And we can now say that the sense of '_a_ is F' affirms the specified alternative and excludes all others of the same range (hence leaving none open), whereas the sense of 'Not: _a_ is F' excludes the specified alternative and leaves open all of the others. Therefore, the negative sense affirms no _particular_ alternative, and herein lies the "indefiniteness" of

the negative sense mentioned back in chapter two. The negative sense, if true, tells us that one of the remaining alternatives may obtain, but, qua negative, it cannot tell us which one (if any) it is. Knowing what would obtain if the negative sense were true is then a far different and more "indefinite" matter from knowing what would--and would not--obtain if the positive sense were true, and this observation may perhaps serve to reinforce the reasons just given as to why there is no explicit indicator of a positive sense; and why, as a result, it is from a "psychological" point of view "natural" to assimilate the CAE with its positive projection.

In any case, it should now be apparent that the job of specifying or describing the alternative or state of affairs is performed primarily by the CAE, and that the logically subsequent features of affirming, excluding, and leaving open arise only in connection with projecting that state of affairs. Thus, as we have seen, the independent criterion can be used to explicate the notion of "projecting" the sense of the CAE--either as being the case or not being the case. For to say that 'a is F' is the positive projection of the state of affairs described by the CAE (i.e. that it projects the state of affairs as being the case) is to say that its sense or thought affirms the described state of affairs, excludes all others of the same range, and hence leaves none of them open. Correspondingly, 'Not: a is F' is the negative projection of

that same state of affairs inasmuch as its sense or thought excludes that state of affairs, leaves open all other of the same range, and hence affirms none of them.

It will perhaps be instructive to consider some of Wittgenstein's remarks on negation and singular propositions. At 3.144 of the Tractatus, he says: "... propositions are like arrows, they have sense." The arrow analogy suggests the idea of propositions as having a "direction"--an idea which in fact crops up several times in the Notebooks as well as the Tractatus. As I read this metaphor, we are being told that a proposition not only specifies a state of affairs, it also projects it either positively or negatively. And this reading is born out by some further passages. For example, on page 26 of the Notebooks we find:

> It might be said: the negation refers to the very logical place which is determined by the negated proposition.
> Only don't lose the solid ground on which you have just been standing!
> The negative proposition determines a different logical place from the negated proposition.
> The negated proposition not only draws the boundary between the negated domain and the rest, it actually points to the negated domain.
> The negating proposition uses the logical place of the negated proposition to determine its own logical place. By describing it as the place that is outside the former.

The passage is intriguing because it can readily be seen why Wittgenstein corrected himself at the beginning. The proposition p and ⌜not-p⌝ determine the same boundary, but the

logical place of ⌜not-p⌝ is outside it, and that of p inside it. Wittgenstein was confusing the CAE shared by p and ⌜not-p⌝ with p and ⌜not-p⌝ themselves. In terms of the plane analogy, upon which the notion of "logical space" may be constructed, ⌜not-p⌝ excludes the interior, and p the exterior. Notice also that the negated proposition "actually points to" the interior area; that is, the proposition p expresses a positive sense, and this is like shading the exterior area so that the interior stands out by way of contrast. It is in this way that propositions resemble arrows; they have "direction" inasmuch as they are projections of states of affairs, and this is born out by 5.2341 where it is said that "negation reverses the sense of the proposition." The sentence-CAE distinction is not explicitly drawn by Wittgenstein, but it is surely implicit in these passages as well as certain others to be cited later on. And though the picture-ilustration distinction is not explicitly recognized, it is clearly evident in a passage on page 27 of the Notebooks: "In order to designate a logical place with the picture, we must attach a way of symbolizing to it (the positive, the negative, etc.)."

So the two dimensions distinguished earlier in connection with pictures are to be found in singular sentences as well, with the result that a good deal of light is shed on various logical features of such sentences, e.g. that they have a sense independently of the facts, that their senses are bearers of truth-value, that the negative of p must logically be

expressed in terms of p, that in ⌜not-p⌝ there is no element in the specified state of affairs corresponding to 'not', and so forth. More needs to be said on these matters of course, but for the present it will suffice to note the analogies holding between the picture-illustration and CAE-sentance distinctions, and the explication of the "existent" and "non-existent" methods of projection in terms of the independent criterion.

It will be worthwhile if we pause at this point to consider the notion of sense more carefully. We have spoken of the sense of the CAE and the sense of (thought expressed by) an indicative sentence. So far, we have very roughly characterized the sense of the CAE as the possible state of affairs the CAE "describes"--this term being analogous to "represent" with respect to a picture. However, this characterization, taken as it stands, requires modification; for to say that a CAE has a sense is not simply to say that it specifies a state of affairs. Rather, to ask which sense it has would only be in part to ask which state of affairs it describes, for the distinct, non-synonymous CAE's of the sentences, 'the Morning Star is a planet' and 'the Evening Star is a planet' nevertheless describe the very same state of affairs. Thus, it must not be supposed that any two CAE's which describe the same state of affairs will have the same sense--though of course the converse of this will hold. Rather, the sense of a CAE is determined in part by the senses of the component

expressions, and therefore it is not really accurate to speak of the state of affairs specified by a picture or CAE as their "sense". However, none of the foregoing arguments are affected by this discrepancy. We could have made this distinction at the outset and then gone on to speak only of states of affairs as being projected; yet our conclusion that a sentence is a projection of a state of affairs and that its sense may accordingly be described as positive and negative (as these notions were introduced via the independent criterion) would still hold. However, to do so would have unduly complicated the exposition, and thus the sense vs. state of affairs dichotomy was not introduced at that time. Not even our initial claim--that since a state of affairs is not the reference of a picture or CAE, it "ought more accurately be construed as its sense"--is really damaged. For though there is no identity of sense and state of affairs, the similarity is so close as to warrant treating them as one for purely expository purposes. For example, because a picture or CAE has a sense independently of the facts, so likewise they must specify a state of affairs independently of the facts.[21]

---

[21] However, we have come across a striking difference between the singular sentences of ordinary discourse and Wittgenstein's elementary propositions. For the latter, since they contain terms which have only reference but no sense (or alternatively, their sense is their reference), would not admit of a distinction between sense and state of affairs. The possible state of affairs quite literally is the sense of the proposition--odd as it is to speak of a "sense" as "existing" or "not existing".

However, since we have now arrived at the trichotomy of CAE, sense, and state of affairs, we shall hereafter speak of the CAE as expressing a sense, and inasmuch as specifying a sense is sufficient for specifying the corresponding state of affairs, we shall say that the CAE, by virtue of its expressing a sense, specifies or describes the state of affairs. Now we have seen that the state of affairs may be projected in two opposed, complementary ways which have now been termed "positive" and "negative". A sentence, then, will be such a projection, and its sense described as positive or negative as the case may be. However, in one manner of speaking, the sentences, p and ⌜not-p⌝, may be said to express the same sense, viz. the sense of the CAE they share. For it is in virtue of their expressing this sense that they are projections of the same state of affairs. However, qua projections, they express different and opposed senses, and these senses are bearers of truth-value and hence are what we express in the course of making assertions. Thus, it should not be supposed that because we can distinguish the CAE from the sentences, p and ⌜not-p⌝, we then have three distinct, coordinate senses: the senses of the CAE, of p, and of ⌜not-p⌝. That is, we do not have "three senses" in the way in which Aristotle, for example, distinguished "three senses" of the word 'healthy'. To suppose this would be rather like supposing that when I hold a coin in my hand, I am holding "three things": the coin, the heads side, and the tails side. Rather, the relation between the positive and negative senses to the sense of the CAE

should be properly analogized to the two shaded drawings of a rectangular figure and the figure left unshaded. It would be absurd to suppose we then have three possibilities: the figure, the figure qua solid-in-an-open-space, and the figure qua open-space-in-a-solid. Likewise, we have one state of affairs and two possibilities: being-the-case and not-being-the-case. We do not have three possibilities, nor would it make any sense to go in the other direction and say we have three "states of affairs": the state of affairs, the state of affairs' being the case, and the state of affairs' not being the case. To refer to these latter two as being themselves "states of affairs" would be to introduce a radically new and different sense to the phrase 'state of affairs'--and one which would have no point. For this sense of 'state of affairs' can be easily explicated in terms of the original sense of that phrase, together with the notion of projecting a state of affairs (in the original sense). A "state of affairs" in this new sense is simply a projection (positive or negative) of a state of affairs in the original sense. But a projection of a state of affairs in the original sense simply is a sentence, and so it is true in a quite trivial way that there are "positive" and "negative" states of affairs in the new sense--true inasmuch as there are, obviously, sentences expressing positive and negative senses. On the other hand, it is quite false to suppose that a "state of affairs" (or "fact") in this new sense can be a non-linguistic entity

which can itself be classified as "positive" or "negative", and which, if it exists, renders true the sense expressed by the "corresponding" positive or negative sentence.[22] What renders the thought or sense true is simply the existence or non-existence of the state of affairs in the original sense of that phrase.

Moreover, it is equally false to suppose that there are "positive" and "negative" states of affairs in the original and bona fide sense. As we have already seen, there is only the one state of affairs in this sense which is specified by the one picture or CAE, and which may then be projected positively or negatively. There are no "negative" pictures or CAE's, and would then be very misleading to speak of them as being "positive" either. For the positive-negative dichotomy is logically subsequent to the picture or the CAE, and hence to the single state of affairs jointly specified by the positive and negative sentences.

These remarks are important, for if they are overlooked, we may be unwittingly led into a mistake that has afflicted philosophers since ancient times but which receives its most sophisticated formulation in the writings of Bertrand Russell

---

[22] Unless, of course, we shift the notion of "projection" to the sense of the sentence; then, a projection would be a "non-linguistic" entity. For our present purposes, however, it matters very little as to whether we construe a sentence or its sense as a "projection". None of the points made here contra "negative states of affairs" would be affected.

and other recent thinkers. It is, to wit, the idea that "negative facts" are somehow an indispensable kind of "entity" in logical analysis.[23] It is indeed natural to suppose that for a sentence (or its sense) to be true, there must exist something (a "fact" or "state of affairs") in the world ("reality") which renders it true or in virtue of which it is true; and once this is granted, it is difficult, if not impossible, to circumvent Russell's very convincing arguments.[24] For if the thought expressed by 'a is not F' is true, there are powerful arguments to the effect that the only "entity" which can correspond to the sentence and render it (or its sense) true is the (existent) fact or state of affairs: a's not being F[25]--a "negative" fact or state of affairs. However, natural though this supposition may be, it nevertheless remains that if the preceding discussion is correct, the

---

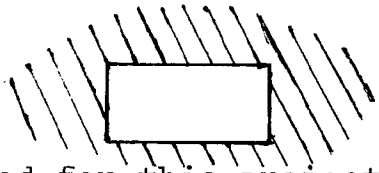
[23] See for example pps. 211-215 of his Logic and Knowledge, edited by R. C. Marsh (London: George Allen and Unwin, 1956). I think that both Russell and Raphael Demos, whose argument against there being negative facts is outlined by Russell in this passage, commit the mistake to be mentioned below. However, if my reading of Wittgenstein is correct, he did not hold the view argued by Russell.

[24] As Russell uses 'fact', it is not redundant (as it would ordinarily be) to speak of a "fact" as existing. Thus, 'It is a fact that p' does not here have the force of 'It is true that p'. I shall follow him in this usage--odd though it is. 'Fact', here, means the same as 'state of affairs' rather than 'existent state of affairs'.

[25] Or as it is sometimes rendered: "that a is not F".

supposition rests on a conceptual mistake. For as we saw, if 'state of affairs' (or 'fact') is used in what was called the "new" sense, then, though we can indeed speak of a "negative state of affairs", in this sense, all this can mean is 'negative sentence'--it cannot mean a non-linguistic entity which "corresponds" to the negative sentence, and which, if it exists, renders the sentence (or its sense) true. And if by 'state of affairs' we mean the "original" sense, then for the reasons already given, the terms 'positive' and 'negative' cannot legitimately be applied to such an entity, but only to the ways in which it can be projected. Of course these senses are easily confused--thus leading to Russell's "natural supposition"--if one fails first of all to draw the CAE-sentence distinction. For it is not a sentence as such to which a state of affairs, qua non-linguistic entity, corresponds, rather the question of "correspondence" properly arises only in connection with the purely descriptive dimension of a sentence; that is, with the CAE. And to say that a state of affairs "corresponds" to a CAE is to say no more than that it is specified or described by the CAE. On the other hand, the question of truth-value can arise only in connection with a projection of the state of affairs--the projection will either accord or fail to accord with "reality", thus either be true or false. It is therefore a mistake to raise the question of correspondence with respect to the projective dimension, i.e. with respect to the sentence per se.

It is a mistake, that is, to look for a further entity--in addition to the specified state of affairs (in the original sense of that phrase)-- in order to render the negative projection of that state of affairs (or its sense) true. To do this is to look for an explanation where none is required; it is like asking for an entity to render true the following projection of a rectangular figure:



All that is required for this projection to be true is that the actual figure specified by ("corresponding to") the figure drawn here be an open space in a solid. We do not require that there be another entity in addition to the specified figure which might then be characterized as: "that-the-figure-is-an-open-space-in-a-solid." Similarly, all that is "required" for the negative projection (or its sense) of the state of affairs, a's being F, to be true is that the state of affairs not exist, but we do not require the further entity: "that-the-state-of-affairs-does-not-exist", or alternatively, "that-a-is-not-F". By the same token, just as we do not require the entity: "that-the-figure-is-a-solid-in-an-open-place", so we do not require the entity:that-the-state-of-affairs-does-exist", or alternatively, "that-a-is-F".

But it ought to be clear by now that there is a sense in which it would be absurd to deny that there are negative facts. If all that is meant by 'negative fact' is what I mean when I

utter expressions like: "It is a fact that a is not F", then obviously there are negative facts in this trivial and philosophically unimportant sense. Nothing that has been said here is incompatible with this truism, and indeed it should be apparent that the truism can easily be explicated on the present analysis. All that has been argued here is that it is a mistake to suppose that we must admit the existence of negative facts as fundamental, irreducible entities in logical theory in order to make intelligible the truth of the thought expressed by a sentence like 'It is not the case that a is F'. I take it as obvious that we should try to avoid admitting such counterintuitive "entities" if possible, for if we do not, many of the puzzles raised at the opening of chapter two must be faced all over again, as well as the uninviting prospect of "disjunctive" facts, "conjunctive" facts, "general" facts, and so furth.

Negation, Projection, and Anscombe's Problem

Now at last we can return to Anscombe's truth-table. It will be recalled that the table looked like this:

I.

p	p	not-p
P	T	F
N	F	T

Anscombe then went on to say: "The proposition which is an element in a truth-function is thus introduced as having two senses, the positive and the negative, rather than two truth-values, T and F." And our original question was whether 'p'

indicates the same "kind of thing" in both its occurrences. Using the distinctions drawn thus far, table (I) may be interpreted as follows: 'p' in its first occurrence indicates what has been called the CAE, and in its second occurrence the positive sentence. The 'P' and 'N', then, will signify the different, possible, and complementary ways in which the state of affairs specified by the CAE may be projected, and the 'T' and 'F' will apply only to the sentences which share the CAE. The sentences, thus, express in turn the two different projections of the specified state of affairs, and inasmuch as the state of affairs will either exist or not, one of these projections must accord with the facts (be true) and the other not accord. Hence, table (I) specifies which sentence is which (i.e., which sentence expresses which sense--positive or negative) by indicating the condition under which it will be true. That 'p' in its second occurrence indicates positive sentence (the one expressing the positive sense) is shown by placing the 'T' in the P-row--which says, in effect, that it is that projection of the state of affairs specified by the CAE which is true when and only when that state of affairs exists; thus, it is that sentence wherein the specified state of affairs is projected as existing or "positively". For as we have already seen, to say that it is the sentence wherein the state of affairs is projected "as existing" or positively is to say that its sense may be legitimately

described as positive according to the independent criterion inasmuch as it is the positive sense that "affirms" the specified state of affairs.[26] Thus, the 'P' and the 'N' stand for the two possible ways in which the state of affairs specified by the CAE may be projected: as existing, and as not existing; and 'p' in its second occurrence indicates the sentence wherein that state of affairs is projected in the former way, and 'not-p' indicates the sentence wherein it is projected in the latter way. That they do so is indicated by the placement of a "T" in the P-row for 'p' and in the N-row for 'not-p', for in so doing, their respective senses--positive and negative--are identified. Anscombe's description of the table is therefore inaccurate, for nothing, strictly speaking, can be described as "having two senses, the positive and the negative". The CAE has one sense in virtue of which it specifies a state of affairs, and this sense is neither positive nor negative. The sense of p (in its second occurrence) is positive however, and the sense of 'not-p' negative. Speaking more loosely, though, it is natural to describe the CAE (which, contrary to what Anscombe says, is not itself a sentence or proposition)

---

[26] Thus, the notion of a sentence as a "positive projection" of a state of affairs--or in other words, projecting it "as existing" or "positively"--is explicated in terms of its expressing a positive sense. However, it must be kept in mind that this is an explication only of the notion of "projecting..as being the case" or "projecting..positively", it is not an explication of the notion of "projecting" as such. The explication of that lies in the arguments given earlier in conjunction with the picture and plane analogies.

as having two senses inasmuch as the state of affairs it specifies may be projected in two possible ways which may reasonably be described as "positive" and "negative".

Let us now return to the coin analogy mentioned earlier in this chapter. We noted certain similarities between truth-tables and "bet-tables", and perhaps now we can explicate them. Consider again the table:

II.

C	B_1	B_2
H	W	L
T	L	W

We will for the present regard the 'C' as standing for a particular coin instead of being a "coin variable", and 'B_1' and 'B_2' will then express the different, possible, particular bets constructed from C. The logical similarities between tables (I) and (II) may now be spelled out as follows: just as 'H' and 'T' signify the different ways in which a coin may land on a given toss, so 'P' and 'N' signify the different ways in which the state of affairs may be projected in an act of assertive utterance; and just as, for example, placing the 'W' of the B_1 column alongside the 'H' of the C-column indicates that the bet signified by 'B_1' is that bet which wins if and only if the coin comes up heads (in effect, that it is "the heads bet"), so likewise placing the 'T' of the second p-column alongside the 'P' of the first p-column indicates that the second 'p' represents that sentence whose sense is true if and only if the specified state of affairs exists--

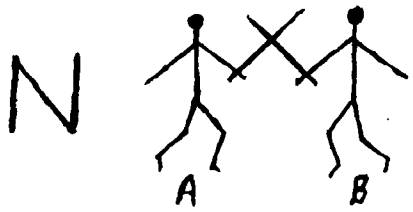
that is, that it expresses the positive sense. Thus, in the same way as table (I) tells us which of the two senses--positive and negative--are expressed by p and by ⌜not-p⌝, table (II) tells us which of the two bets--the "heads" bet and the "tails" bet--are expressed by 'B_1' and by 'B_2', for in the one case we are given the conditions under which the sense will be true or false, and in the other the conditions under which the bet will be a winning one or a losing one. Thus, the CAE stands to the sentence in much the same logical way as the sign for a coin stands to the sign for a bet made from the coin. For the sense of the bet-sign (i.e. which bet it expresses, the heads bet or the tails bet) is specified in table (II) in a way precisely analogous to the way in which the sense of a sentence (i.e. which thought is expresses, the positive or the negative) is specified in table (I), and this should serve to give us a pretty clear idea of why the two occurrences of 'p' in table (I) indicate different "kinds of things".

It will be recalled that in chapter two it was shown that on Frege's theory of assertion, the two different aspects of truth and falsehood, viz. qua success in assertive utterance and qua accordance with the facts, parallel the relation between the "win-lose" and "heads-tails" concepts respectively, and on this basis the theory was found defective. And it now appears that the proper analogy with the "win-lose" and "heads-tails" concepts lies between the latter aspect of "true-false"

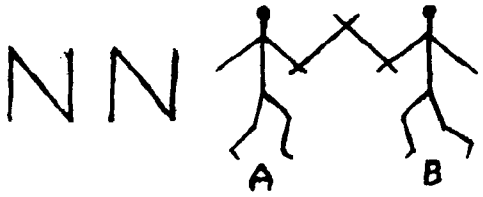
and the positive and negative methods of projection. Nevertheless, it ought to be kept in mind that there still remains a point of similarity between winning and truth in the former aspect, viz. that both are what constitute successful execution in their respective "games".

Extensional and Non-extensional Aspects of 'Not'

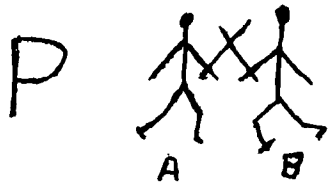
So far, we have explained 'not' as it occurs in singular sentences in terms of being an indicator of one of the two opposed ways in which the state of affairs specified by a CAE may be projected or logically "thought out". That is, it was analogized to the role played by 'N' in the illustration:



However, as the reader has no doubt already perceived, there is an obvious objection that warrants discussion at this point. For one thing, the account offered thus far seems not to be compatible with the Law of Double Negation, and more generally, the fact that 'not' can be reiterated any number of times. For on the view presented here, it seems quite mysterious as to what sense could be made of the sentence ⌜not-not-p⌝, and moreover, why it expresses the same sense as p (considered as the positive sentence, not the CAE). Since there are only two ways of projecting a state of affairs, the "illustration":



certainly cannot express a third kind of projection, and yet it would be absurd to suggest that it expresses the same projection as the positive illustration:



for it surely sounds unintelligible to say that projecting the state of affairs as not existing twice over is the same as projecting it once over as existing. And this fact certainly lends support to the view that our original intuitions concerning truth-tables were correct despite the points made so far. That is to say, that the p's and q's do stand for the same kinds of things in all their occurrences, and that the truth-function of negation is a function from truth-values to truth-values, and hence that the 'p' in ⌜not-p⌝ is itself a sentence or propositive, not a CAE.

Interestingly enough, many of these considerations are actually brought forth by Wittgenstein in the Notebooks, and strangely, argued in the very passages from which the quotations cited earlier were taken.

> Any proposition can be negated. And this shows that "true" and "false" mean the same for all propositions..
>
> Only a finished proposition can be negated. (And similarly for all truth-functions.)...

> Negation refers to the finished sense of the negated proposition and not to its way of presenting.
> If a picture presents what-is-not-the-case in the forementioned way, this only happens through its presenting that which is not the case.
> For the picture says, as it were, "This is how it is not", and to the question "How is it not?" just the positive proposition is the answer.
>
> The possibility of negating the negated proposition in its turn shows that what is negated is already a proposition and not merely the preliminary to a proposition. 27

By a "finished" proposition, Wittgenstein seems to mean the sentence as opposed to the CAE--it is "finished" inasmuch as the state of affairs has been projected as well as specified or described. And what he has in mind, I think, in speaking of the "presenting" aspect of a picture is what a picture has qua picture but not qua illustration, i.e. its purely "representative" aspect. Correspondingly, what he calls the "preliminary" to a proposition is just what we have been calling the CAE--the "unfinished" proposition implied in the earlier passage. And he is thus maintaining that the 'p' in '⌜not-p⌝' must be the positive sentence rather than the CAE.

Now all of this is perfectly correct and I do not wish to deny any of it. But I shall now try to show that, contrary to what it might at first seem, none of it does any damage to the points made so far, and indeed, complements them instead. Let us return to the coin-bet analogy and contrast table (II)

---

27 Notebooks, pages 21, 25, and 27. The italics are his.

with the following:

III.

B_1	B_2
W	L
L	W

If I know antecedently which bet B_1 is (i.e.know under what conditions it will win and lose), then I can specify B_2 by saying it is that bet which wins just in case B_1 loses and vice-versa. That is, I can say which bet B_2 is by specifying it in terms of B_1. However, if I specify B_2 in terms of B_1, then I clearly cannot specify B_1 in terms of B_2, and this means that to specify B_1, I must return to table (II) and say it is that bet which wins just in case the coin, C, comes up heads and loses otherwise. Thus B_2 can be specified in two different ways which we shall call the "direct" and the "indirect" ways. In the former, we specify the bet in terms of the coin coming up heads or tails, and in the latter, we specify it in terms of the conditions under which B_1 will win and lose. Secondly, if B_2 is specified indirectly (by means of table (III) instead of table (II)), then the sign 'B_2' is not the most perspicuous way of expressing to the bet in question. Rather, since it specified in terms of B_1, it is far more informative to express this bet in terms of the sign 'B_1'. We could, then, rewrite table (III) as :

IIIa.

B_1	$\#(B_1)$
W	L
L	W

where the '#' is a functional sign indicating that the bet expressed by the complex sign formed from the conjunction of '#' with the sign for some other bet always has as its value the opposite "win-value" of that other bet. '#' is therefore a sign with logical significance and '#()' is a genuine functional expression. Moreover, '#' can be reiterated; the sign '#(#(B_1))' expresses that bet which has the value, win, whenever #(B_1) has the value, lose, and vice-versa. Here then, we have a function from bets to bets (or from "win-values" to "win-values"), and the signs in table (IIIa) all stand for the same "kinds of things". And yet the relation between tables (IIIa) and (II) is quite clear, and certainly there is nothing about table (IIIa) which invalidates our earlier conclusions concerning table (II).

Now let us contrast table (I) with the following:

IV.

p	not-p
T	F
F	T

It can readily be seen, I think, that there is a paralled between tables (I) and (IV) on the one hand and tables (II) and (IIIa) on the other. To be sure, there are some dissimilarities too, and these will be discussed eventually, but I hope to show eventually that these dissimilarities do not in any way damage the analogies that do exist. Let us, then, first of all fasten upon the similarities.

Clearly 'not' plays the same role in table (IV) that '#' plays in (IIIa)--it is a functional sign such that when conjoined with a sentence produces a new sentence whose sense has the opposite truth-value from that of the original. The expression 'not', then, like '#', can be reiterated inasmuch as it indicates a genuine function from truth-values to truth-values (a way of producing, so to speak, one sense from another). ⌜Not-(not-p)⌝ is a sentence whose sense has the opposite truth-value from that of ⌜not-p⌝, and if ⌜not-p⌝ is a negative sentence (expresses a negative thought), then 'not-(not-p)' will be a positive sentence since it expresses the same (positive) sense as p. 'Not', then, as it occurs in table (IV), expresses a function by which a negative sentence may be produced from a positive one and vice-versa; thus, though it may be reiterated countless times, each sentence so produced will express one of only two possible senses--in the same way that any of the countless expressions formed with '#' will express one of only two possible bets.

More importantly, just as the "tails bet" can be specified directly or indirectly, so likewise can a negative sense be specified directly, as in table (I), or indirectly, as in table (IV). However, the relation between the direct and indirect specifications must be explored further, for the "p" in 'not-p', as it occurs in table (I), is the CAE, whereas in table (IV) it is the positive sentence. And what this shows, I believe, is that 'not', as it occurs in ordinary discourse,

simultaneously plays two conceptually distinct, logical roles. The first has already been discussed at some length; it is that 'not' functions as an indicator of a negative sense--of the state of affairs specified by the CAE being projected negatively as this notion has been explicated. In its second role, brought out by table (IV), it serves as a genuine functional sign by which a sentence expressing a negative sense may be constructed from one expressing a positive sense and vice-versa, and it is in this capacity that 'not-p' (or '~p') is introduced in logic texts. However, there is another feature of this role to be brought out, viz. that a negative sense is ultimately specified in terms of the corresponding positive one, just as bet B_2, the tails bet, is ultimately specified in terms of B_1. That is to say, though 'not' may be reiterated any number of times--sometimes to produce a negative sentence (and hence, negative sense) from a positive, and sometimes to produce a positive from a negative (as in ⌜not-(not-p)⌝ from ⌜not-p⌝)--its very first application must be to produce a negative from a positive, so that the positive sentence inserted into 'not-()' does not itself contain 'not'. Thus, a negative sense is derivative from a positive one--it is specified "indirectly" inasmuch as it is specified in terms of the positive sense being true or false, not the state of affairs being projected one way or the other.

Earlier, we compared 'not' with the functional sign '#' in table (IIIa). Let us now pursue this a little further. It

has already been noted that the CAE stands to its two complimentary sentences as the sign for the coin in table (II) stands to the signs 'B_1' and 'B_2', and it should be clear that we can dispense with the 'B_1' and 'B_2' significations and construct new signs for the bets in questions on the basis of the sign for the coin. Thus, table (II) can be rewritten:

IIa.

C	*C	#C
H	W	L
T	L	W

where the '*' in effect functions as an indicator of the kind of bet wherein heads is correlated with winning, and '#' is then an indicator of the kind of bet wherein tails is so correlated. '#C' is thus a sign expressing one of the two significant kinds of bets that can be made with the single coin C, and this parallels the function of the sign 'not-p' in table (I)--though not of course in table (IV)--inasmuch as 'not' there operates as an indicator of one of the two significant kinds of senses assertible with respect to a particular state of affairs p. In other words, '#C' expresses that particular "tails-type" bet which can be made with a particular coin C just as 'not-p' in table (I) expresses that particular "negative-type" sense which may be asserted about the particular state of affairs p. The sign '#', therefore, will play the very same dual roles in tables (IIa) and (IIIa) that 'not' plays in tables (I) and (IV). In table (IIa), the tails bet

is specified directly whereas in (IIIa) it is specified indirectly, and same applies to the negative sense with respect to tables (I) and (IV) respectively. It should be noticed, however, that if we are to carry out this new notation fully, then the sign for the tails bet in table (IIIa) must be altered once more. Instead of '#(B_1)', we will now write '#(*C)' to make it clear that we are expressing one bet in terms of another but where both involve the same coin. Thus, table (IIIa) now becomes:

IIIb.

*C	#(*C)
W	L
L	W

By the same token, the "p" in 'not-p' as it occurs in table (IV) is, as we have already seen, the positive sentence derived from the CAE, not the CAE itself. For in the 'not-p' of table (IV)--not table (I)--we are expressing one projection in terms of another, but it is the same state of affairs projected in the two cases. Now let us consider a language, "English-P", which is in every way like English but which contains in addition an explicit sign, 'pos', which functions as an indicator of a positive (type of) sense. Clearly in this language, the equivalent of table (I) would be:

V.

p	pos-p	not-p
P	T	F
N	F	T

and the equivalent of table (IV) would be:

	pos-p	not-(pos-p)
VI.	T	F
	F	T

Here, it is made explicit that 'not' when used to specify a negative sense indirectly (as in table (VI)) can attach only to a sentence, and the dual roles played by 'not' can be seen in a far clearer way in tables (V) and (VI) than in the corresponding tables for English. There is, thus, a striking parallelism holding between tables (IIa) and (IIIb) on the one hand and (respectively) tables (V) and (VI) on the other. And because of this, I think we are now in a better position to investigate the relation between the two roles of 'not' by first comparing them with the roles played by '#' in the more readily intelligible coin-betting case.

First of all, let us call the role played by '#' and 'not' in tables (IIIb) and (VI) respectively their extensional role, and the role played in tables (IIa) and (V) their <u>non-extensional</u> role. Thus in its extensional role, '#' serves as a functional sign which can attach only to a sign for a bet so as to produce a new such sign, and may be reiterated any number of times. It represents a function from "win-values" to "win-values" (in effect, from bets to bets) and has been defined in such a way that the new bet-sign thus produced expresses a bet which always has the opposite win-value from that of the bet expressed by the sign which occurs as the argument-expression. For this reason, the tails bet is

specified indirectly inasmuch as '#' must always attach to a bet-sign--not a "coin-sign"--and so that bet can be expressed only by '#(*C)' or any further expression containing an odd number of occurrences of '#'. Hence, if we are to give an English rendering of '#()' in its extensional use, clearly the phrase to choose would be 'opposite bet from...'. But to say that B_2 is the "opposite bet from B_1" is to say nothing about <u>which</u> bets B_1 and B_2 are--which is the heads bet and which the tails bet--or even whether they are different <u>kinds</u> of bets at all. For all we know, B_1 and B_2 might <u>both</u> be heads bets, though of course based on different coins, C and D let us say. And for some odd reason, it might just be that B_1, or *C has the value, win, whenever B_2, or *D, has the value, lose, and <u>vice-versa</u>. That is, whenever coins C and D are tossed, C always comes up heads whenever D comes up tails and <u>vice-versa</u>. The signs '#(B_1)' and 'opposite bet from B_1' only tell us that <u>whatever</u> B_1 is, B_2 will always have the opposite win-value, but in order to know that B_2 is in fact the <u>tails bet</u> based on C, we would at least have to reexpress 'B_1' as '*C' as a first step. But even now in the signs '#(*C)' and 'opposite bet from the heads bet based on C', we have not <u>uniquely</u> specified the tails bet based on C, for the heads bet based on D <u>also</u> has the value, win, whenever the heads bet based on C has the value, lose, and <u>vice-versa</u>. Thus, it is wrong to suppose that for any given bet, there will always be just one other bet which is <u>the</u> opposite bet

from the given one. Now there can be no doubt that the tails bet based on C and the heads bet based on D <u>are</u> distinct bets despite the misleading similarity in their "win-conditions". For one thing, they are based on different coins, and for another, our natural inclination is to say that the tails bet based on C stands in the "opposite bet" relation to the heads bet based on C <u>of necessity</u> (a coin must come up either heads or tails), whereas with the heads bet based on D it is only a rather unusual matter of fact. But the most telling point is that they <u>differ</u> in their win-conditions according to their <u>direct</u> specifications; that is, *D, or the heads bet based on D, is that bet which wins when the coin D comes up heads and loses when it comes up tails, but the tails bet based on C is just what the phrase suggests--it wins whenever coin C comes up tails and loses when it comes up heads. The direct specifications, in other words, show us that we have two different bets because it shows us that they are intrinsically different <u>kinds</u> of bets--a heads bet and a tails bet.

So far, then, we have established two things: first, the tails bet based on C and the heads bet based on D are distinct bets, and second, while '*D', as introduced in the table:

D	*(D)
H	W
T	L

suffices to directly specify the heads bet based on D, the extensional use of '#'--as it occurs in '#(*C)', '#(#(#(*C)))',

etc. and meaning only "opposite bet of..."--does not suffice to uniquely mark off the tails bet based on C from all other possible bets. Clearly, then, something more is required, and it should now be apparent that what is needed in addition to the extensional use of '#' by which the tails bet is indirectly specified is the non-extensional use wherein it is specified directly; that is, the use it has in table (IIa). Whereas the extensional use of '#' tells us nothing about the bets B_1 and B_2 such that $B_2=\#(B_1)$, the whole point of the non-extensional use as it occurs in (IIa) is to function as an indicator of a "tails-type" of bet--it expresses one of the two possible kinds of bets based on a single coin C, and hence '#' attaches to the sign for the coin and not the sign for the other bet. Therefore, as far as identifying or specifying a tails bet is concerned, the non-extensional use of '#' takes up the slack left by the extensional use. That is to say, the indirect specification of a tails bet cannot always do the job we want it too, viz. uniquely specify that bet, but the direct specification will. Hence, whereas '#(*C)' can be rendered 'opposite bet from the heads bet based on C', the sign '#C', where '#' now attaches to the sign for the coin, simply expresses the tails bet based on C, i.e. that bet based on C such that it wins if and only if C comes up tails.

Now instead of talking about coins, bet-signs, bets, win-values, and '#', we can state all of this in terms of states of affairs, sentences, their senses, truth-values, and 'not'

respectively--with the result of shedding a great deal of light on the nature of negation with respect to singular sentences or propositions. It should be clear already that some of our conclusions will be the same as those developed in the discussion of negation back in chapter two. The point of restating them here is threefold: first, to state these conclusions in a clearer and more incisive way; second, to dispose of the apparent incompatibility between the use of 'not' in table (I) (or table (V)) and its use in table (IV) (or table (VI)); and third, to exhibit these conclusions as a consequence of the analysis of singular sentences developed in this chapter, and hence that this analysis enables us to resolve the questions about 'not' raised in chapter two in connection with Frege's extensional account of negation.

First, then, in its extensional use 'not' serves as a functional sign (a "logical connective") which may be iterated any number of times, and which can attach only to a sentence as as to produce a new sentence. It is a function from truth-values to truth-values (in effect, from senses to senses), and has been so defined that this new sentence expresses a sense which always has the opposite truth-value from that of the sentence occurring as argument-expression. Thus in tables (IV) and (VI), the negative sense is specified indirectly inasmuch as 'not' must always attach to a sentence, and that sense can only be expressed (in English-P) by ⌜not-(pos-p)⌝ or any other such expression containing an odd number of

occurrences of 'not'. It should be clear, then, that in its extensional use, 'not-()' means "opposite sense of...", or more naturally, "contradictory of...". But as in the betting example, to say of two senses, S_1 and S_2, that S_2 is the "contradictory of S_1" is to say nothing about which kinds of senses they are--positive or negative--or even, for that matter, whether they are different kinds at all. As we saw in chapter two, S_1 and S_2 might both be positive senses--derived, let us say, from the different though "exhaustive"CAE's, p and q--and because of their exhaustive feature, the sense of ⌜pos-p⌝ (or S_1) has the value, true, whenever the sense of ⌜pos-q⌝ or S_2 has the value, false, and vice-versa. Our example in chapter two, stated now in our present terminology, was: 'a is (pos) hot' and 'a is (pos) cold', where hot and cold jointly exhausted the thermal spectrum. The sign 'not-'[28]

---

[28]There is a fairly obvious point to be raised with respect to construing 'not-()' or 'it is not the case that ()' as meaning simply 'contradictory of the thought that ()'. For the phrase 'contradictory of the thought that Socrates is wise'--with or without a preceding "the"--is not a sentence whereas 'it is not the case that Socrates is wise' certainly is a sentence, and the two expressions thus fail the test of intersubstitutability salva veritate. However, this does not undermine the points made thus far, for 'not', in the extensional use, expresses a truth-function, and as we shall see in chapter four, truth-functions are examples of what Wittgenstein calls operations. And it is characteristic of operations that their values are always of the same logical type as their arguments. Thus, strictly speaking, 'not-S_1' cannot be rendered 'contradictory of the thought that S_1', but for expository purposes, I shall do so anyway. See the discussion of operations in chapter four.

S_1', where 'not' occurs in its purely extensional use and thus meaning in effect 'contradictory of S_1', merely tells us that <u>whatever</u> sense S_1 is, S_2 will always have the opposite truth-value. But if in fact S_2 is the <u>negative</u> of S_1, then in order to know that S_2 is the negative sense derived from the CAE, <u>p</u>, we must first reexpress S_1 as 'pos-<u>p</u>'. But even now in the signs 'not-(pos-<u>p</u>)' and 'contradictory of (the thought that) pos-<u>p</u>', we have not <u>uniquely</u> specified the corresponding negative sense--not, that is, if 'not' is construed in its purely extensional use. For the positive sense derived from the CAE, <u>q</u>, also has the value, true, whenever the positive sense derived from <u>p</u> has the value, false, and <u>vice-versa</u>. We cannot say, the, that there is just <u>one</u> sense which is <u>the</u> contradictory of ⌜pos-<u>p</u>⌝--no more than we can say that there is just one number which is <u>the</u> square root of 4. It is clear, however, from the arguments employed in chapter two that the positive sense derived from <u>q</u> is a <u>different</u> sense from the negative sense derived from <u>p</u>; that is, that the sense of '<u>a</u> is (pos) hot' is distinct from that of 'Not: <u>a</u> is (pos) cold' despite the misleading similarity in their truth-conditions. Those arguments will not be repeated here; instead, we need only note that the fact that the two senses are distinct follows from the difference in their <u>direct</u> specifications. The sense of ⌜pos-<u>q</u>⌝ is that sense which is true when and only when the state of affairs specified by the component CAE exists (⌜pos-<u>q</u>⌝ is in effect the "positive projection" of the state of affairs) whereas the negative sense

derived from p is just what the phrase suggests, viz. that sense which is true when and only when the (distinct) state of affairs specified by p does not exist.

Thus, we have established that ⌜pos-q⌝ is a distinct sense from the negative sense derived from p, and moreover, though the sign 'pos-q', as introduced in the table:

q	pos-q
P	T
N	F

suffices to specify the positive sense derived from q, the extensional use of 'not'--wherein ⌜not-(pos-p)⌝ means only 'contradictory of (the thought that)pos-p' or 'opposite sense from that of 'pos-p'--does not suffice to mark off uniquely the negative sense derived from p from all other possible senses. And it should now be clear that what is needed to accomplish this task is the non-extensional use of 'not' wherein this sense is specified directly, i.e. the use it has in tables (IV) and (VI) and which has been spelled out by means of the plane and picture analogies. For the extensional use of 'not' tells us nothing about the senses S_1 and S_2 such that S_2=not-S_1, but the whole point of the non-extensional use as it occurs in (I) and (V) is to function as an indicator of a "negative-type" of sense--it expresses, that is, one of the two possible methods of projecting the state of affairs specified by the component CAE, and for that reason, 'not' attaches to the CAE instead of the sentence expressing the

other sense. Thus, in order to specify a negative sense, we must fall back on the non-extensional use of 'not', for the extensional use as such gives us only an indirect specification of that sense (what was called in chapter two Frege's "purely extensional account" of negation in terms of the contradictory relation)--a specification that a sense other than the negative sense in question could satisfy. The non-extensional use, on the other hand, captures what was called the "intrinsic" or "non-extensional" features of negative thoughts or senses. Thus, whereas 'not-(pos-p)' in the extensional use means only 'contradictory of the thought that pos-p', the sign 'not-p' in tables (I) and (V) simply expresses the negative sense derived from p. Again, a negative sense is not merely a contradictory of a positive one (another sense could also be one: hence 'a', not 'the'), it is also the negative of that sense. And it is so because it can properly be characterized as negative per se--not just as the negative of the other sense.[29]

It should be noted that in English-P, 'pos' may also have both an extensional and non-extensional use. We are already familiar with its non-extensional use; where 'p' is a CAE, ⌜pos-p⌝ will be a sentence expressing a positive sense,

---

[29] However, 'the negative of the thought that ()', like 'contradictory of the thought that ()', can also produce a positive sense from a negative one. That is, the sense of ⌜pos-p⌝ is also "the negative of" the sense of ⌜not-p⌝, for it is the same sense as that of ⌜not-(not-(pos-p))⌝. More of this in chapter four.

and can occur as an argument-expression in 'not-()' where 'not' occurs extensionally. However, we might also introduce a truth-function signified by 'pos-()', and here 'pos' will of course be used extensionally. The function in question is that function which has the value, true, if and only if the argument has the value, true. Thus, the sense of the derived value-expression will always have the same truth-value of the sense of the argument-expression. 'Pos-()', therefore, will operate like Frege's "horizontal" (i.e. '——()') except that the only admissible argument-expressions are those whose senses have a truth-value, and of course that the resulting value-expression will not refer to a "truth-value" or any other "object". The sign 'pos-()' will thus express the idea behind the claim that "every proposition is a truth-function of itself". Hence, where the left-hand occurrence of 'pos' is used in this extensional way (i.e. to indicate the truth-function specified above), ⌜pos-(pos-p)⌝ expresses the same (positive) sense as does ⌜pos-p⌝, and ⌜pos-(not-(pos-p))⌝ expresses the same (negative) sense as ⌜not-(pos-p)⌝, where of course 'not' is used extensionally as well. The extensional uses of 'pos' and 'not', then, express the two significant[30]

---

[30]It should be clear that the foregoing discussion concerning the extensional and non-extensional roles of 'not' and the related distinction between direct and indirect specifications of sense suffice to resolve the problems raised at the beginning of this chapter concerning matrices and their interpretation. More specifically, they enable us to see just why the "general thesis" discussed on page 165 is false.

truth-functions of one argument (excluding, that is, tautology and contradiction.

Some Consequences of the Foregoing Analysis

A great deal more needs to be said about the extensional and non-extensional uses of 'not', and how, specifically, they relate to our actual use of 'not' in ordinary English. But since these topics have a direct bearing on the notion of assertion and the proper analysis of the singular sentence or proposition, they will be discussed in chapter four. In the meantime, we shall examine some of the more immediate consequences of what has already been said. Now in chapter two, we had occasion to mention a language, "English-N", wherein the senses of any two sentences of the form, p and ⌜not-p⌝, were "reversed", and the issue was whether or not there could be a language-game in which the speakers aimed for falsehood instead of truth. That discussion, and the conclusions drawn, were prompted by the following passages in the Tractatus:

> 4.062 Can we not make ourselves understood with false propositions just as we have done up till now with true ones?--So long as it is known that they are meant to be false.--No! For a proposition is true if we use it to say that things stand in a certain way, and they do; and if by 'p' we mean not-p and things stand as we mean that they do, then, construed in the new way, 'p' is true and not false.
>
> 4.0621 But it is important that the signs 'p' and '~p' can say the same thing. For it shows that nothing in reality corresponds to the sign '~'. The occurrence of negation in a proposition is not enough to characterize its sense.

The second and third sentences of 4.0621 have not been fully explained as yet, and we shall have to wait until chapter four for that explanation; however the other remarks should be fairly clear in light of the foregoing discussion. Even so, the notion of a language like English-N raises several interesting problems which we are now in a position to handle. Perhaps the best way of presenting these problems is to examine Max Black's critical comments of 4.0621. Black sets up the notion of English-N--which he calls "E'" in contrast to "E" which of course is English--and uses the sign '¬' where we have used 'not'. He then says:

> W.'s thought now seems to be that we might discover the following semantical correlations:
> p (in E') is true when and only when ~~p (in E) is true.
> ¬p (in E') is true when and only when p (in E) is true.
>
> If this were true, we might certainly be inclined to say that the role of '¬' in E' is exactly the reverse of the role of the '~' in E. Just as the presence of '~' in E means negation, so the absence of '¬' in E' means negation--and correspondingly for affirmation.
>
> But let us pursue the matter further. It is a characteristic of '~' in E that ~~p has the same truth-conditions as p while '~p' does not. But, according to the second of our semantical correlations, ¬p in E' has the same truth-value as p in E and ¬¬p therefore has the same truth-value as ¬p, i.e. the same truth-value as p again. In other words, repeated applications of the angle sign '¬' in E' reduce to a single application of it. On the other hand, nothing answers in E' to the repeated applications of the curl '~' with alternate changes of truth-value in E. Since the correlate in E' of ~p in E is simply p, in which there is no operator sign present, it is impossible to show in E' anything answering to the

> repeated application of the curl sign in E, and there is no correlate in E' for the distinction in sense between ∿∿p and ∿p.
>
> All this being so, we would have no right to say that the angle sign in E' constituted a negation sign used in reverse. The proper verdict would be that its use in E' constituted a special device for negating or affirming a single proposition, with no provision for the possibilities of repeated application that define the formal properties of the curl operator in E...
>
> My conclusion is that there is an important sense in which W. was wrong in saying that p and ∿p "can say the same thing" (though also a trivial sense in which he was right). Given our present meaning for '∿', we cannot "reverse its role" in the manner supposed. 31

Black's main point seems to be that since '¬' can not admit of alternate changes of truth-value on successive re-applications, that p not ⌜not-p⌝ cannot be used to "say the same thing" in English and English-N (E')respectively. However, Black's premiss is mistaken, for the iterability feature applies only to 'not' in its extensional use, and '¬' could (though not necessarily) signify the very same truth-function in English-N as '∿' signifies in English, yet English-N would still obey Black's semantical correlations. What makes English-N the "negative language" of English is the fact that p expresses a negative sense rather than a positive one. That is, just as in English there is no explicit indicator of a positive sense, so likewise there is in English-N no explicit

---

31 Max Black, A Companion to Wittgenstein's Tractatus (Ithaca: Cornell University Press, 1964) p. 180.

indicator of a negative sense. But in its extensional use (i.e. as a "logical connective") '¬' in English-N can mean the same as '~' or 'not' in their extensional uses in English; roughly, "contradictory of...". Thus, if p in English-N expresses a negative sense, ⌜¬p⌝ will express a positive one if '¬' does in fact signify the very same truth-function, and by the same token, ⌜¬¬p⌝ will express the same (negative) sense as p, but not the same as ⌜¬p⌝. Hence, 'not' (or '~') in its extensional use in English would mean the same as '¬' in English -N, viz. a function from truth-values to truth-values such that both ⌜¬p⌝ (in English-N) and 'not-p' (in English) express a sense which is true if and only if the sense of p is false and vice-versa. However, 'not' in its non-extensional use in English--that is, to specify (directly) the negative sense with respect to a certain state of affairs--does not mean the same as '¬' in what would be its non-extensional use in English-N. For in the non-extensional use of '¬', ⌜¬p⌝ expresses in English-N a positive sense, i.e. it is synonymous with 'pos' in its non-extensional use in English-P. Thus, ⌜¬p⌝, regardless of whether '¬' occurs extensionally or non-extensionally (as attaching to a sentence expressing a negative sense or to a CAE), expresses the positive sense with respect to the state of affairs specified by the CAE in question; in the former case, that sense is specified indirectly in terms of the sentence p which expresses the corresponding negative sense, but in the latter,

it is specified directly in terms of p--this time construed as the CAE. Likewise, ⌜¬¬p⌝ will express the same (negative) sense as the sentence p, contrary to what Black says; for there is nothing wrong from a logical point of view in supposing that '¬' in its extensional use in English-N can mean exactly the same as 'not' in its extensional use in English even though their non-extensional uses are directly opposed to each other. Of course, '¬' could be used extensionally to mean just what Black wants it to, viz. the same as what 'pos' in its extensional use in English-P means--specifically, a function from truth-values to truth-values such that the sense of ⌜¬p⌝ always has the same truth-value as the sense of p; in other words, Frege's "horizontal" with the qualifications noted earlier. Then obviously Black would be correct in saying that '¬' does not operate in English-N as 'not' or '∿' do in English. But as we have now seen, he is quite wrong in supposing that '¬' (in its extensional use) must operate this way in English-N in order for that language to satisfy the required semantical correlations. The reason for his mistake should now be clear: he has failed to draw the extensional-non-extensional distinction; but once the distinction is made, it is then easy to see that it is quite possible for '¬' in English-N to signify in its extensional use the very same truth-function signified by 'not' in English in its extensional use rather than the function Black thought it had to signify, and at the same time for

English-N properly to be the "negative language" with respect to English (i.e. to satisfy the "semantical correlations" as Black has stated them).

To expand on this, let us now drop Black's angle sign and use only 'not', and let us further consider a language, English-NP, which is just like English-N except for containing an explicit indicator of a negative sense. That is, English-NP stands to English-N as English-P stands to English. We can now express the difference between English-N and English more clearly by contrasting English-NP with English-P. It will be recalled that in table (V), both 'pos' and 'not' are used non-extensionally, and the negative sense is thus specified directly as well as the positive one. In table (VI) 'not' occurs extensionally, and hence the negative sense is specified indirectly. Now let us suppose that just as English-P contains 'pos' to explicitly distinguish the sentence expressing the positive sense from the CAE (something not done in table (I)), so likewise English-NP contains 'pos' to distinguish the sentence expressing the negative sense from the CAE (something not done in English-N). Then, corresponding to table (V) for English-P, we would have for English-NP:

p	not-p	pos-p
P	T	F
N	F	T

and thus in their non-extensional uses ⌜not-p⌝ in English-NP has the same sense as ⌜pos-p⌝ in English-P and vice-versa.

And now we can reexpress the objection to Black's remarks in a clearer way. The extensional use of 'not' in English-P is given in table (VI), and it should now be apparent that this use in English-NP can be exactly the same as for English-P; that is, table (VI) would also specify the extensional use of 'not' for English-NP as well as English-P. The difference between the two languages lies of course in the fact that ⌜not-(pos-p)⌝ in English-P expresses a negative sense whereas it expresses a positive sense in English-NP, but this is because the argument-expression ⌜pos-p⌝ as it occurs in the complex expression expresses a positive sense in English-P and a negative sense in English-NP. Nevertheless, the sense of 'not-()', as it occurs extensionally in both English-P and English-NP, is the same for the two languages--in other words, it signifies precisely the same truth-function of one argument for English-P and English-NP. Therefore, that the two languages are, so to speak, "negatives" of one another consists solely in the fact that the senses of ⌜not-p⌝ and ⌜pos-p⌝ in their non-extensional uses are "switched around". Again, 'not' in its extensional use in English-NP could have meant what 'pos' in its extensional use in English-P means, viz. the truth-function specified by the following table:

pos-p	not-(pos-p)
T	T
F	F

and then, as Black rightly notes, 'not' (or '¬') in English-N

and English-NP would not admit of alternate changes of truth-value on repeated applications. But the important point is that it can in its extensional use express the same sense in both English-P and English-NP, and that suffices to destroy Black's objection to the remarks at 4.062 and 4.0621. The notion of a language like English-N where ⌜not-p⌝ expresses the same sense as p in English and vice-versa, and where 'not' also admits of alternate changes of truth-value on repeated applications is an eminently intelligible notion. Therefore, p and ⌜not-p⌝ can say the same thing in the very strong sense in which Wittgenstein intended that remark.

As I read 4.062, Wittgenstein is saying that it is impossible for there to be a community the speakers which aim for falsehood when making assertions. Instead of using, say, English and aiming for falsehood, their linguistic behavior ought properly to be described in terms of using English-N and aiming for truth. This, with the qualifications stated there, was the same conclusion reached in chapter two after a rather lengthy course of argument, and it follows that it is at least logically possible for there to be a community which speaks English-N or some other such "negative" language. We might, then, ask why in fact there is no community speaking such a language.

However, let us first examine the question more closely. By 'English-N' or 'French-N' or 'German-N' etc. will be meant a language such that 'not', 'ne...pas', 'nicht', etc. in their

extensional uses mean just what they do in English, French, and German respectively, viz. a function from truth-values to truth-values such that the value of the function for a given argument is always the opposite truth-value from the argument. However, in their non-extensional uses, 'not, 'ne...pas', 'nicht', etc. serve to express a positive sense rather than a negative one, and moreover there is no explicit indicator for the corresponding negative sense--just as in English, French, and German there is no indicator of a positive sense. Clearly we can generalize beyond the case of existent natural languages to the case of a community which speaks such a negative language but where there is no other community which employs the corresponding "positive" one. Now though it is logically possible to come across a community speaking such a negative language, the chances are, to say the very least, remote; and to now returnto the original question, let us ask why. Why, that is, are there in fact no negative languages in use?

Now I think that intuitively we would all be inclined to say that there is a "reason" why there are no such languages, for it is certainly too much to suppose that it is just coincidence that all of us speak "positive" languages. And here we have finally reached the point where the coin analogy breaks down--though not in a way that damages the similarities already noted. For it is a purely arbitrary matter as to whether I specify the tails bet in terms of the heads bet or vice-versa. In fact, it was the tails bet which was specified

indirectly in our example, but it could just as easily have been the other way around. But the inclination to feel that there is a reason why there are no negative languages in use unquestionably indicates that we feel it is not "just as easy" to specify the positive sense indirectly instead of the negative, but this is of course just what a negative language does. Now I do not think we have to look very far to find the reason why it is the positive sense that is specified directly in natural languages rather than the negative--in fact, we have already touched on it. It is that there is an essential "indefiniteness" about a negative sense inasmuch as it affirms no particular alternative and leaves open all of the alternatives of a given range except the one specified by the component CAE--which is of course excluded. And we expressed this earlier by saying that a negative sense is definite only about what is <u>not</u> the case. Now with a coin, the heads and tails sides are logically on a par with each other; there are no logical features pertaining to one of them but not to the other. But the independent criterion and the picture and plane analogies make it quite clear that there is an important logical contrast to be drawn between a positive and a negative sense. On the latter analogy, the CAE is like the inscribing of a figure on a plane--in Wittgenstein's model, the CAE marks out a definite area in "logical space". The figure thus determines two areas, but only the interior area is definite, the surrounding, exterior area is limitless.

And it is this latter area, which encompasses all the remaining alternatives to the given state of affairs, which is left unshaded in the drawing corresponding to the negative sense. In this way, a negative sense is indefinite about what is the case. But a positive sense, since it leaves nothing open, is definite both about what is and what is not the case inasmuch as however many further alternatives there are to the specified state of affairs (however many further areas could be drawn on the plane), each and every one of them does not obtain if the positive sense is true. Just as the CAE marks out a definite area of logical space (specifies a particular state of affairs), so the positive sentence (that projection which expresses a positive sense) indicates something definite about both the interior and exterior areas. There would seem then, to be an intimate connection between the CAE and its positive sentence--so much so, in fact, that in natural languages the two are not explicitly distinguished.

I doubt if any of this can be made much clearer--we seem to have hit rock bottom as far as explanations are concerned--but it might help if we reexamine the picture analogy. A picture represents a state of affairs--something which answers to the generalized description: "such-and-such being so-and-so". And the state of affairs may be projected or logically thought out in two possible ways: as being the case or as not being the case. Moreover, it is natural to want to re-express these possibilities as "such-and-such being so-and-so"

and "such-and-such not being so-and-so". The first of these expressions thus can be used to characterize both the positive illustration (and by analogy, the positive sentence) and the state of affairs independently of its being projected or thought out at all. Of course 'being' carries a different sense in the two cases; in the one case it merely serves to give a generalized characterization of the notion of a state of affairs, while in the other it carries the force of 'exists' or 'obtains'. Even so, these uses of 'being' are easily confused, thus giving rise to a great many philosophic problems.[32] The fact that these two ontological notions are not readily distinguished, together with the point made earlier that a positive sense is definite about what is as well as what is not the case, perhaps provides us with the "reason" why in natural languages the CAE and its positive sentence are never distinguished from one another, and why the positive sense is specified directly instead of the negative. It should not be surprising, then, that there are in fact no "negative" natural languages, nor that the natural languages that do exist contain no explicit indicator of a positive sense.

---

32 Among other things, giving rise to the notion of "positive" and "negative" facts or states of affairs. But as we have seen, there is no kind of entity satisfying the generalized description: "something's-not-being-so-and-so". However, we can think out something's-being-so-and-so as not being the case, for this is just what a negative sense is.

So whereas the heads and tails possibilities with respect to a coin are on an equal logical par with each other, there is no such parity with respect to the "being-the-case" and "not-being-the-case" possibilities with respect to a state of affairs. And this shows us why it is not a purely arbitrary matter as to whether we specify the negative sense indirectly or the positive one. But the conclusions arrived at earlier by exploring the coin analogy still holds, for they merely depend on the fact that "heads" and "tails" are the only two possibilities on a toss of the coin and that "being the case" and "not being the case" are the only two possibilities with respect to projecting a state of affairs. Hence, the heads and tails bets stand to each other in a broadly-speaking "logical" relation which is the precise analogue of the contradictory relation for the positive and negative senses.

There are still some questions to be resolved concerning the extensional-non-extensional distinction as it has been developed thus far. They will constitute the first order of business in the next chapter. Thereafter, we will consider some of the further ramifications of that distinction, for I hope to show that it will lead us to some important insights concerning the topic of assertion, and indeed, concerning the logical nature of the sentence or proposition as such.

> He [Wittgenstein] also told me how wonderful Bertrand Russell was in those days...For a long time they could agree about nothing, but at last they both came to see that the "assertion sign" which Whitehead and Russell took over from Frege, had no meaning whatever. This they both regarded as a great triumph...
>
> --Karl Britton in Portrait of a Philosopher (his italics).

CHAPTER FOUR

SENTENCES, OPERATIONS, AND ASSERTION

Extensional and Non-extensional Aspects of Negation in English

We have distinguished two "uses" of 'not', the extensional and the non-extensional. However, this distinction is of a largely conceptual nature as far as ordinary English is concerned. That is, it would be wrong to suppose that given an occurrence of 'not' in a singular sentence of English, we can then raise the question of whether it occurs extensionally or non-extensionally. Rather, it is not an "either-or" matter at all; the extensional versus non-extensional distinction brings to the surface two distinct but intimately related logical features of 'not' as it occurs in ordinary discourse. For inasmuch as the sentence:

1. Socrates is not wise

or the equivalent form:

1a. It is not the case that Socrates is wise

expresses a negative sense according to the independent

criterion, and is not merely the contradictory of the sense of 'Socrates is wise', it is clear that these sentences can be logically analyzed or decomposed in the manner outlined in the previous chapter, to wit:

2. It is not the case: Socrates' being wise.

Here, 'Socrates' being wise' functions as what was called the "CAE" (the purely descriptive dimension of a sentence) and 'it is not the case' supplies both the "projective" dimension, wherein the state of affairs specified by the CAE is projected negatively, and also the indicative mood. thus forming an expression whose sense is a bearer of truth-value. Here, 'it is not the case' functions just like 'N' in the "illustration" mentioned in chapter three, and by analyzing (1) or (1a) into (2) we specify the (negative) sense in question "directly" according to the table:

Socrates' being wise	Socrates is not wise
P	F
N	T

On the other hand, inasmuch as (1) and (1a) can be themselves negated so as to produce sentences which express a positive sense, it is plain that they can be analyzed in the standard, textbook manner. That is, where the more logically perspicuous (1a) is broken down into 'it is not the case that...' and 'Socrates is wise'. The unsaturated expression is a genuine functional sign expressing that truth-function of one argument usually represented by the tilde in mathematical logic. The latter expression is not a CAE, but a sentence

which expresses the opposite (positive) sense from that of (1a). Hence, on this analysis the sense of (1a) is specified "indirectly" according to the table:

Socrates is wise	Not: Socrates is wise
T	F
F	T

Thus, 'not', where it has a single occurrence in an ordinary, singular English sentence, serves both as an indicator of a negative sense and to express a particular truth-function of one argument. A satisfactory logical analysis of negation in natural languages such as English must account, first, for the fact that 'not' (and its equivalents in other languages) may be reiterated any number of times; second, for the fact that however many such iterations are made, the same two senses are alternately re-expressed with corresponding change of truth-value; and third, for the fact that these two senses may properly be described as <u>the</u> negatives of one another. And this last point, as we saw long ago, requires the positive-negative distinction as it has been expressed in the independent criterion. Now the extensional and non-extensional features of 'not', as they were introduced in the chapter three, do, I feel, suffice to give us the desired account, but neither of these features by itself is sufficient, and it would thus be a mistake to ask whether 'not', as it occurs in (1) and (1a), is used extensionally <u>or</u> non-extensionally. We can for philosophical purposes isolate these features, and having done so speak of them

(as in chapter three) as "uses". Thus, we can say that 'not' is used (or occurs) non-extensionally in (2), and is used (or occurs) extensionally in (1a) where (1a) is here regarded as a value-expression formed from 'it is not the case that...", a synonym for Frege's ' $\neg\xi$' or the '$\sim()$' of present day logical notation. But if (1a) is regarded not as an analysis of (1), but simply as a synonym requiring an analysis, then it would be wrong to suppose that 'not' is "used" or "occurs" in just one of these ways to the exclusion of the other.

However, this matter should be pursued a little further. Though an occurrence of the English 'not' in a sentence like (1) or (1a) may be said to play two distinct logical roles--and thus that the negative sense they express can be regarded as having been specified either directly or indirectly--it nonetheless remains that because of the extensional role, wherein the negative sense is specified indirectly by means of the function expressed by 'not' and the corresponding positive sense, the positive sense acquires a certain logical primacy over the negative. As we saw at the end of the preceding chapter, English is so to speak a "positive language"; it is in terms of positive senses that the truth-conditions for negative and other senses are specified. In English-N, on the other hand, the truth-conditions of positive and other senses would be expressed in terms of negative ones. This primacy of the positive sense over the negative in

natural languages introduces certain complications with respect to the foregoing remarks, and we ought now to attempt at least briefly to unravel them.

Though a positive sense can be expressed by many logically distinct varieties of sentences,[1] it is paradigmatically expressed (in English) by singular, uncompounded sentences like:

3. Socrates is wise.

Such a sentence, first, contains no explicit indicator of a positive sense, and second, does not express its sense as a (truth-) function of the sense of any other sentence, i.e. it contains no expression which could be called a "logical connective". In both these respects, then, (3) differs from (1) and (1a), and it is in terms of sentences like (3) and the senses they characteristically express that the senses of logically complex sentences are specified. Since (3) contains no logical connective, it cannot be analyzed "extensionally" in the manner of (1) and (1a), but inasmuch as it projects the state of affairs specified by the component CAE "positively" or as "being the case", its non-extensional analysis would be:

4. It is the case: Socrates' being wise,

where this matches (2) as an analysis of (1) and (1a). This

---

[1]E.g. 'It is not the case that Socrates is not wise' and 'Either Socrates is wise or Socrates is wise.'

much of course is clear from chapter three, but let us reconsider the language there described as "English-P". Unlike ordinary English, English-P contains an explicit sign for a positive sense, viz. 'pos'. Thus the equivalents of (3) and (4) respectively would be:

5. Socrates is pos wise.

6. It is pos the case: Socrates' being wise.

Notice that 'pos', as it occurs in (5), has a purely non-extensional occurrence in marked contrast to 'not' as it occurs in (1). For here, 'pos' serves only to indicate one of the two possible projections the state of affairs in question; it does not exhibit the sense of (5) as a truth-function of some other thought, and hence cannot be construed as a logical connective. We do not derive (5) from (3) in English-P by the insertion of 'pos', rather we have (5) instead of (3) in that language; that is, (5) serves as the basic sentential unit for constructing further sentences in English-P just as (3) does in English. 'Pos' is therefore used in a purely non-extensional capacity in (5) to express in English-P what is left implicit in English. This fact is important, for it shows that in English-P we can classify occurrences of 'pos' and 'not' as extensional or non-extensional whereas in English, as we saw, the occurrence of 'not' in a sentence like (1) or (1a) does not admit of such one-way classifications. Let us now follow this up in more detail.

Like English, English-P is a "positive language", and this means that in its extensional role, 'not', in its initial application to a sentence, is used to express a negative sense indirectly in terms of a positive one rather than vice-versa as in English-N. Thus, unlike English, we can classify occurrences of 'not' in English-P as extensional or non-extensional according to whether the linguistic unit to which it attaches contains an occurrence of 'pos'--thus indicating whether the unit is itself a sentence or not. Hence, in the sentences:

7. Socrates is not pos wise, and

7a. It is not the case that Socrates is pos wise,

'not' occurs in a purely extensional role, for the occurrence of 'pos' makes quite clear the fact that it is the sentence 'Socrates is pos wise' to which 'not' is being attached, not the CAE. Thus, 'not' here functions only as a logical connective, not as a direct indicator of a negative sense. In other words, the negative sense expressed by (7) and (7a) is specified indirectly owing to the occurrence of 'not' as a logical connective--it is specified in terms of (as a truth-function of) the positive sense expressed by (5), not as one of the two possible ways of logically conceiving or projecting the specified state of affairs. If there were a means of directly specifying a negative sense in English-P, then (1) and (1a) would probably serve as that means, for there the lack of 'pos' would reveal that 'not' does not

express a truth-function from thoughts to thoughts inasmuch as it does not attach to a sentence. Rather, as the analysis into (2) shows, it serves as an indicator of one of the two possible projections of the state of affairs specified by the CAE 'Socrates' being wise'. Thus, if (1) or (1a) were allowable expressions in English-P, the occurrence of 'not' in those contexts would be purely non-extensional. But they would not have to occur in English-P at all, for the negative sense can already be specified indirectly as it is in (7) and (7a). The important point, however, is that in English-P the extensional-non-extensional distinction would not merely be a conceptual one as it is in English and other natural languages. In order for 'not' in its extensional role--and indeed for any logical connective--to occur in a sentence of English-P, we must first already have a sentence (or sentences in the case of two or more arguments) such as (5) which expresses a positive sense specified directly by means of the indicator 'pos'.

As we saw briefly in chapter three, 'pos' could also have an extensional use in English-P, though there would seem to be little point in it. It could stand for the other significant truth-function of one argument, viz. that function wherein every sense or thought is a truth-function of itself, and for which there is no sign either in natural languages or any standard logical notation except Frege's, where <u>mutatis mutandis</u> the horizontal stroke plays this role.

Even here, it would be an easy job to separate out extensional and non-extensional uses of 'pos'. In the sentences:

8. Socrates is pos pos wise, and

8a. It is pos the case that Socrates is pos wise,

only the right-hand occurrences of 'pos' would be considered non-extensional. For the sense of (8) or (8a) is derived from (5) in the same manner as (7) and (7a) were--the difference being that the same (positive) sense is expressed once again. Like any other logical connective, the extensional 'pos' can attach only to a sentence, and the basic sentential units of English-P are sentences like (5) which must contain the non-extensional 'pos'. And as a connective, the extensional 'pos' can of course attach to more complex units. Thus, the sentence:

9. It is pos the case that Socrates is not pos wise

is a value-expression derived from (7) as argument-expression, and expresses the same (negative) sense as (7). The unsaturated expression 'it is pos the case that...' would function as a genuine logical connective corresponding to Frege's '———ξ' just as 'it is not the case that...' corresponds to '—¬—ξ' or '¬()'.

Operations

'Not', in its extensional role, expresses that Wittgenstein calls an operation--as do all the logical connectives. This notion is an extremely important one and is discussed in the Tractatus from 5.2 through 5.442. It is not feasible

to reproduce all of these entries here; instead, the reader is asked to familiarize himself with them, and I shall cite or quote them in the ensuing discussion as the need arises. As a first step toward gaining an understanding of what operations are, we must get clear on the terminology Wittgenstein employs in this passage.

So far, we have used 'function' in the very broad sense of an entity which takes values for arguments and is signified by an unsaturated expression--one requiring "completion". In this sense, operations are a special case of functions. We shall eventually consider the question of what marks them off from other functions, but for the present we must note that in the passage under discussion, Wittgenstein largely uses 'function' in a much narrower sense than we have been using it. In his usage, an operation is not a function, rather a function is something to be contrasted with an operation. More specifically, a function in his sense is any function in our broad sense which is not an operation. This can be seen clearly at 5.25 where he warns us not to confuse functions and operations, and yet the entire discussion makes it quite plain that the kind of operations with which he is most concerned are what would usually be termed "truth-functions" (negation, conjunction, etc.). Wittgenstein does use the term 'truth-function' in two closely related ways--neither of them, however, meaning what we would call a function. First, '$p \& q$' is said to be a

truth-function of p and q, and for Wittgenstein this means that the sentence or proposition ⌜p&q⌝ is (in our terminology) a value-expression derived from p and q as argument-expressions. In his terminology, it is also described as the result of an operation (viz. the "truth-operation" of conjunction) being applied to p and q as bases. Second, he speaks of signs like 'p&q' as a truth-function, not as a truth-function of something else. In this usage, 'truth-function' means simply 'result of a truth-operation'. By a "truth-function" then, Wittgenstein does not mean a truth-function in the usual sense; rather he means the value-expression produced by combining the logical connective with the bases. What is usually conceived as the function itself is what Wittgenstein calls the "operation"--in the case of truth-functions, "truth-operations". However, to complicate matters further, we find him saying at 5.2341: "The sense of a truth-function of p is a function of the sense of p." Here, he uses the term ' is a function of' to express the operation or function in the broad sense. To take an example, he is saying that the sense of ⌜∿p⌝ (a "truth-function of p") is a function of the sense of p. Which function? The truth-operation of negation. That is, the sign '∿p' is the result of the operation of negation being applied to p as the base. And as if all this were not confusing enough, Wittgenstein sometimes uses 'operation' to mean the sign (e.g. 'not' or '∿') but more often to mean that which is

signified (the operation or what we have been calling the truth-function). Finally, at 5.44 where he says: "Truth-functions are not material functions...", the ensuing discussion indicates that by 'truth-function' he means what before he had called "truth-operations" (in other words, "truth-function" in our sense), and by 'material function' what he had previously called simply "functions".

In view of all this, I shall adopt the following terminology. First, I will use 'function' in the broad, generic sense just as before, and hence operations will be a special case of functions. Those functions which are not operations will be termed "material functions". I will use the terms 'truth-operation' and 'truth-function' to mean only functions like negation, conjunction, etc., not expressions like '$p \& q$' and '$\sim p$'. I will use both of these terms in the absolute sense (as in 'negation is a truth-operation'), but only 'truth-function' will be used in the relative sense ("is a truth-function of..."). The signs '$\&$', '$\sim$', 'and', 'not', etc. will be termed "logical connectives" or "signs for operations". Expressions like 'ξ and ζ', '$\sim\xi$', 'not ξ', etc. will be called "functional expressions" or "operation-expressions". Sentences and their schematic counterparts like '$p \& q$', 'p and q', '$\sim p$', etc. will be termed "results" or "bases" of operations as the context requires. The sense of a sentence will continue to be termed a "thought".

Now let us try to determine what operations are and how

they are distinguished from other functions. About all we know so far is that truth-functions are a typical example of them. A complete, technical explication of the difference between operations and material functions lies beyond the scope of this essay, and hence I shall here attempt only to mark the contrast in an informal way. Of the various distinctive features Wittgenstein attributes to operations, some require for their clarification an understanding of other, equally problematic passages of the Tractatus. Hence, I shall pass over these features and concentrate on the more readily intelligible ones.[2]

Let us begin by considering some functions other than truth-functions which I believe would also be regarded by Wittgenstein as operations. The most familiar of these are what are often termed "arithmetic operations", examples of which would be '$\sqrt{\xi}$' and '$\xi+\zeta$'. We shall also consider an example employed by Anscombe,[3] viz., the notion of conversion as it applies to first-order dyadic relations. That is, if

---

[2]The less obvious features are:
a. an operation expresses an internal relation between the structures of its result and of its base(s) (5.2-5.22),
b. an operation manifests itself in a variable (5.24), and
c. operations do not indicate a form, but only a difference between forms (5.241).

[3]Introduction, chapter 8, page 117. Many of the conclusions to be developed here were inspired by Miss Anscombe's comments in this chapter. However, it should become clear to anyone familiar with her remarks that there are several important areas of disagreement between her and myself.

we consider a singular, relational sentence of the form:

$$\underline{a} \text{ is husband of } \underline{b}$$

or in logical notation:

$$H(\underline{a},\underline{b})$$

we can then introduce a functional sign, say 'Cnv', which, when attached to the relational expression, expresses the converse of that relation--in the example cited, "is wife of". Thus we would have:

$$\underline{a} \text{ is Cnv:husband of } \underline{b}$$

or in logical notation:[4]

$$\text{Cnv:}H(\underline{a},\underline{b}).$$

These expressions will then express the same sense as is expressed by:

$$\underline{a} \text{ is wife of } \underline{b}$$

and

$$W(\underline{a},\underline{b}).$$

The operator 'Cnv' will express an operation in much the same way as the arithmetic schemata mentioned earlier. Finally, let us return to the coin example employed in

---

[4]In effect, conversion is here being introduced as what Frege would term a "second-level" concept expressed in logical notation by the unsaturated, functional expression:

$$\text{Cnv:}\_\_(\underline{a},\underline{b})$$

and the gap indicates the place where a first-level relational expression may be inserted. Inserting '$H(\xi,\zeta)$' gives us:

$$\text{Cnv:}H(\underline{a},\underline{b})$$

chapter three. There, it will be recalled, we used the sign '#' in its extensional role to express the "tails bet" made with a coin, C, in terms of the "heads bet" based on the same coin. That is, instead of "directly" specifying the tails bet by means of the table:

I.

C	
H	L
T	W

we specified the heads bet directly and specified the tails bet "indirectly"; in other words, the latter bet was specified in terms of the conditions under which the heads bet won or lost instead of the coin's coming up heads or tails. If we express the heads bet again as '*C', then the tails bet is specified indirectly by means of the table:

II.

*C	#(*C)
W	L
L	W

'#ξ' is thus a functional sign meaning roughly: 'opposite bet from...'; that is, for any bet, B, #(B) is that bet which loses when B wins and wins when B loses.[5] The operator '#' expresses an operation from bets to bets just as '$\sqrt{\xi}$' expresses an operation from numbers to numbers, and 'Cnv: __(a,b)' from relations to relations. Let us now consider a further example. We shall introduce a bet based on two

---

[5] The essentials of the discussion concerning extensional and non-extensional uses of 'not' and direct vs. indirect specifications of a negative sense back in chapter three should be kept in mind here.

coins, C and D, and which is directly specified by the table:

III.

C	D	
H	H	W
H	T	L
T	H	L
T	T	L

We shall hereafter refer to this bet as the "double-heads bet" inasmuch as it wins just in case both coins come up heads. Now just as with the tails bet based on C, we can introduce an operation by means of which the double-heads bet based on C and D may be specified indirectly, i.e. in terms of the heads bet on C and the heads bet on D. We shall use the sign '$' for this purpose, and it will express that bet-function of two arguments such that the bet (*C)$(*D) has the value, win, when the bets *C and *D both have the value, win, and the value, lose, in all other cases. Hence, the table by which the double-heads bet is specified indirectly is:

IV.

*C	*D	(*C)$(*D)
W	W	W
W	L	L
L	W	L
L	L	L

Thus, 'ξ$ζ' will be an operation-expression in the same way as '#ξ'. The functions expressed by these unsaturated expressions as well as by 'Cnv: __(a,b)', '$\sqrt{\xi}$', and 'ξ+ζ' are all, I believe, examples of what Wittgenstein calls an operation, and I shall now try to explain why.

First of all, then, let us consider the arithmetic operations. They are functions from numbers to numbers, and

the expressions representing them may be reiterated any number of times. Thus, '$\sqrt{4}$' designates 2 and '3+2' designates 5. Moreover, 2 may also be designated by '$\sqrt{\sqrt{16}}$' and 5 by '1+(2+2)'. The reiterative character is distinctive of all operations (5.251), and in this respect, operation-expressions like '$\sqrt{\xi}$' and '$\xi+\zeta$' differ sharply from what are often called "arithmetic predicates"--examples of which would be '$\xi^2 = 1$' and '$\xi > 5$'. Such predicates represent in the realm of arithmetic functions which are not operations--they are examples of what Wittgenstein calls "material function" at 5.44. The value-expressions '$3^2 = 1$' and '$7 > 5$', unlike '$\sqrt{4}$' and '3+2', do not designate numbers, and indeed, if we adhere to the conclusions established in chapter two, do not designate at all. Rather, they express thoughts (senses which are true or false), and hence the respective functions involved are not functions from numbers to <u>numbers</u>. On the other hand, the sense of '3+2' consists, roughly speaking, in exhibiting the number 5 as a certain function of the ordered pair <3,2> (specifically, as the sum of 3 and 2), and such senses clearly are not bearers of truth-value. Moreover, expressions like '$3^2 = 1$' and '$7 > 5$' cannot themselves serve as argument-expressions (bases) for '$\xi^2 = 1$' and '$\xi > 5$'. The expressions '$(3^2 = 1)^2 = 1$'

and '(7>5)>5' are obvious nonsense[6], and hence the functions $\xi^2 = 1$ and $\xi > 5$ do not admit of reiteration. In an arithmetic operation, however, we can move from a designation of a number (the base) to a compound designation of a number (the result) which can in turn serve as the base for constructing yet another compound designation.

This should serve to give us at least a preliminary idea of the difference between an operation and a material function. Let us now turn to the more important examples of conversion and coin-bets. In the former, we have a function from thoughts to thoughts, in the latter, from bets to bets (or win-values to win-values). Further, the functional expressions all admit of reiteration; from 'Cnv:H(a,b)' as base, we may derive the result 'Cnv:Cnv:H (a,b)' which expresses the same sense (thought) as 'H(a,b)', and from '#(*C)' and '(*C)$(*D)', we may derive respectively '#(#(*C))' and, say, '(*C)$((*C)$(*D))'. Similarly, truth-functions are functions from thoughts (or truth-values) to thoughts (or truth-values), and operation-expressions like '∿ξ' and 'ξεζ' obviously can be reiterated. But all of this contrasts sharply with what Frege calls "concepts" and

---

[6]As we saw in chapter one, such expressions are allowed in Frege's formal system, and because they make no sense in ordinary mathematical discourse, are arbitrarily assigned references. The rationale, as we also saw, stems mainly from Frege's rather stringent demand for "rigor". It should already be apparent that this demand contributes to obliterating the sharp distinction Wittgenstein wishes to draw between operations and material functions.

"relations"[7], and the corresponding concept- and relation-expressions such as 'ξ is wise' and 'ξ is to the right of ζ'. Whatever the correct analysis of the functions represented by such expressions (whether they are references or senses, and what exactly their values and arguments are), two things are clear. First, unlike the operation-expressions just cited, their argument- and value-expressions are of logically, different categories. In an operation-expression, the result and base(s) are always the same kind of expression--sentences to sentences, bet-signs to bet-signs, etc. But with an expression like 'ξ is wise', the argument-expression must be a singular term and the value-expression produced by completing 'ξ is wise' with, say, 'Socrates' is a sentence--a complex expression of a markedly different order from its components inasmuch as it expresses a sense which is true or false, can be asserted, has a mood, and so forth. And it is obvious that the value-expression 'Socrates is wise' cannot in turn serve as an argument-expression, for we would then derive the nonsensical

---

[7]It may be recalled from chapter one that because of the demand for "rigor" (a function must be defined in terms of the values it takes for all "possible" arguments), Frege defines truth-functions as a special case of concepts and relations. I shall try eventually to bring out the philosophical objections to his doing so; for now, however, let it be noted that from here on, I shall use 'concept' and 'relation' to cover only those functions represented by unsaturated expressions of the form 'ξ is Ø' and 'ξRζ', etc. That is, by expressions which have often been termed "propositional functions" or, in a highly technical use, "predicates".

expression: 'Socrates is wise is wise'. That is, concept and relation expressions do not admit of reiteration, and such expressions would then represent material functions, not operations.

So far, though, we have only scratched the surface. The most important distinguishing feature of operations is described by Wittgenstein at 5.25:

> The occurrence of an operation does not characterize the sense of a proposition.
> Indeed, no statement is made by an operation, but only by its result, and this depends on the bases of the operation. (Operation and function must not be confused with one another).

Though Wittgenstein seems to be talking here about operations generally, his reference to propositions in the first sentence indicates that he chiefly has in mind operations by which expressions whose senses have a truth-value are produced from other such senses--more specifically, he has in mind truth-functions or "truth-operations". And I believe the force of the second remark in this passage is that although completing an expression for a material function, e.g. 'ξ is wise', by a singular term gives us a new kind of expression--one which expresses a thought or "makes a statement"[8]--no such thing happens in the case of an operation-

[8]The translation here is not very literal; I think a more accurate rendering of: "Die Operation sagt ja nichts aus" might be: "The operation does not say anything"--thus meaning in the whole context that an operation merely enables us to move from expressions which already "say something" (sentences) to other such expressions. The operation thus does not contribute the truth-claiming feature of the result.

expression. Rather, if we begin with sentences as bases, we will wind up with sentences as results. Hence, in an operation, sentences (and their senses) are not "formed" or "generated" from other kinds of expressions (and their senses); instead, the operation merely enables us to express the sense of one sentence (the result) as a function of the sense(s) of another (others). In a truth-operation, we only move from expressions which already say something to further such expressions--we do not <u>generate</u> such expressions from expressions of other logical categories. Thus, the "truth-claiming" feature of the result of a truth-operation is not supplied by the operation, it must already have been present in the base or bases.

Now let us turn to the first remark in 5.25. Wittgenstein's thought here is not very easily expressed, but it is very important and we find him reaffirming it with special reference to negation at 5.44 and 4.0621. We can perhaps understand it better if we begin with the coin example. Earlier, we introduced the connective '$' to express the double-heads bet <u>indirectly</u>, i.e. in terms of two heads bets rather than the two coins themselves. Now the following table directly specifies yet another ("kind of") bet based on two coins, C and D:

V.

C	D	
H	H	L
H	T	W
T	H	L
T	T	L

Thus, this bet (hereafter, the "heads-then-tails" bet) wins only when coin C comes up heads and coin D comes up tails. Now, as with the double-heads bet, we can express this new bet indirectly in terms of the heads bet on C and the heads bet on D, and we might introduce the sign '¢' as the connective in terms of which the heads-then-tails bet can be specified indirectly. Thus, we would have:

VI.

*C	*D	(*C)¢(*D)
W	W	L
W	L	W
L	W	L
L	L	L

Here, the heads-then-tails bet is specified indirectly as that bet which has the value, win, if and only if the heads bet on C has the value, win, and the heads bet on D has the value, lose. But of course we could have dispensed with the '¢' and indirectly specified the heads-then-tails bet in terms of the connectives already at our disposal. Instead of expressing this bet by the sign '(*C)¢(*D)', we can express it by '(*C)$(#(*D))'--as that bet which wins if and only if the heads bet on C wins and the <u>tails</u> bet on D (here also specified indirectly) wins. This can be seen by the following table:

*C	*D	#(*D)	(*C)$(*D)	(*C)$(#(*D))
W	W	L	W	L
W	L	W	L	W
L	W	L	L	L
L	L	W	L	L

And of course, if we had chosen, we could have retained the '¢', dispensed with '$', and indirectly expressed the

double-heads bet solely in terms of '¢' and '#'. The table will not be given here, but it can easily be verified that the double-heads bet is expressible as '(*C)¢(#(*D))'.

Now consider the thought specified directly in the table:

p'	q'	
P	P	T
P	N	F
N	P	F
N	N	F

Here, 'p'' and 'q'' represent CAE's, not sentences, and the specified thought will be true just in case both specified states of affairs obtain. Accordingly, thoughts of this kind will be termed "conjunctive" thoughts, and they are clearly the analogue of the double-heads kind of bets. And just as '$' is used to express that kind of bet indirectly, so the familiar logical connective '&' (or ':'), as well as 'and' in one of its most important uses in English, expresses such a thought indirectly. That is, instead of specifying the thought in terms of whether the specified states of affairs obtain or not, we may specify it indirectly as a function of the positive thoughts based on p' and q'--just as the double-heads bet was expressed indirectly as a function of the heads bets on coins C and D. If we now use 'p' and 'q' to represent sentences expressing the positive thoughts in question, the table would be:

IX.

p	q	p&q
T	T	T
T	F	F
F	T	F
F	F	F

Here the thought in question is expressed as a (truth-) function of the positive thoughts expressed by '$\underline{p}$' and '$\underline{q}$'. And just as the "bet-connectives" '#' and '$' may be used to express (indirectly) other bets based on two coins, so '∿' and '&' can be used to indirectly express other thoughts based on two CAE's. Thus, instead of introducing a further connective for indirectly specifying the thought directly specified by the table:

X.

p'	q'	
P	P	T
P	N	T
N	P	T
N	N	F

we may indirectly specify this thought solely in terms of '∿' and '&'; specifically, as '∿(∿$\underline{p}$&∿$\underline{q}$)'. However, if we do introduce a special connective, as is commonly done in logic texts with the 'v', clearly we can dispense with '&' and re-express the conjunctive thought as '∿(∿$\underline{p}$v∿$\underline{q}$)'. The tables by which this can be verified will be found in any elementary logic text, and hence will not be reproduced here.

The signs '#', '$', and '¢', as well as '∿', '&', and 'v', all express operations; in the former case "bet- (or win-) operations", in the latter "truth-operations". The interdefinability of signs for operations is, as Wittgenstein notes at 5.42, one of their most important features, and I think it will help us to understand the first remark in 5.25. The double-heads bet, as we saw, may be expressed by either '(*C)$(*D)' or '(*C)¢(#(*D))'; the same bet is expressed

(specified indirectly) by both. What, then, is the difference? Well, in the former case the double-heads bet is expressed as a certain function of two arguments--of the heads bets, *C and *D. In the latter, it is expressed as a different function of two arguments--this time of the heads bet, *C, and the tails bet, #(*D). In the one case, it is expressed by the result of performing the operation represented by '$', and defined in table (IV), on '*C' and '*D', and in the other, it is expressed by the result of performing the operation represented by '¢', and defined by table (VI), on '*C' and the result of performing the operation represented by '#' on '*D'. By the same token, the same thought is expressed by both '$p \vee q$' and '$\sim(\sim p \& \sim q)$'; in the former case, it is expressed by the result of performing the operation commonly called "disjunction" on 'p' and 'q' as bases, and in the latter case it is expressed by the result of performing the operation called "negation" on the result of performing the operation called "conjunction" on the results of performing the operation of negation on 'p' and 'q' respectively. But again, it is the same thought. Which thought? The thought directly specified in table (X), and which might well be termed a "disjunctive" thought. But if so, it must be kept in mind that the connective 'v' is not in any way essential to expressing such a thought--the occurrence of the operation commonly called "disjunction" is not a distinguishing mark of the thought expressed by the complex

sign in which it occurs. For example, the thought expressed by '∿(∿p∨∿q)' is what was earlier termed a "conjunctive" thought specified directly in table (VIII). Indeed, no logical connective ever serves to mark the logical kind of thought expressed by the truth-functional compound in which it occurs; rather, whatever kind of thought it may be, the operation-sign merely serves to express it as a function one or more other thoughts. Now we could, of course, have introduced the expression 'p'vq'' in table (X) for the purpose of indicating the logical kind of thought directly specified in that table. However, here the sign 'v' would not behave as an operation-sign (logical connective), but rather as an indicator of one of the fourteen possible ways in which two states of affairs (here expressed by the CAE's, 'p'' and 'q'') may jointly be projected or logically conceived. That is, 'v' would here have a non-extensional use, and we could, as we did for 'not' or '∿' in chapter three, distinguish between its role as an indicator of a possible projection of two states of affairs and its extensional role as signifying a certain operation or truth-function of two arguments.

This is, I believe, what Wittgenstein is driving at in 5.25 when he says, "The occurrence of an operation does not characterize the sense of a proposition." He means that the operation-sign does not serve to identify the logical kind of thought expressed by the complex sign which is the result of combining the operation-sign with the bases. Instead,

the thought (whatever it is, positive, negative, conjunctive, etc.) is specified merely as a function of one or more other thoughts. An operation only serves to specify indirectly one thought in terms of one or more others, it does not identify the thought in terms of its inherent, logical features. Thus, at 4.0621 we find: "That **negation** occurs in a proposition is no characteristic of its sense ($\sim\sim p = p$)."[9] This is clear from the fact that '$\sim\sim p$' expresses a positive sense, and hence '$\sim$' or 'not' in their purely extensional roles as logical connectives do not serve to characterize the kind of thought that is expressed by a complex sign in which they occur. And this is to say that '$\sim$' (or 'not') in that role does not serve to pick out or identify a special class of thoughts, viz. negative thoughts. Rather, this job, as we saw back in chapter three, is performed by '$\sim$' in its non-extensional role--as an indicator of one of the two possible projections of a single state of affairs. The same considerations apply to all the logical connectives, and clearly to the "bet-connectives" as well. '#' in its extensional role (as a bet-connective) has the force of 'opposite bet from...', as can be seen from table (II). But there is no special class of bets called "bets which are

---

[9]This is the Ogden translation; the Pears-McGuinness translation is not accurate, for it gives the impression that Wittgenstein is saying that '$\sim$' contributes at least something to the sense of a proposition. But the German text does not contain this suggestion; rather, it flatly states that '$\sim$' contributes nothing at all.

opposites", for every bet has an opposite. There are, however, two possible kinds of bets based on one coin: heads bets and tails bets; and in chapter three, '#' was also used as an indicator of the latter. Here then, it played a non-extensional role; it directly specified the tails bet instead of expressing it merely as a function of (specifically, "opposite bet from") the heads bet. In like manner, '~' in its non-extensional role directly specifies a negative sense or thought, but as a logical connective, it can only express such a sense as a function of (contradictory of) a positive sense; and that Wittgenstein is right at 4.0621 is easily seen from the fact that there is no special class of thoughts called "thoughts which are contradictories". Every thought has a contradictory; but for all that, the independent criterion shows us that there is a special class of thoughts which may justly be termed "negative thoughts". Thus, in the non-extensional role, '~' or 'not' certainly does contribute to the sense of the larger expression in which it occurs. Just how it does was discussed back in chapter three.

Moreover, it is clear that in:

Cnv:H(a,b)

a certain dyadic relation, wife of, is specified indirectly; that is, as a function of the relation husband of. And it is also clear that the occurrence of the operation-sign 'Cnv' is not a distinguishing mark of the sense of the entire expression, nor that 'Cnv' is in any way essential for

expressing that sense. For the sense may be specified (expressed) directly by:

$$W(\underline{a},\underline{b}).$$

And it is obvious that 'Cnv' does not serve to single out a special class of relations, viz. "relations which are converses". Every relation has a converse. Likewise, in '3+2', the operation-sign '+' is not essential for designating that number which is the reference of the entire expression. The occurrence of '+' in the whole only serves to exhibit the number as a certain function of <3,2>, and therefore the number is specified indirectly. We may obviously specify it directly by the simple numeral '5'. That we have an indirect specification is apparent from the fact that '+' does not pick out a class of numbers called "numbers which are sums". The arithmetic operations differ from the others only in the extrinsic respect of being functions from references (numbers) to references rather than from senses (or their win- or truth-values) to senses.

All of this stands in marked contrast to "material functions". We have seen that '$\sim$' in its extensional role as an operation-sign is not a distinguishing mark of the sense of the whole expression in which it may occur. This is certainly not the case with 'wise', and thus there is a sharp contrast to be drawn between functional expressions like '$\sim\xi$' (or 'not-ξ') and 'ξ is wise'. Functional expressions of this latter variety are clearly distinguishing

marks of the senses of the value-expressions produced by completing them with a singular term. A singular, uncompounded sentence--an expression which, unlike its components, may be asserted--must consist of an unsaturated term, and inasmuch as sentences are what paradigmatically express those senses called thoughts, the sense of the unsaturated expression certainly contributes directly to the thought expressed by the resulting sentence. Thus, functional expressions like 'ξ is wise', when completed with a singular term, express a sense of a unique sort; that is, functional expressions of this kind serve to pick out a special class of senses, viz. thoughts. But an operation from thoughts to thoughts, like that expressed by '∿ξ', does not, as we have already seen, serve to pick out a special class of thoughts, viz. negative thoughts. We shall now pursue this contrast a little further.

We are here drawing a very sharp distinction (and one which Frege ignores) between combining '∿ξ' and 'p' to get '∿p' (or alternatively, inserting 'not' into 'Socrates is wise' to get 'Socrates is not wise') and completing 'ξ is wise' with 'Socrates' to get 'Socrates is wise'. In the latter case, the sense of the value-expression, as we have seen, possesses important logical features not found in the senses of the component expressions, but it must nonetheless be specified in terms of the senses of these components. Indeed, it is natural and appropriate to speak of the senses of 'Socrates' and 'ξ is wise' as themselves "components" or

"parts" of the sense of 'Socrates is wise'. In fact, it would seem that the only sense that can be attached to these metaphors is the following: the sense of 'X' is a component or part of the sense of 'Y' if and only if specifying the sense of 'X' is essential for (a necessary condition for) specifying the sense of 'Y'. For this reason, it would be quite absurd to speak of 'Socrates is wise' as an indirect specification of its sense; rather, the direct vs. indirect dichotomy does not apply here at all. It is of course correct to say that the sense of 'Socrates is wise' is a function of the sense of 'Socrates', just as the sense of $\ulcorner \sim p \urcorner$ is a function of the sense of p, but surely no one would wish to maintain that the role of 'ξ is wise' is simply to specify indirectly the former sense in terms of the latter--as if there were yet a further means by which it could be expressed "directly", i.e. without making reference to the latter sense.

Herein lies the underlying difference between operations and material functions. Operations are essentially means by which a certain entity (a thought, a bet, a number, etc.) may be indirectly specified in terms of one or more other such entities. The immediate consequence of this is of great importance philosophically, but difficult to articulate. Whereas it is intelligible to speak of the sense of 'Socrates is wise' as a "compound" of the senses of its components, it is quite wrong to suppose, as Frege and Russell did, that

because the sign '∿p' is a compound of '∿ξ' and 'p', that the sense or thought expressed by the whole is for that reason a "compound sense".[10] The same holds of course for all the logical connectives, and for the English words 'not', 'and', and 'or' in their purely extensional aspect. The sentence ⌜∿p⌝ does not express a sense which is a compound of the senses of '∿ξ' and p; rather, it merely expresses a negative sense, but as a function of the corresponding positive sense. As we saw in chapter three, a negative sense is one of the two possible, complementary projections of a single state of affairs, but for that very reason it does not contain the positive sense as a "part". The role of '∿', then, qua logical connective, is to express one sense as bearing a certain functional relation to another sense--it does not "combine" with that latter sense to "produce" the former one. This would be like supposing that '$\sqrt{4}$' must designate a "compound number" or that 'Cnv:H(a,b)' must express a "complex relation" or even that '#(*C)' must express a "compound bet". The force of '$\sqrt{\ }$' is simply to exhibit the

---

[10] That Frege held this view should be apparent from the discussion in chapter one, but if further verification is desired, see for example his "Compound Thoughts", translated by R. Stoothoff in Mind, 1963. Russell, so far as he is intelligible on these matters, seems committed to it in section A of part I of Principia Mathematica, as well as chapter III of "The Philosophy of Logical Atomism", reprinted in Logic and Knowledge, R. Marsh, ed. For a further comparison of their views with Wittgenstein's, see chapter eight of Anscombe's Introduction, pps. 118-120. We shall have more to say on Frege's views later.

number 2 (which cannot sensibly be said to be simple or compound) as a function of 4, and the sense of p no more occurs as a "component" of the sense of ⌜∿p⌝ than 4 occurs as a "component" of 2. By the same token, 'Cnv' in 'Cnv:H(a,b)' serves to express (indirectly) one relation, wife of, as a function of the relation husband of, but it is nonsense to suppose that the former relation is "compounded" out of the latter plus the sense of 'Cnv'. Presumably, we would then also have to maintain that in 'Cnv;Cnv:H(a,b)' the relation husband of occurs as a component of itself, and the same would hold for ⌜∿∿p⌝ . And it should be quite obvious by now that this point would still hold even if we were to regard, as Frege did, the relation as the reference of 'H(a,b)' instead of its sense.

However, the most fruitful analogy to pursue here is the betting case. '#(*C)', far from expressing a "compound bet", merely specifies indirectly the tails bet on C--indirectly because that bet is expressed as a certain function of the heads bet. The heads bet is thus not a "component" of the tails bet, but they nevertheless do share something in common. They are the two possible, significant bets that can be made with the same coin. Likewise, the positive sense expressed by 'p' is not a "component" of the negative sense expressed by ⌜∿p⌝, but these senses are the two possible, significant projections of the same state of affairs. And as we saw long ago, it always is possible to specify the

tails bet and the negative thought directly, i.e. without reference to the heads bet or the positive thought at all. And because of this, it is absurd to suppose that the tails bet is a compound bet and the negative thought a compound thought.

Moreover, '(*C)$(*D)' and ⌜p&q⌝ do not express respectively a compound bet or a compound thought--at least not in the sense Frege had in mind. Rather, the first of these expressions specifies indirectly one of the fourteen possible bets that can be made with two coins--specifically, the "double-heads" bet specified directly in table (IV). Equally, ⌜p&q⌝ specifies indirectly one of the fourteen possible projections (thoughts) with respect to the two states of affairs, p' and q'. The thought in question is the "conjunctive" thought directly specified in table (VIII). The result 'p&q' expresses the conjunctive thought here as a certain function of two arguments of the positive thoughts expressed by the bases 'p' and 'q', but it would be wrong to suppose that these latter two thoughts are "components" of the conjunctive thought. As our analysis of operations has shown, it does not follow from the fact that p and q occur as physical parts of ⌜p&q⌝ that their senses must occur as "components" of the sense of ⌜p&q⌝--no more than because '3+2' designates 5, that 3 and 2 are "components" of 5, or that because '(*C)$(*D)' contains two bet signs as parts, the bets they express are somehow parts of the bet expressed by the

whole expression. The only sense that can be made out of calling the sense of '(*C)$(*D)' a compound bet or the sense of ⌜p&q⌝ a compound thought is that in the one case we have a bet made with two coins and in the other a thought which is a projection of two states of affairs. But just as a bet is not composed of other bets, so a thought is not composed of other thoughts.

Miss Anscombe remarks that when we are introduced to the propositional calculus in introductory logic texts, we are usually told that propositions can be combined in certain ways to form further propositions, and the idea is thus created "that propositions combine to form further propositions, much as metals combine to form alloys."[11] And she continues, "In short, is there not an impression as it were of logical chemistry about (such) explanations?" There is indeed such an impression, and it is one which I believe Wittgenstein wished to counteract in his remarks on operations. However, as we shall see presently, these remarks

[11] Introduction, page 53.

have still more far-reaching consequences.[12]

Frege's Theory of Assertion Again

Now at last we can return to Frege's theory of assertion, and the foregoing remarks concerning operations will be of use in showing what is wrong with it. It will be recalled that in chapter two, three purposes behind the theory were delineated: first, to distinguish between assertive and non-assertive uses of sentences; second, to dissolve various logical puzzles which arise because of the failure to distinguish between assertion and predication; and third, to account for the fact that truth is not simply the counterpart of falsehood, but also constitutes "success" in assertive utterance. In chapter two, this third purpose was examined in some detail, and a *reductio* argument against Frege's theory was eventually produced, along with the further result of casting doubt on Frege's analysis of sentences and the

---

[12]The foregoing remarks should serve to illuminate the truth-condition principle; specifically, with respect to the question of why the principle cannot be applied in a non-circular fashion to the singular, uncompounded sentences of ordinary discourse. For a specification of the sense of a sentence in terms of this principle is essentially a specification in terms of what other *thoughts* must be true. Therefore, to specify the sense of a sentence in this manner is, according to the foregoing remarks, to specify it *indirectly*. And since logically simple sentences of the form '*a* is Ø' do not require for the truth of the thoughts they express the truth of any *further* thoughts, they cannot properly be said to admit of "truth-conditions" in this sense. Since such sentences do not contain a logical connective--an expression by which the thought expressed is specified as a function of one or more other thoughts--the truth-condition principle cannot be applied to them. The thought expressed is specified *directly*.

senses they express. Now a reductio argument, if successful, only shows us that a theory is wrong, not what the error is. Nonetheless, it often provides some clues, and in the aforementioned chapter, it was suggested that Frege, in distinguishing between a thought and a judgment concerning it (grasping a sense and "acknowledging its truth") had so to speak created a much wider gap between a thought and its truth-value than is warranted--that judging is not the kind of "advance" from a thought to a truth-value Frege conceived of, and that there may be an assertoric element in the sentence or thought that he has wrongly ascribed to the judging relation. These suggestions will be followed up toward the end of this chapter, but for the present we must return to the first two purposes for which Frege introduced the theory. For in analyzing them, we will, I believe, be able to see wherein his theory goes wrong.

The critical remarks I wish to make concerning Frege's theory are derived in part from my interpretation of the following comments by Wittgenstein. I will quote the passages here in full, and refer back to them by number. I shall also letter the paragraphs for future reference.

I. The assertion-sign is logically quite without significance. It only shows, in Frege and in Whitehead and Russell, that these authors hold the propositions so indicated to be true. '⊢', therefore, belongs as little to a proposition as (say) the number of a proposition. A proposition cannot possibly assert of itself that it is true. Assertion is merely psychological. There are only unasserted propositions. Judgment, question, and command all stand on the same level; but all have in common the propositional form, and that alone

interests us. What interests logic are only the unasserted propositions.[13]

II. a. Frege's idea that every assertion contains an assumption which is the thing that is asserted, really rests on the possibility found in our language of writing every statement in the form: "It is asserted that such-and-such is the case." But "that such-and-such is the case" is not a sentence in our language--so far it is not a move in the language-game. And if I write, not "It is asserted that..", but "It is asserted: such-and-such is the case", the words "It is asserted" simply become superfluous.
b. We might very well also write every statement in the form of a question followed by a "Yes"; for instance: "Is it raining? Yes!" Would this show that every statement contained a question?
c. Of course we have the right to use an assertion-sign in contrast with a question-mark, for example, or if we want to distinguish an assertion from a fiction or a supposition. It is only a mistake if one thinks that the assertion consists of two actions, entertaining and asserting (assigning a truth-value or something of the kind), and that in performing these actions we follow the propositional sign roughly as we sing from the musical score. Reading the written sentence loud or soft is indeed comparable with singing from a musical score, but "meaning" (thinking) the sentence that is read is not.
d. Frege's assertion sign marks the beginning of the sentence. Thus its function is like that of a full-stop. It distinguishes the whole period from a clause within the period. If I hear someone say "it is raining" but do not know whether I have heard the beginning and end of the period, so far this sentence does not serve to tell me anything.[14]

---

[13] "Notes on Logic", in Notebooks, page 96. A very similar remark occurs at 4.42 in the Tractatus.

[14] Philosophical Investigations, paragraph 22. The italics are Wittgenstein's. For the word 'assumption', one should read simply 'thought' or 'sense of a sentence'. His use of 'assumption', as Anscombe argues, derives from his reading Russell's notoriously inaccurate account of Frege's logical writings in the Principles of Mathematics, (London: George Allen and Unwin Co., 1903) Appendix A. See Anscombe's Introduction, pps. 105-6.

Now the two purposes mentioned above have a common starting point, viz. the distinction between grasping a thought and judging it. If this distinction is not made, then as Frege rightly argues, we shall be drawn into all kinds of philosophical muddles. It is thus a necessary distinction for logical theory, but one which, as I hope to show, Frege fails to draw correctly. Let us now turn to the second of the two purposes and examine it more carefully. We shall return to the first purpose later.

It will be recalled that the second purpose was to resolve various logical puzzles which arise from the confusion between assertion and predication. If we fail to adhere to the distinction between grasping and judging a thought, we may be led, in the manner of the traditional, pseudo-Aristotelian logic texts,[15] to the view that in the "proposition", the predicate is "asserted with respect to the subject." If this were the case, then as we saw in chapter one, all valid arguments would be petitio's. For example, in the disjunctive syllogism, since the conclusion occurs in one of the premisses, it would follow that it has already been "asserted" in that premiss. And since no one would wish to deny that the sentence expresses the same thought in both occurrences, the argument would thus be circular. It is for this reason that Frege introduces the assertion-sign

---

[15] See, for example, H.W.B. Joseph, An Introduction to Logic (Oxford: Oxford University Press, 1906), chapter seven.

'⊢'.[16] When the argument is properly reformulated as:

P1⊢(Either P or Q)

P2⊢Not-P

C ⊢Q

it is made clear, via the scope of the '⊢', that it is the thought expressed by Pa as a whole that carries assertoric force. It is the "compound thought" expressed by ⌜Either P or Q⌝ that is properly the content of judgment. The sign ⌜Either ⊢P or ⊢Q⌝ is not just an incorrect expression of the first premiss of disjunctive syllogism, it is plain nonsense inasmuch as one cannot possibly assert Q in the course of saying "either-or". Thus, while we must grasp the thought expressed by Q in order to grasp the thought expressed by ⌜Either P or Q⌝, only the latter is so to speak "in a position" to be judged, and this fact is made clear by the '⊢'. On the other hand, Q in its occurrence as conclusion of the argument, does not merely express its thought--as it does in P1--rather, that thought may be a content of judgment as well. The distinction between the two kinds of occurrences of Q in the above argument should be reasonably clear, but it is not an easy one to express in words. Frege introduces no technical expression to mark it, but Russell, who closely

---

[16] In chapter one, it was pointed out that '⊢' is actually a compound of the "horizontal" stroke '—' and the vertical, assertoric stroke '|'. However, nothing is really lost here in speaking of '⊢' as if it were a single, logical unit.

follows Frege on these matters, expresses it as the difference between an "asserted" (in C) and "unasserted" (in P1) occurrence of a sentence or proposition.[17] Thus, the sign '⊢' is here introduced for the logical purpose of marking an "asserted" occurrence of a sentence--the sort of occurrence wherein the thought expressed may be judged as well as merely grasped or understood. It marks, that is, whatever further role a sentence has in the context over and above its expressing a sense (and for Frege at least, designating a truth-value). But it does not signify anything that gets added to the sense in such a way as to produce a new sense, for that would mean that Q in the conclusion does not have exactly the same sense it has in P1--thus vitiating the whole enterprise.[18]

---

[17] See section 38 of Principles of Mathematics, as well as Principia Mathematica, with A.N. Whitehead, (Cambridge: Cambridge University Press, 1913), page 8 and pages 92-93. It should be noted that these passages are filled with what would now be looked on as formal and theoretical errors which may tend to impede the reader's understanding of the distinction Russell wishes to draw. It should also be noted that some commentators, most notably Miss Anscombe, have interpreted the Frege-Russell view as follows: "Being asserted (in this "logical sense") is, for both Russell and Frege, something that cannot possibly attach to a proposition unless it is true." This is quite wrong, as the discussion in chapter one should make clear--wrong at least as far as Frege is concerned. It can, however, be fairly easily read into some of Russell's rather obscure remarks. See pages 114-117 of Anscombe's Introduction.

[18] Hence Frege's later doctrine that '⊢' is in no way equivalent to the predicate 'is true', and correspondingly that Q expresses the very same sense as 'The thought that Q is true'. Both expresses a content of judgment, while '⊢' pertains to the act of judging the content.

Frege's reasoning here seems very plausible, but in (I) Wittgenstein claims that '⊢' is "logically quite without significance". Why? Part of the reason can be found in (IIc) and (IId), but the essential reason, I think, is contained in his notion of an operation as it has just been explicated. We have seen that the truth-operations expressed by signs like '&', '∿', etc. are fundamentally means by which one thought may be expressed indirectly in terms of one or more others, and that such a sign is not therefore a distinguishing mark of the thought expressed by the entire expression in which it occurs. And as we have seen, the consequence of this is that in, say ⌜P&Q⌝ the thoughts expressed by the component sentences are not themselves "components" of the thought expressed by the entire expression. However, it ought to be quite clear by now that Frege's argument turns on the supposition that they <u>are</u> components--that they are essential for the specification of the thought expressed by the entire expression. For the distinction between asserted and unasserted occurrences of Q in the above example was based on the notion that the sense or thought expressed by Q may in the one case be judged whereas in the other it cannot, and the reason why it cannot, according to Frege, is that it occurs as part of a larger whole--a "compound thought" whose specification essentially depends on specifying those thoughts of which it is "composed". But we have seen that a specification in terms of <u>other thoughts</u> is a purely indirect one;

hence, for the reasons given earlier, the signs used in such a specification, inasmuch as they are not marks of the logical kind of thought expressed by the whole (do not tell us what its intrinsic logical features are), are not essential for expressing the thought in question. Rather, it is by means of the direct specification that the logical character of the thought expressed is made clear, and such a specification does not make reference to any further thoughts. Therefore, while it is true that Q expresses the same thought in both its occurrences, and while we may still distinguish between occurrences wherein Q may be asserted (the thought judged) and those where it cannot, the explanation Frege offers for this is quite wrong. It is not because the thought occurs as part of a larger whole in the unasserted case, it is because that thought does not "occur" at all. The occurrence of the sentence Q as a physical part of ⌜P&Q⌝ does not signify the occurrence of its sense as a "part" or "component" of the sense of ⌜P&Q⌝. Instead, the sense or thought expressed in P1 is merely being specified indirectly, i.e. in terms of or as a certain function of two other senses. The sole reason, then, for the occurrence of P and Q in P1 is to express those senses of which the sense expressed in P1 is a function (viz. a "conjunction"), just as the sole reason for '4' in '$\sqrt{4}$' is to indicate that number of which the number 2 is a square root. To know that something is the value of a given function, one must obviously know what the

corresponding arguments are. And again, it in no way follows from this that the senses of P and of Q occur "in" or as "components" of the sense of ⌜P&Q⌝. All of this is decidedly unlike the occurrence of 'Socrates' in 'Socrates is wise' (though Frege wrongly tries to subsume this case and the conjunction case under the heading of "concepts and relations"), for the latter as we have seen is not an indirect specification of its sense at all.

The foregoing considerations, I believe, show in part why Frege's '⊢' is "logically quite without significance". I say "in part" because Frege had other reasons for introducing that sign, and we shall discuss them later. But it should now be clear that while there may be a point in introducing '⊢' to distinguish among occurrences of a sentence, there can be no point for introducing it to distinguish among "occurrences" of the thought expressed. We may wish to distinguish between cases where the sentence Q may be asserted and those where it may not owing to its occurrence in a larger context, but since the thought expressed by Q does not occur in or as a component of ⌜P&Q⌝, there can be no point in using '⊢' to mark off those cases where the thought may be judged and those where it may not. And this is to say that while there may be a grammatical purpose for introducing '⊢', there is no logical point to it as Frege and Russell had supposed. Its only purpose is to indicate the scope of the entire remark or sentence to be asserted; that is, its

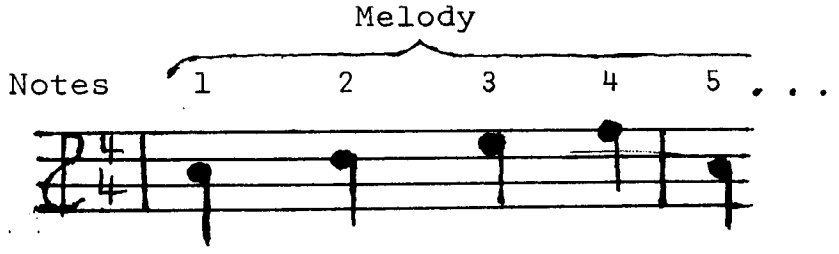
function, as Wittgenstein says in (IId), is like that of a full-stop--to indicate the whole period and thereby to mark the occurrence of a sentence like Q in P1 as a clause within the period. But this is a purely grammatical task, and one for which, in natural languages at least, we already have a sign, viz. the period. Thus, when it is introduced into the formal languages of the Grundgesetze and Principia, its only purpose (and one which is not strictly necessary in such languages) can be to act as the period does in English. Logically, the sign has no point; that the thought expressed by Q cannot be judged in ⌜P&Q⌝ is already apparent from the occurrence of '&'--a sign expressing a particular operation on P and Q. The same result holds of course for English expressions like 'and' and 'not' in their purely extensional aspect. That such expressions also have a non-extensional role and in addition carry certain non-logical connotations not relevant to the determination of the truth-conditions of the expressions in which they occur is beside the point here. For these considerations quite obviously have no bearing on the use of '⊢'.[19]

In (IIc), Wittgenstein lists two mistakes involved in

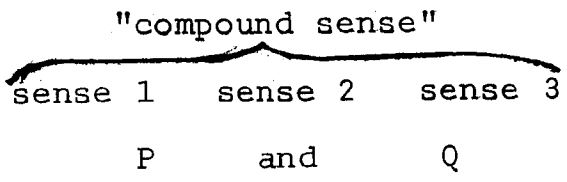
---

[19]As Wittgenstein expounds the matter at (IId), he may have in mind a much broader application of the point under discussion. His example of "it is raining" seems to indicate that it could occur in purely intensional contexts like 'Jones believes that it is raining'. However, we are not concerned with such cases in this essay. Moreover, in the Tractatus, Wittgenstein maintains that all functions of propositions are extensional.

the use of the assertion-sign. The first has to do with the idea of judging as an "advance" from the grasping of a thought to the "acknowledgment of its truth", and we must for the present forego a discussion of this point. But the second one provides the underlying basis for the remarks in (IId). That is, that in grasping (and subsequently, judging) the thought, we follow the sentence much as we sing from a musical score. The mistake, that is, consists in supposing that just as we grasp a melody by attending to the "note-signs" of the score in a step-by-step, left-to-right fashion, so we grasp the sense of a truth-functionally compound sentence by similarly attending to the component expressions. Reading the sentence is like following the score, but "thinking it" (grasping its sense) is not at all like grasping the melody. This is not a psychological point; Wittgenstein is making the logical point that the sense of the compound sentence is <u>not</u> composed of the senses of the component sentences in the quite straightforward way in which a melody is composed of an ordered series of notes. We might think of the matter in the following way:



However, the following is <u>not</u> an accurate picture:



But this, of course, is the consequence of the Frege-Russell view: there is a sort of isomorphism holding between the compound sentence and the thought it expresses--a view which is easily discredited once the nature of a truth-operation has been understood.

Final Comments on the Theory of Assertion

It is now time to return to the first purpose underlying Frege's theory of assertion. It was noted in chapter one that Frege meant to include under the general heading of "asserting" all those illocutionary acts wherein our concern is for truth, i.e. wherein we commit ourselves to the truth of the thought expressed by the spoken or written sentence. Thus, acts such as reporting, stating, affirming, denying, claiming, etc. would all be cases of "asserting" in this broad sense, and in such cases the corresponding thought may be said to be "judged". However, as Frege rightly notes, there are many contexts where our concern is not for the truth of the thought expressed. Thus, though ordinarily an actor in a play may be said to be "reporting" something, he would not really be asserting in Frege's sense. For neither he nor anyone else in such a context is concerned with the truth-value of the thought expressed (though the actor may

perhaps be "ostensibly" concerned with it). Thus, asserting, as Frege has characterized it, occurs only when one of a certain braod range of illocutionary acts has been performed in what we shall hereafter call a "standard context".

Now this special, Fregean sense of 'assert' was distinguished from the ordinary sense (i.e. as just one of the various illocutionary acts in question) back in chapter one, and the difference between them was described as "logical" vs. "psychological" respectively. However, it might well be questioned how "logical" Frege's sense really is. A philosophical investigation into this subject would certainly fall within the realm of the philosophy of language, but it is by no means obvious that it ought to be regarded as a central topic in the philosophy of logic. An examination of the contextual conditions of utterance and of the various problems concerning the illocutionary acts just cited (e.g. how exactly they are to be individuated) does not seem to fall within the province of the latter discipline. And this raises the more fundamental question of to what extent the concept of assertion is a logical concept at all rather than a "psychological" or linguistic one--especially in view of the fact that the alleged "logical" purpose discussed above was seen to be pointless.

One can of course well understand why Frege wished to introduce this sense of 'assert', for as we have seen he wished to emphasize the distinction between merely grasping

a thought and making a judgment concerning it, and hence it is natural for him to want to delineate those uses of a sentence ("assertive" uses) whereby the thought is being judged by the speaker rather than, for example, being put forward for consideration. But though Frege thus had a reason for at least mentioning this sort of "assertion", it remains difficult to understand why he should attach any logical importance to it. It is of course the counterpart of the (presumably mental) act of judging, but this just raises the question all over again--to what extent, if any, is such an "act" a logical notion? More specifically, it is difficult to understand what exactly is the role of the assertion-sign in this connection. Is it that when placed in front of a written sentence, it indicates that the writer holds the thought expressed to be true--that in other words the author would be prepared to utter that sentence in the course of performing one of a variety of illocutionary acts in a standard context? But if so, then, as Wittgenstein says in (I), there would seem to be no <u>logical</u> role fulfilled by '⊢', and though it could perform a "psychological" role (in the very broad sense of that term intended by Wittgenstein), there would seem to be no necessity for introducing it on these grounds. After all, we are all of us pretty adept at determining whether an author or speaker is asserting a sentence or not, and it is doubtful that the introduction of '⊢' (or a corresponding set of sounds) would be of any assistance in

this regard. We can usually determine in advance--and with little or no effort--whether a book is intended as fiction and if a context is a standard one. And in a work of non-fiction and a standard context, there is the general presumption that the author or speaker means to assert the written or spoken sentences which are, so to speak, "full-stopped" unless there is an explicit indication to the contrary. Moreover, the "impersonal" character of the assertion-sign would render its usefulness at best minimal. If we do not already know the author of a given sentence, the assertion-sign will not help us; the closest English equivalent of '⊢' would seem to be 'it is asserted'. But it should be noted in this regard that whereas ⌜if it is asserted that P, then...⌝ is legitimate, we may not write ⌜if ⊢P, then ...⌝. And it should also be noted that given Frege's psychological role for '⊢', it would be a gross mistake for an actor in a play to use the assertion-sign to distinguish between asserted and unasserted occurrences of the sentences he is uttering or writing down--thus illustrating the enormous disparity between the psychological and logical roles Frege has assigned to '⊢'. In any case, since Frege's "logical" grounds for distinguishing between asserted and unasserted occurrences of a sentence rest on a conceptual mistake, one can easily understand why Wittgenstein claims that assertion has only a psychological sense, and hence the assertion-sign can at best perform the rather unnecessary

and non-logical task of indicating that the author has in fact judged the thought expressed by the flagged sentence to be true--that he has asserted or is at least prepared to assert the sentence in question.

And yet we cannot dismiss the matter so easily. For Frege certainly did think that '⊢' plays an important logical role, and one which is distinct from (though closely related to) the logical role discussed earlier in connection with the distinction between asserted and unasserted occurrences of a sentence. He may not have actually distinguished the two roles as such, but there can be no doubt that he is committed to both. This second "logical" role was mentioned briefly in chapter one, and is succinctly stated by Geach:[20]

> In a language where the assertion-sign is used, a proposition that lacks the assertion-sign will in fact be a name of the True or the False, but in writing it down we are so far not doing anything with it, no more than if we wrote down a designation of any other object; in Wittgenstein's phrase, we are only setting out the pieces on a board, not making a move in the game. What we do by writing the assertion-sign in front of the proposition is to warrant it as a name of the True.

A name as such, in addition to expressing a sense, merely designates an object, and sentences, according to Frege, are names for truth-values. However, these names are unique in that they are commonly what get asserted in an act of asserting, and their senses, unlike those of other names,

---

[20]Geach and Anscombe, Three Philosophers, page 153, his italics. See also Montgomery Furth's introduction to his translation of Frege's Grundgesetze, section 7.

may be judged. Such features, thus, do not apply to the sentence qua name and to the thought qua sense of a name. It is natural, then, for Frege to want to introduce a new expression which stands to asserting (and hence judging) as names stand to designating or referring. Thus we are led to the category of assertions--expressions consisting of a sentence preceded by '⊢'. This sign is therefore neither a name nor a variable, nor is it a metalinguistic sign meaning 'it is a theorem that...'. Rather, it is an essential and thoroughly unique unit of the object-language, and the difference between a sentence or name of a truth-value like:

A. Socrates is wise.

and the assertion:

B. ⊢Socrates is wise.

is that in (A) we are merely inscribing a name of a truth-value, but in (B) we are at the same time saying or asserting that the truth-value is the True. Thus, as we saw more fully in chapter one, an assertion is not a name or a referring expression, rather, as Frege says, it "asserts something".[21] Thus, (B) differs from (A) in somewhat the same manner as Q in the conclusion of:

P1 Either P or Q
P2 Not-P
C Q

---

[21] "Function and Concept", in Philosophical Works, page 34. The preceding sentence is almost a direct quote from the same passage.

differs from Q as it occurs in P1. In the conclusion, Q does not merely express a thought, that thought may also be judged. Similarly, in (A), we have a name which merely expresses the thought that Socrates is wise, whereas (B) expresses the judgment that Socrates is wise. Herein lies the difference between the two logical roles: the thought that Q may be judged in the conclusion (though not in P1), but in (B) we have the judging of a thought.[22] This should not be taken to mean necessarily that someone has in fact judged the thought at a certain time and place; rather, what Frege has in mind here (rightly or wrongly) is the idea that since thoughts may be judged as well as grasped, we need a further kind of expression that stands to judging as a name stands to grasping (grasping the sense of the sentential name = grasping its thought). The '⊢', when attached to a sentence to get an assertion such as (B), does not merely indicate that the author holds the thought expressed to be true; rather, according to Frege, we derive a unique kind of expression which instead of naming or designating, performs the quite different function of asserting or saying that such-and-such is the case--and the gloss on this is: the sentence is warranted as a name of the True. This is the further "logical" role performed by '⊢', and I shall now

---

[22]The conflation of these two logical roles on Frege's part may perhaps be the reason why some have interpreted him as saying that "being asserted" cannot attach to a sentence unless its thought is true.

try to show that it is just as spurious as the first.

Let us begin by taking a closer look at the notions of a sentence, an assertion, and (an act of) asserting; we shall then move on and consider the mentalistic counterparts of thought and judgment. The distinction between "assertion" and "asserting" should be clear by now; the former is a type of (well-formed) expression, e.g. (B) above, whereas the latter refers to the various illocutionary acts that may be performed by a speaker in a standard context, and in which a (normally indicative) sentence is uttered or written. Thus I reserve the noun, as Frege does, for expressions like (B), and use only the verb to indicate the act.

Toward the end of chapter one, it was pointed out that on Frege's analysis, sentences are construed as value-expressions designating "objects" in exactly the same manner as other complex expressions, e.g. definite descriptions. In a word, they are "names", and such an analysis tends to pass over, or at least minimize, the rather distinctive, logical features of sentences as such--for example, they are, as Wittgenstein says, moves in the game, and those which normally perform assertive moves (those in the indicative mood) express senses which are bearers of truth-value. They thus are uniquely the kind of expression which is uttered in an act of asserting. And we further saw that it is in virtue of this latter feature that an indicative sentence can be described as "saying" or "asserting" something.

For this use of 'say' and 'assert' seems to derive from that use wherein a speaker uses a sentence to "say" or "assert" that such-and-such is the case. We must, then, in an act of asserting use an expression whose sense can bear a truth-value;[23] it would be absurd to suppose that a definite description could be used in an act of asserting or that one could meaningfully attach an assertion-sign to it. Only an expression which may be used by a speaker in an act of asserting can itself be sensibly described as "saying" or "asserting" something, and such an expression will express a thought. Sentences, then, are what we might now call assertable expressions--they are the kind of expression required for performing any of the illocutionary acts falling under Frege's general heading of "asserting". Definite descriptions, ordinary proper names, and the like will thus be examples of unassertable expressions. Assertable expressions hence are what Wittgenstein at (IIa) calls moves in the language-game; thus, 'that Socrates is wise', 'the thought that Socrates is wise', and 'Socrates' being wise', though in some way expressing the thought expressed by (A), are nevertheless not assertable, and hence do not assert or say that Socrates is wise.

Now (A) itself would certainly seem to be an assertable

---

[23]This is not to deny, of course, that there are assertive contexts where it would be both appropriate and natural for a different kind of expression to be used elliptically for a sentence; for example, one may assert something by using a definite description to answer a question.

expression, but on Frege's theory it is also a name. And a name, besides expressing a sense, merely designates an object --it merely picks out that object by means of its sense (the reference of an expression being a function of its sense.) And to the extent to which, for whatever plausible reason, we analogize sentences with the paradigmatic cases of names (ordinary names, definite descriptions, etc.), then to just that extent does the assertable feature of sentences become problematic for us, and it is at just this point that Frege introduces the assertion-sign to take up the slack. Furth expresses it this way:

> ...the distinction (between assertions and the expressions to which the assertion-sign attaches) is what makes it possible to press the analogy between sentences and names, the advantages of which have already been discussed. Otherwise, we should find ourselves repeatedly blocked by the fact that precisely what differentiates sentences from other expressions is the fact that they can be used to assert something, a use in which the analogy with names is much less persuasive.[24]

Now there can be no doubt that Frege pressed the analogy as far as he could, and thus a problem arises. On the one hand, there is the seemingly obvious fact that (indicative) sentences are paradigm cases of assertable expressions--indeed, in English and similar natural languages, they seem to be the only such cases--and on the other, that the analysis of sentences as value-expressions designating the True or the False requires that they be classed with the

---

[24] The Basic Laws of Arithmetic, Introduction, pages 51-52.

paradigm cases of unassertable expressions. Thus Frege must decide whether to accept the straightforward intuitive considerations for analyzing sentences as expressions of an essentially different logical genre from other complex expressions--in which case their assertable feature is retained and the analogy with the standard cases of names rejected--or accept the consequences of applying the theory of functions and the sense-reference distinction to sentences--in which case they must be construed as "names" and some further, additional element must be introduced to account for their assertable character. Clearly he opted for the latter alternative; and he proposes to explain away the assertable feature of this very unusual type of complex name in the following way: the sentence per se, being only a complex name, is despite appearances to the contrary an unassertable expression. Something therefore must be added to the sentence, thought not in such a way as to alter its sense--in order to make an assertable expression. And by introducing the '⊢' to perform the function of warranting such a complex name as a designation of a certain object (the True), we thereby can derive a further kind of expression which does bear the assertable feature. As Frege says, we move from a name to an assertion; in producing (B) from (A), we move from an expression which merely expresses the thought that Socrates is wise and designated the True (or the False) to one which asserts that Socrates is wise--one which certifies the name's reference to the True.

But what, then, about the plain and incontrovertible fact that in ordinary discourse it is sentences without such an addition that get uttered or written in an act of asserting? Why do we assert expressions like (A) rather than (B)? Why is it that we do not preface an allegedly unassertable expression like (A) with some written or verbal counterpart of the '⊢'? Frege is not very explicit on this point, but he seems to hold that what is signified by his special sign '⊢' occurs implicitly in the assertive utterance or inscription of sentences of ordinary discourse. In other words, it signifies explicitly what is revealed to a hearer or reader by those contextual considerations, seriousness of intent, etc. which normally accompany the utterance or inscription of a sentence wherein the speaker commits himself to the truth of the thought expressed rather than, for example, where the thought is merely put forth for consideration. But now the purpose of '⊢' is no longer to produce an assertable expression from an unassertable one; rather the purpose is purely "psychological", viz. to indicate someone's actually asserting an expression which must, then, already be assertable. That is, the difference between p and ⌜⊢p⌝ is no longer that of name of a truth-value vs. expression which asserts that p. Rather, since '⊢' is supposed to be a symbolic rendering of what is essential in an act of asserting, it is the difference between the mere occurrence (spoken or written) of an assertable expression

and the actual asserting of that expression--with the tacit understanding of a particular speaker or writer on a particular occasion. If the '⊢' is supposed to function as the formal counterpart of what occurs in an act of asserting, then the "p" in "⊢p" must already be assertable inasmuch as it is what gets asserted; it cannot be an unassertable expression which, when '⊢' is added, gets transformed into an assertable expression. Frege has thus given us two conflicting interpretations of the sign "⊢p" which correspond to the psychological role and the seond logical role just introduced. The latter role seems plausible enough until we ask why it is that allegedly unassertable expressions like (A) are in fact assertively uttered instead of artificial expressions like (B) which have no counterpart in natural languages but which, according to Frege, are really the assertable expressions. And since the two roles were never distinguished as such by Frege, he attempted to explain away the lack of any sign or set of sounds corresponding to his '⊢' by saying that this sign is the formal counterpart of what occurs in those utterances or inscriptions properly called "assertive".

But it may be objected here that though Frege has assigned two distinct tasks to '⊢', they are not really incompatible. It might be argued that two things are accomplished in an act of asserting: first, the sentential name

that is uttered becomes, by virtue of its being uttered in the course of performing that kind of act, an assertable expression; and secondly, the sentence, so transformed, is actually asserted. The sign '⊢', hence, ferrets out two distinct features of the assertive act, and Frege's only error was in using the same sign for both features; but the "p" in ⌜⊢p⌝ is always a name, not an assertable expression as such. Now whatever further problems there might be with such a view, two questions arise immediately:

a. Why build the "assertable" feature into the act of asserting rather than the sentence uttered in the course of performing that act?

b. What exactly is the relation of '⊢' in ⌜⊢p⌝ to the second feature of the assertive act (the actual "asserting")?

The answer to (a) should by now be obvious. On Frege's view sentences are names, and names are unassertable. Hence, the assertable feature, contrary to our intuitions, does not belong to the sentence per se, and the only other way of accounting for that feature is to build it into the act. Hence, if sentences are not names, there is no longer any reason for supposing that the assertable feature belongs to the act, and every intuitive reason for supposing what seemed to be obvious anyway, viz. that the feature belongs to the sentence as such.

We shall return to this point later.

Concerning (b), does the '⊢' in ⌜⊢p⌝ merely perform the psychological task of indicating that the author holds the thought expressed to be true? Frege often speaks as if

in ⌜⊢p⌝, p is actually being asserted, or that it "asserts" that p, where this does not merely mean that p is an assertable expression. Again, '⊢' performs the purely psychological task, ⌜⊢p⌝ would be itself a content; specifically the content: 'I(the author) hold the thought expressed by p to be true', or perhaps 'I judge (have judged?) that p', or, given the presumably "impersonal" character of the sign, 'It is asserted (by someone) that p'.

But surely Frege is reading far more into the sign ⌜⊢p⌝ than is possible. For to say that ⌜⊢p⌝ asserts that p in a sense other than merely being an assertable expression is to conflate the weak sense of 'assert' or 'say', wherein an expression may be described as saying or asserting (an "assertable" expression), with the strong sense wherein a person uses such an expression to say or assert that something is the case--to express his commitment to the truth of the thought expressed. But Frege sometimes speaks as if he regards the sign ⌜⊢p⌝ not simply as an assertable expression (one which "asserts" in the weak sense) not as merely recording the fact of an actual assertive utterance of p performed by a particular person on a particular occasion nor as merely indicating in a performative way that the author holds the thought expressed to be true ("I hereby assert..."), but rather--inasmuch as '⊢' is supposed to warrant p as a name of the True--as if ⌜⊢p⌝ somehow asserts p in the sense in which a person may be said to do

this.[25] Again, '⊢' is not supposed to be an assertable content or part of such, it is supposed to express the asserting of such a content. In the process, it is forgotten that only a person, by his use of an expression (a "content") on a particular occasion, can be said to "assert a content"--assert in the strong sense. But such things are easy to forget if one remains content to describe ⌜⊢p⌝ in vague terms like "the assertion that p". For here the contrast between p and ⌜⊢p⌝ can either be one of sentence versus the asserting of the sentence or, at least in the Fregean theory, of an unassertable expression (a "name") versus an assertable one. In the former case, we have the strong sense of 'assert' with tacit reference to a particular person and circumstance, and it would be absurd to suppose that ⌜⊢p⌝ can assert in this sense. At best, it asserts in the weak sense that someone did in fact assert (in the strong sense) that p, or at least is asserting in that sense (as in: 'I hereby assert...'). Indeed, no expression can be devised which asserts in the strong sense.

---

[25] I am not saying here that Frege explicitly believed that ⌜⊢p⌝ asserts p in the strong sense, but rather that he failed to draw the strong-weak distinction, and this is the source of his difficulties. It is because of his failure to draw the distinction that he came to think that the expression ⌜⊢p⌝ asserts p in a sense stronger than that of merely being assertable. That he thought this seems evident from the passage of "Function and Concept" cited earlier, as well as section 5 of the Grundgesetze where he says: "We therefore require (a) special sign to be able to assert something as true."

As Wittgenstein says in (I), no proposition can assert of itself that it is true. But for all that, a sentence like (A) not only expresses its sense, it says or asserts that Socrates is wise--again, we should keep in mind 4.022 of the Tractatus. And the difference between (I) and 4.022 is that in the former case it is denied that a proposition asserts in the strong sense, but in the latter it is affirmed that it asserts in the weak sense--and without requiring the addition of '⊢'.

Hence Frege's ⌜⊢p⌝ can either be an assertable expression or it can perform the quite distinct role of indicating that p in fact is being or has been asserted by someone (in the strong sense)--that he has verbally or in written form expressed his judgment that p. It can perhaps also serve simply to express the general idea of asserting as opposed to uttering a sentence for some other purpose. But none of these roles amounts to ⌜⊢p⌝ asserting in the strong sense that p, for there is no sense in which an expression can be said to do this despite the seemingly intelligible notion of '⊢' as "warranting" p as a name of the True. As used by someone in a particular context, '⊢' can express the person's warranting of p; it can perhaps also perform the task of indicating that someone has warranted the name. But it cannot, in the sense in which the person does, warrant the expression, and for this reason, ⌜⊢p⌝ is itself an expression which can be warranted (a "content"). Again, ⌜⊢p⌝ can

indeed express the notion of asserting a content, p, and it is of course necessary to distinguish between a content and an act; but it does not follow that simply because '⊢' is intended as the formal counterpart of the act of asserting, that it, when placed before the sentence, asserts the sentence--that we thereby have an execution of the act. Or again, ⌜⊢p⌝ can express the fact of someone's judging or having judged the thought expressed by p, but though it can be vaguely described as "judging that p"--which is not itself a content but a content with something being done to it--it does not follow that in ⌜⊢p⌝, the thought expressed by p is being judged. There is nothing wrong in introducing '⊢' as the formal counterpart of the act of asserting--and correspondingly, to express the act of judging--there is nothing wrong, that is, in introducing a sign for an act and contrast it with the other signs of the language which express contents of the act. But what does not follow is that when this sign is actually conjoined with a content-sign, that we thereby have performance of the act with that content as its object; rather, we can only produce another content. '⊢' may express or be the "formal counterpart of" the assertive act, but it cannot accomplish what is accomplished by someone's performing the act. It looks possible for ⌜⊢p⌝ to assert p in a sense stronger than merely producing an assertable expression only because the psychological and logical roles of '⊢' have been conflated; the strong and

weak senses of 'assert' and 'say' now become to all intents and purposes fused together.

Thus, Wittgenstein's accusation in (I) is correct; Frege winds up committing the same error he repeatedly warned against, viz. supposing that an expression can not only be assertable, but also an "asserting" of something. But with Frege, the mistake is not the simple-minded one of confusing assertion with prediction, and he repeatedly emphasizes that ⌜⊢p⌝ cannot be equated with any of the following:

> The thought that p is true.
> I assert that p.
> It is asserted that p.
> p is hereby warranted as a name of the True.
> It is true that p.

All of these are, in one way or another, assertable contents --they express thoughts which are contents of judgment, for they can occur within larger contexts, whether extensional or intensional.[26] Moreover, all could be uttered or written in a purely non-assertive context like a play or a novel. What is supposed to make ⌜⊢p⌝ unique--and '⊢' sui generis--is that here we do not have a content to be asserted, but in some sense "the asserting of a content"--an act and a content. But as we just saw, though '⊢' can signify the act (serve

---

[26]The performative examples in the above list raise some further problems, and I do not wish to suggest that in being assertable contents they are in all respects similar to the more familiar examples like the first item. Such problems, however, are beyond the scope of this essay.

as the linguistic representative or counterpart of the act), it cannot do what gets done by someone's performing the act. It is quite impossible for a sign to do this. But if one needs a sign to produce an assertable expression from a name, and at the same time overlooks the difference between being assertable and asserting, one might well suppose that it can be done. And it will then be quite tempting to dismiss the nagging question of just how '⊢' can be a legitimate expression of the object-language and yet not behave logically like 'it is true that...', 'it is asserted that...', etc. by passing it off as "unique" and sui generis.

Frege's idea that the sign ⌜⊢p⌝ actually asserts something in a sense stronger than simply being an assertable expression thus arises from his confusing the quite distinct roles assigned to '⊢'. On the one hand, there is the purely logical role of producing an assertable from an unassertable expression (the second of the two "logical" roles), and on the other, there is the broadly speaking "pschological" role of functioning as the linguistic or formal counterpart of the assertive act. In addition, we earlier considered the first of the logical roles played by '⊢' wherein it was supposed to mark an "asserted occurrence" of a sentence. We have already seen that this first logical role is spurious, and now it is an easy matter to dispose of the second. For it clearly turns on the supposition that sentences are names--that they designate the True or the False, just as ordinary

proper names designate their bearers. As such, the sentence per se must, contrary to our intuitions, be classified as unassertable, and its seemingly assertable character--the brute fact that it is what gets assertively uttered--must now be accounted for by introducing a further logical element. But as we saw back in chapter two, sentences are not names--the supposition that they are is incompatible with the truth-condition principle (which the remarks on "indirect specification of sense" have now clarified) and the principle that the sense of an expression cannot be specified by merely specifying what its reference will be under certain conditions. But if sentences are not names, there is no longer any reason for denying that they are assertable expressions--no longer any reason for denying what seemed intuitively obvious in the first place. The only reason for denying the obvious--that sentences are assertable--was that Frege's admittedly plausible doctrine of functions and sense-reference entails that sentences are names. The second logical role of '⊢' thus disappears, for once it has been clearly distinguished from the psychological role, its sole raison d' etre is to supply the indicative sentence with something it already has, viz. the assertable feature. Therefore, both of the allegedly "logical" roles of the assertion-sign are unnecessary, and '⊢' is "logically quite without significance". As Wittgenstein says in (I)[28]

[28]See page 270.

assertion is merely psychological; judgment stands on the same "psychological" level with question and command, and '⊢' can only indicate that the author holds the thought expressed by the sentence it flags to be true. And as he suggests in (IIc), we can use '⊢' in this psychological role, but for the reasons given earlier, there would seem to be no practical gain in so doing.

That '⊢' is supposed to produce an assertable from an unassertable expression is brought out well by Wittgenstein at (IIa). Since for Frege sentences are names, and hence unassertable, he must conceive of them in terms of unassertable correlates of actual sentences; phrases like 'that Socrates is wise'. Thus, something more is needed to produce an assertable expression--what Wittgenstein here calls a "move" in the language-game.[29] The addition of '⊢' or 'it is asserted' is thus required for producing an assertable expression, and one can indeed imagine a language,which, instead of straightforward sentences like 'Socrates is wise', contains only that-clauses and appropriate completing expressions like '⊢' or 'it is asserted'. The only trouble is that '⊢' could not also serve to distinguish the assertive act from what is asserted. For that which is asserted

[29]The term 'move' admits of a strong-weak distinction just as 'assert' and 'say'. When Wittgenstein says a sentence is a "move", he is using it in the weak sense. However, a sentence or assertable expression may be described as the sort of expression necessary for a speaker to make a move in the game. Here, 'move' is used in the strong sense.

(the "content") is produced by combining '⊢' or 'it is asserted' with 'that Socrates is wise'. As Wittgenstein here suggests, Frege has tried to graft the content versus asserting-of-a-content distinction (or thought versus judging-a-thought distinction) onto the assertable expression versus name distinction--as if the name was the content instead of the assertable expression, and that providing the assertable feature is one and the same with "the asserting of the content". We have already seen the error in this, but it is worth noting that if '⊢' is somehow supposed to express the act of asserting, then as Frege notes, the expression ⌜if ⊢p then ⊢q⌝ is nonsense. But if it is supposed to operate like 'it is asserted' in the kind of language just mentioned, then since the antecedent and consequent of the conditional are presumably themselves assertable expressions, ⌜if ⊢that-p then ⊢that-q⌝ would certainly not be nonsense.

In (IIc), Wittgenstein begins by admitting that one can use '⊢' in the psychological role, but then, in what is certainly a reference to Frege, goes on to mention two "mistakes" involved in its use. We have already examined the second of these, and we are now in a position to give a brief account of the first--that is, the mistake of supposing that "the assertion consists of two actions, entertaining and asserting (assigning a truth-value, or something of the kind)."

It was said earlier that Frege's theory of assertion

has as its genesis the important but often overlooked distinction between grasping a thought and judging it.[30] Now it ought to be stressed here that none of the points made so far *contra* Frege in any way undermines this distinction, but they do, I believe, put it in proper perspective. For one thing, it is basically a "psychological" or perhaps epistemic distinction--between understanding what is expressed by a sentence on an occasion of its use and holding that which is expressed to be true. As such, it is not a logical distinction, but it is nonetheless of great use in refuting various quasi-Aristotelian theories of the proposition.[31] However, Frege thought it was a logical distinction, and we need not search very far to discover why. For it should now be clear that the distinction was in Frege's mind inextricably fused with the admittedly logical distinction between an expression which merely designates an object (a name, an unassertable expression) and one which in a sense Frege never clearly understood "asserts" or "says" something. Since sentences are names, to grasp or "entertain" a thought is only to grasp or entertain the sense of a name. But since names are unassertable, something further is required to get an assertable

---

[30] For an interesting account of some of the errors that have been committed by overlooking the distinction, see Peter Geach's essay "Assertion", Philosophical Review, 1965. Geach seems to swallow whole Frege's doctrine concerning '⊢', but this does not affect his remarks on the distinction in question.

[31] See the discussion on page 272 of this chapter.

expression, and because of the conflation of the strong and weak senses of 'assert', this "something further" produces not just an assertable expression (one which is the proper content of the assertive act), but also an "asserting" in the strong sense--an expression which is supposed to do in the symbolism what only a person can do by uttering or writing the sentence in a standard context. Hence it was natural for Frege to correlate the act of "judging that p" with the expression '⊢p'. Thus the grasping versus judging distinction is wrongly grafted onto the name versus "assertion" distinction, with the result that the act of judging iw now mistakenly glossed as an "advance from a thought to a truth-value"--from merely grasping the sense of a certain kind of name to the warranting of a certain object as its reference (what Wittgenstein calls "assigning a truth-value" and what Frege himself calls "acknowledging the truth of the thought"). The grasping versus judging distinction thus seems to take on a logical air, and the interpretation or glossing of judgment as an advance from a thought to a truth-value--as having moved from a mere entertaining of the sense of a name to warranting the True as the name's reference--follows quite readily once one has first accepted the notion of a sentence as a name (and hence of having both a sense and a reference), and second, confused the strong and weak senses of 'assert'. It is this construal of the notion of judging which Wittgenstein rejects at (IIc); judging (or as

he there terms it, "asserting") is only psychological, and what is judged is not the sense of a name, but the sense of an expression which is already assertable. Since it is assertable, no further logical element like '⊢' is needed.

Toward the end of the discussion of Frege's theory of assertion in chapter two, it was suggested that Frege had overlooked an assertoric feature of the sentence as such, and had wrongly built it into the act of asserting (and by analogy, the act of judging). It should be clear by now that this "feature" is the assertable character of an indicative sentence which the sense-reference distinction requires Frege to play down. It is what distinguishes the sentence (a "move" in the game) from a corresponding that-clause, and from the more standard cases of names. An exploration of the concept of an assertable expression is beyond the bounds of this essay, but as the preceding discussion has shown, it is of great importance for logical theory that the concept not be overlooked. It was also suggested in chapter two that Frege's theory creates too sharp a gap between a thought and its truth-value--that the connection between the two is much closer than one of sense to reference. An expression to which the sense-reference distinction applies may yet have a sense but in fact no reference, and if sentences are names of truth-values, then as we saw more fully in chapter one, they at least can lack a truth-value owing to lack of reference on the part of a

component expression. Thus, there can be thoughts which are neither true nor false, and in "On Sense and Reference" Frege works out a Strawson-like "presupposition" doctrine.[32] But once we deny that sentences are names, not only do we undermine the supposedly "logical" features of asserting and judging, we also remove Frege's basis for claiming that a thought stands to its truth-value as a sense to a reference,[33] and by extension his ground for saying that there are thoughts which are neither true nor false. In contrast, we find Wittgenstein in several places suggesting (thought not in this terminology) that a thought, qua sense of a sentence or assertable expression, must be essentially true-false,[34] and indeed the picture analysis of sentences presented in chapter three, and based on some of Wittgenstein's remarks, does seem to imply this. We cannot, however, discuss this question any further; it must suffice to note that Frege's

[32] I do not wish to suggest that Strawson regards sentences as names, for it is clear from "On Referring" that he does not (Mind, Vol. 59, 1950). But it should be noted that Frege at least gives a theoretical backing to his presupposition theory--he can offer an explanation of it in terms of the sense-reference distinction and the theory of functions generally. This is more than can be said for Strawson who seems to think it is enough to appeal to "what we would say".

[33] This does not strictly follow, for there are expressions other than names to which Frege applies the sense-reference distinction, viz. unsaturated or functional expressions. But it should be obvious from chapter one that sentences cannot have "references" of this sort either; such references cannot be the True or the False.

[34] See for example 4.063 and 6.111 of the Tractatus, and also paragraphs 135 and 136 of the Philosophical Investigations.

interpretation of the act of judging as an advance from a thought to a truth-value--from a sense to a reference--is wrong, and hence some other ground (if there is any) must be supplied to support the Frege-Strawson thesis concerning the sharp separation of thought and truth-value.[35]

Sentences as Names of Truth-values

In chapter two, a reductio argument against Frege's doctrine of sentences as names of truth-values was proposed based on Wittgenstein's remark at 4.431. The argument showed that the doctrine is wrong, but did not provide many clues as to what is wrong with it. In what follows, I shall try, in a rather brief and capsulized way, to indicate what the error is.

We saw in chapter one that it will not do to object to Frege's doctrine simply on the ground that it runs counter to our intuitions. For Frege has given us a plausible argument, based on the theory of functions, that sentences are value-expressions referring to or designating objects which are values of certain kinds of functions for their arguments. And from here, as we also saw, it quite strictly follows that these objects are exactly two in number--one the reference of all sentences expressing true thoughts, the other the reference of all those expressing false ones. We need not worry about this latter thesis, for it falls if the

---

[35]For an intriguing discussion of this matter, see chapter three of Anscombe's Introduction.

former one does. Let us then briefly review Frege's reason for regarding sentences as value-expressions referring to or designating objects.

Briefly, the reason consists in Frege's extension of the theory of functions from purely arithmetic functions, like $\xi+2$, where the value-expression designated a number, to functions where the value-expression expresses a sense which is true or false, e.g. $\xi^2 = 1$ and ξ is wise. The arithmetic function is discerned in the following list:

1+2
2+2
3+2
etc.

and represented by the unsaturated functional expression '$\xi+2$'. By completing this expression with an argument-expression, say '3', we produce the value-expression '3+2' which designates the number 5. Thus, 5 is the reference of '3+2', 3 the reference of '3', and the function itself the reference of '$\xi+2$'. Frege quite naturally extended these ideas so that we have functions revealed in the lists:

$1^2 = 1$	Socrates is wise
$2^2 = 1$	Plato is wise
$3^2 = 1$	Aristotle is wise
etc.	etc.

Here again, the functions are represented by (are the references of) the unsaturated expressions '$\xi^2 = 1$' and 'ξ is wise', and the arguments of these functions (numbers and men) are referred to by the argument-expressions '1', '2', 'Socrates', etc. But if we are to carry this out fully,

each of the value-expressions in the above lists must designate objects which are the values of these functions for the arguments in question--just as '3+2' designates the value of the function $\xi+2$ for the argument 3. Hence the idea of sentences as value-expressions designating objects of one sort or another.

Thus, the degree to which it is plausible to treat sentences as names depends on the degree of plausibility in construing $\xi^2 = 1$ and ξ is wise as functions differing in no important logical respects from the paradigm cases like $\xi+2$. But however plausible such a construal might at first appear, it will not stand up under more intense examination. For the paradigm cases are operations whereas the others are examples of what were earlier called "material functions", and hence the latter functions differ from the arithmetic one in all the various ways delineated earlier in this chapter. But Frege attempted to treat the material functions along exactly the same lines as the arithmetic operations, and so just as '3+2' designates an object, so '$2^2 = 1$' and 'Socrates is wise' must designate objects. He thus failed to notice the important distinction between operations and material functions, and indeed his demand for "rigor" in specifying functions leads him to overlook it. For it is just this demand which prompts him to ascribe characteristics uniquely possessed by operations to material functions as well. For example, reiteration properly applies only to

operation-expressions like '$\xi+2$', but the demand for rigor requires that a function be specified for all "possible" arguments. And because Frege attempted to introduce material functions along exactly the same lines as arithmetic operations, the value-expression '$(2^2 = 1)^2 = 1$', though nonsense in standard mathematical discourse, becomes a legitimate well-formed expression of his formal system to which a reference must be assigned. The material function $\xi^2 = 1$ must be specified not just for numbers as arguments, but for all possible arguments ("objects") including the True and the False which are allegedly designated by expressions like '$2^2 = 1$'. By the same token, '$\text{Socrates}^2 = 1$' and '$(\text{Socrates is wise})^2 = 1$' are legitimate as well. Thus, material functions are forced into the mold of operations, and the genuine discrepancies between the two, like reiteration, are legislated away.

But this is not all. For though material functions are pressed into the mold of arithmetic operations--thus artificially acquiring some of their distinctive characteristics --the otherwise commendable demand for rigor also leads Frege to take away from operations some of their other, equally distinctive, features. For genuine operations, whether they be arithmetic ones or truth-functional ones like $\neg\xi$, are also specified by Frege for all possible arguments, and thus "value-expressions" like 'Jones+2' and '$\neg 4$' are arbitrarily assigned references. But as we

saw, it is essential in an operation that it be a function whose values and arguments be of the same genre--from numbers to numbers, or from thoughts (truth-values) to thoughts (truth-values), etc. For operations are essentially the means by which we may indirectly specify one entity of a particular genre in terms of (as a function of) one or more other such entities. And this feature gets lost when, for example, Frege specifies functions like $\overline{\ \ }\xi$ as if they were "concepts", i.e. exactly analogous to the stock examples like ξ is wise.

Thus, sentences are construed as names because material functions like $\xi^2 = 1$ and ξ is wise are treated as if they were arithmetic operations in certain key respects. In an arithmetic operation, we move from expressions designating numbers to other such expressions, and hence for Frege sentences analogously come out as designations of the True or the False. The qualification "arithmetic" is important here. For once we understand the idea of an operation, it becomes clear that the operation tells us nothing about the kinds of expressions which occur as bases and results. It should not be thought that because we move from names to names in arithmetic, that we must do the same with truth-operations. Rather, as we saw earlier, we may in this latter case legitimately hold that the expressions which serve as bases and results (sentences) have no references at all. Frege thought they did because truth-operations were wrongly

assimilated to material functions (the standard "concepts" and "relations") which, in turn, were wrongly assimilated to the paradigmatic arithmetic operations where the notion of designating certainly occurs. As Wittgenstein implies at 5.515, the range of significant substitutions in an operation-expression is revealed by the symbolism itself; there is no need, therefore, for stipulating such ranges (in the manner of Russell), nor need we go in the other direction and succomb to Frege's unnecessary, and in fact mistaken, demand for specifying a value for all arguments whatsoever. Thus, the operation-material function distinction not only casts doubt on Frege's attempt to analogize sentences with complex arithmetic designations, it also, if our explication of operations is anywhere near correct, shows that it is positively wrong to attempt to draw such an analogy in the first place. The singular, uncompounded sentences of ordinary discourse are certainly value-expressions, but since they are generated from unsaturated expressions representing material functions--functions remarkably different from those represented by arithmetic operation-expressions--this fact does not in any way make them referring expressions or "names".

The foregoing is of course very rough and sketchy; it is not so much an attack as a proposal for pursuing a line of attack. A systematic development of the ideas suggested here would take us far beyond the scope of this essay.

Nevertheless, I hope it provides at the very least some hints as to what is wrong with the Fregean line of argument whereby it is concluded that sentences are names of truth-values.

We have come a long way since first considering Wittgenstein's reductio back in chapter two. From there we produced a reductio against Frege's theory of assertion, in the course of which we had occasion to introduce the independent criterion. That criterion was then explored further with the result that the standard extensional account of 'not' was found to be insufficient. We pursued this line further in chapters three and four and gained a more satisfactory account of negation--and by extension, the other truth-functional concepts--as well as some new insights into the nature of singular propositions generally. The distinction between extensional and non-extensional aspects of 'not'--and the related notions of direct and indirect specifications of sense--was then utilized in explicating Wittgenstein's difficult notion of an operation by means of which we were able to refute one of the most important features of Frege's theory of assertion. And we were also by that time in a position to explore the rest of theory as well--thus enabling us to discern why the reductio of chapter two was so effective. Finally, even though they have not been fully worked out, we have acquired some insights concerning the philosophical errors involved in treating sentences as names. And this is of no little consequence in a time when there

are still philosophers and logicians who seriously hold this view, and casually dismiss objections to it by saying that the question of whether to analyze sentences of ordinary discourse in this way, or to accept or reject formal systems wherein sentences are so construed, is a purely "external" one--a pragmatic "choice" to be decided by weighing the relative advantages and disadvantages. One might as well claim that it is a similarly "pragmatic" question as to whether the earth is flat.

SELECTED BIBLIOGRAPHY

Anscombe, G. E. M. An Introduction to Wittgenstein's Tractatus. London: Hutchinson University Library, 1959.

Anscombe, G. E. M. and Geash, P. T. Three Philosophers. Ithaca: Cornell University Press, 1961.

Black, Max A Companion to Wittgenstein's Tractatus. Ithaca: Cornell University Press, 1964.

Church, Alonzo Introduction to Mathematical Logic. Princton: Princeton University Press, 1956.

Frege, Gottlob The Basic Laws of Arithmetic, edited and translated by Montgomery Farth. Berkeley: University of California Press, 1964.

Frege, Gottlob "The Thought", translated by Anthony Quinton in Mind, vol. 65 (1956).

Geach, P. T. and Black, Max, eds. Philosophical Works of Gottlob Frege. Oxford: Basil Blackwell, 1960.

Lewis, C. I. and Langford, C. H. Symbolic Logic. New York: Appleton Century Crofts, 1932.

Quine, W. V. O. Mathematical Logic. Cambridge: Harvard University Press, 1940.

Russell, Bertrand Logic and Knowledge, edited by Robert Marsh. London: George Allen and Unwin, 1956.

Wittgenstein, Ludwig Notebooks: 1914-1916, edited by G. E. M. Anscombe. Oxford: Basil Blackwell, 1961.

Wittgenstein, Ludwig Philosophical Investigations, edited by G. E. M. Anscombe. New York: Macmillan Co., 1953.

Wittgenstein, Ludwig Tractatus Logico-Philosophicus, translated by David Pears and B. F. McGuinness. New York: Humanities Press, 1961.

Wittgenstein, Ludwig Tractatus Logico-Philosophicus. London: Routledge and Kegan Paul, 1922.