

Deflationism about Truth and Meaning

By

Onyoung Oh

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

The Graduate Center

City University of New York

© 2013

Onyoung Oh

All Rights Reserved

This manuscript has been read and accepted for the Graduate
Faculty in Philosophy in satisfaction of the dissertation
requirement for the degree of Doctor of Philosophy

Jan. 31. 2013

Stephen Neale

Chair of the Examining Committee

Jan. 31. 2013

Iakovos Vasiliou

Executive Officer

Paul Horwich
Michael Levin
Alex Orenstein

Supervisory Committee

The City University of New York

Abstract
Deflationism about Truth and Meaning

By
Onyoung Oh

Adviser: Professor Paul Horwich

Abstract

The aim of my thesis is to defend a deflationary view of truth and meaning. I characterize the view as a doctrine holding that truth is a purely logical notion, and truth-theoretic notions don't play a serious explanatory role in an account of meaning and content. We use truth-terms (e.g. 'true') everywhere, from the discourse of ordinary conversation to those of the hard science and morality. The ubiquity of truth-terms gives rise to the impression that truth is a profound notion playing substantive explanatory roles. This impression, say deflationists, is unduly inflated—the ubiquity of truth-terms is not a sign of the richness but thinness of the concept of truth. In my thesis, I aim to defend this view by responding to some of its well-known objections. To defend a view often involves a modification, which is especially relevant to the case of deflationism due to the plethora of its variants. I have chosen two variants—Horwich's and Field's—in order to find out what features are to be had by a well-rounded variant of deflationism. My special interest is on the merits of a deflationary theory of truth as it is applied to an account of meaning and content. The specifics of each chapter are summarized in the following.

In chapter 1, I discuss the background of a deflationary theory of truth by examining the problems with a correspondence theory of truth. I divide a correspondence theory into two kinds:

a fact-based theory and an object-based theory. As examples of a fact-based correspondence theory, Russell's and early Wittgenstein's theories are given a critical examination. I then turn to Tarski's semantic definition of truth. I argue that Tarski's definition has offered a way to develop a correspondence theory without invoking a fact or fact-like entity. I argue, however, that even a correspondence theory of a Tarski-style is vulnerable to a certain problem—the problem raised by Field. I then turn to reductive/physicalistic theories of reference—Kripke-Putnam's causal theory of reference, the information theory, and the teleological theory of representation. By arguing against each of these theories, I conclude that the prospect of a correspondence theory of truth is dim. I end this chapter by discussing how the dismal prospect of a correspondence theory of truth has motivated a deflationary theory of truth.

In chapter 2, I embark upon the core project of my thesis—developing and defending a deflationary theory of truth and meaning. I devote chapter 2 mainly to the discussion of Field's pure disquotational theory of truth. According to this view of truth, the concept of truth is at bottom purely disquotational. In this chapter, I try to elaborate and clarify the central ideas underlying this radical version of a deflationary theory of truth. To do so, I focus on some objections leveled against this view: that it cannot accommodate the modal properties of truth and logical derivations involving an attribution of truth to sentences that one does not understand. After criticizing Field's solutions to these problems, I propose my own solutions.

The topic of chapter 3 is the success argument against a deflationary theory of truth, according to which a deflationist cannot make sense of the explanatory role of truth in an account of the success of behavior or theories. In the first half of this chapter, I examine Nic Damnjanovic's supervenience/compatibilist objection to deflationism. I argue that a supervenience approach to truth is incompatible with a deflationary theory of truth. In the second

half, I discuss Kitcher's realist objection to deflationism. Drawing upon the role of truth in an account of the success of scientific theories, Kitcher contends that realism requires a non-deflationary—correspondence—concept of truth. I criticize Kitcher's argument on the grounds that it conflates the objectivity requirement with the systematicity requirement. I argue that only the first is needed to accommodate the role of truth in an account of the success of a scientific theory.

In chapter 4, I aim to defend a deflationary theory of meaning and content. To this end, I carry out three projects—first, defending Horwich's use theory of meaning against Kripke's skeptical challenge; second, bringing out the commonalities between Horwich's and Field's views of meaning and content; and third, arguing for Field's deflationary analysis of the role of truth-conditions in psychological explanations. More precisely, I try to bring out the core ideas running through some deflationists' views of meaning and content such as late Wittgenstein, Horwich, and Field. By doing so, I aim to explain what it involves to state that truth-theoretic notions don't play a central role in an account of meaning and content, which is the main thesis of Horwich's and Field's deflationary views. I end this chapter by defending Field's view of truth-conditions—not only truth but also truth-conditions are expressive, not explanatory, devices aiding generalization.

I dedicate this dissertation to my late grandparents.

Acknowledgements

First and foremost I want to express my heartfelt gratitude towards my adviser, Paul Horwich. It has not only been an honor but a life-changing experience to work with him. Thanks to his guidance, I could recognize the sloppiness of my arguments and the biases crippling my understanding. His rigorous, but patient supervision has strengthened my philosophical muscles and widened my philosophical perspectives.

I also want to express my gratitude towards Michael Levin for his insightful and relentless comments. Some of the discussions in my thesis would not have existed without his comments. I am also greatly indebted to Michael Devitt for his detailed comments. In particular, his comments on the debate between scientific realism and deflationism helped me clarify my position. I would like to thank Alex Orenstein for reading and commenting on earlier drafts, and he encouraged me when I got off the track. I could not have completed this thesis without Stephen Neale's understanding and support. As the chair of the examining committee, he has offered me the precious opportunity to complete the revision of my thesis. I am also greatly indebted to Iakovos Vasiliou, the executive officer of the department of philosophy at the Graduate Center of the City University of New York. His warm understanding and encouragement has helped me endure the tough times of the Ph. D pursuit.

It would like to express my deepest gratitude towards John Greenwood. His invaluable support has rescued me from despair, self-doubt, and many temptations to give up the Ph. D pursuit. I cannot thank him enough. I also want to express my gratitude towards Hartry Field. I became interested in deflationism about truth and meaning by reading his works on the topic. His candid but often intriguing approach has greatly inspired me, and I have enjoyed solving

philosophical puzzles raised by him in the way Sherlock Holmes solves his cases. I am also happy to inform my former professor, Stephen Schiffer, of the completion of my thesis.

I also thank Gary Seay at Medgar Evers College of the City University of New York. Teaching at Medgar Evers College has opened my eyes to American society. I also thank Matthew Moore (at Brooklyn College), Enrique Chavez-Arviso, Jonathan Jacobs, and John Pitman (at John Jay College) for their exceptionally warm support while I was struggling with teaching and working on my thesis. I am deeply appreciative of the help and understanding offered by Mr. Douglas Ewing at the Graduate Center.

My dear friends Carrie Figdor and Jared Blank spared their time to read and comment on parts of my thesis. My friend Karen Ramkissoon has always been there when I needed someone to talk to.

Lastly, I want to thank my family for their love and support. I thank my father for being a Socratic gadfly to me. I am appreciative of the valuable advices my Mom gave me when I was lost in the dark. I thank my sister and my brother-in-law's warm support. Above all, I thank my nephew, Louis, for having been my best friend during past four years.

Table of Contents

<u>Abstract</u>	iv
<u>Acknowledgements</u>	vi
<u>Table of Contents</u>	viii
<u>Introduction</u>	1
Chapter 1: <u>The Correspondence Theory of Truth and the Motivations for Deflationism</u> .	8
1.1 <u>The Correspondence Theory of Truth</u>	
1.1.1 <u>Russell’s Correspondence Theory of Truth</u>	
1.1.2 <u>Wittgenstein’s Correspondence Theory of Truth</u>	
1.2 <u>Field’s Criticism of Tarski’s Theory of Truth</u>	
1.3 <u>The Motivations for the Deflationary Theory of Truth</u>	
1.3.1 <u>The Motivation I: Problems with Reductive Theories of Reference</u>	
1.3.2 <u>The Motivation II: The Role of Truth</u>	
Chapter 2: <u>Field’s Pure Disquotationalism: Separation of Truth from Meaning</u>	66
2.1 <u>Introduction</u>	
2.2 <u>The Modal Objection</u>	
2.2.1 <u>The Problem and a Misunderstanding</u>	
2.2.2 <u>The Quasi-Disquotational Solution</u>	

2.3 Attributions of Truth to Sentences in Another Language

2.3.1 Attributing Truth to a Sentence in a Language that One Understands

2.3.2 Attributions of Truth to Sentences in a Language That One Doesn't Understand

2.3.3 A Different Approach

Chapter 3: Deflationism and Success Arguments 125

3.1 Introduction

3.2 Damnjanovic's Criticism of Deflationism

3.2.1 The Standard Deflationist Response to the Success Argument

3.2.2 Damnjanovic's Criticism of the Standard Deflationist Response

A. The Distinction Between Causal Efficacy and Causal Relevancy

B. Deflationary Truth is a Supervenient Property

3.2.3 The Problems with Damnjanovic's Argument

3.3 Kitcher's Realist Attack against Deflationism: A Case For the Causal-Explanatory Role Of

Correspondence Truth In the Success-To-Truth Rule

3.3.1 Introduction

3.3.2 Horwich's View on Truth and the Realism Debate

3.3.3 Kitcher's Arguments

A. Kitcher's Criticism of Horwich's Explanation

B. The Map-Example

3.3.4 The Criticisms of Kitcher's Argument

Chapter 4: <u>Deflationist Views of Meaning and Content</u>	178
4.1 <u>Introduction</u>	
4.2 <u>Horwich's Use Theory of Meaning and its Response to Kripke's Skeptical Paradox</u>	
4.3 <u>The Commonalities Between Horwich's and Field's Deflationist Views of Meaning</u>	
4.3.1 <u>Meaning-Entities</u>	
4.3.2 <u>A Fact of The Matter</u>	
4.4 <u>Field's Response to The Success Argument: A Different Version</u>	
<u>Bibliography</u>	237

Introduction

The objective of this dissertation is to examine what a deflationary theory of truth and meaning is about. The term ‘deflationary’ is now used so abundantly that it may have little explanatory function—to call a theory ‘deflationary’ may say little about the theory. That said, in a debate about truth, it has a well-rounded meaning—that truth is a purely logical concept whose sole purpose is to express generalizations over statements, beliefs, sentences, propositions, claims or assertions. To take a mundane example, we sometimes wish to express our attitude toward a belief, claim or proposition, but we find ourselves not being able to articulate the belief, claim or proposition because we don’t clearly remember it. In such a case, the notion of truth is handy—by uttering a truth-statement such as

What John said is true,

we can avoid making an infinite conjunction such as

If John believes that snow is white, then snow is white; if John believes that grass is green, then grass is green; and so on.

To say that truth is a purely logical notion or property is to say that the role of truth is exhausted by a generalization of this kind.

An advocate of a deflationary theory of truth defends the deflationary conception of truth just characterized—that truth is nothing but a logical concept the sole purpose of which is to express generalizations over beliefs, sentences, or propositions. There are both positive and negative sides of this theory. The positive side is that if this deflationary concept of truth is accepted, it should be quite easy to develop an adequate theory of truth—an adequate theory of truth just needs to explain how the notion of truth has the generalizing function. Thus, a deflationary theory typically offers a very simple explanation of the concept of truth; for example, the disquotational schema such as ‘ p ’ is true iff p or the equivalence schema such as the proposition that P is true iff P . The negative side of the deflationary conception of truth is that it doesn’t square well with the time-honored view of truth in western philosophy. The time-honored view is that truth plays important roles in explaining certain aspects of thoughts and language; for example, the notion of truth plays an important role in explaining the effectiveness of beliefs in achieving our goals; it plays a central role in explaining the meaning of an expression and the content of a cognitive state (e.g., a belief or desire state); it plays an important role in explaining the reasonableness of logical principles; and so on. The notion of truth also appears in almost all philosophical discussions of importance; for example, the debates such as whether moral claims can be said to be true, whether we have an access to truths about unobservable entities, whether there is an objective truth in an absolute sense, etc. Even outside philosophy, truth is deemed as a serious and preponderant notion; truth is considered as having an important value (e.g. ‘true friend’, ‘truthful mind’, etc.).

From a deflationist point of view, however, it is merely a misconception to consider the frequent employment of a truth-term as the sign for truth being a deep philosophical concept waiting for a substantive analysis. The reason why the notion of truth is invoked so frequently is

not that truth is a property with a deep nature but that it has the generalization function. Truth pervades philosophical discussions and ordinary conversations for the same reason that expressions such as ‘it’ and ‘that’ appear frequently in English language; it is a generalizing device.

A deflationary theory comes in a wide variety.¹ The discussions in this thesis will mainly be concerned with the ones developed by two contemporary deflationists—Paul Horwich and Hartry Field. I will defend their views—in particular, I will defend their views by arguing against some objections raised about them. Given that the notion of truth is invoked in the discussions of all major philosophical topics, the implication of endorsing a deflationary theory can be discussed with respect to a wide range of philosophical problems. In discussing Horwich’s and Field’s deflationary theories, however, I will only be concerned with its application to semantic problems—the problems of meaning and content. It is undeniable that the notions of truth and meaning are closely connected. For this reason, it is inevitable that a theory of meaning be affected by a theory of truth; for example, a theory of meaning presupposing a correspondence theory of truth is expected to meet certain conditions that are not required when a deflationary theory is assumed. An aim of this thesis is to examine what a theory of meaning should be like

¹ Some of the varieties of deflationary theories of truth are the redundancy (or the disappearance) theories of truth advocated by Ramsey and Frege; the performative theory of truth formulated by Strawson; the prosentential theory of truth developed by Belnap, Camp, and Grover; the simple theories of Prior and Williams; the disquotational theory of truth that can be found in Quine, Tarski, Ayer, and Field; and Horwich’s minimalist or deflationary theory of truth. See F. Ramsey, “Facts and Propositions,” 1927, in D. H. Mellor, ed., *Philosophical Papers*. Cambridge University Press, 1990; G. Frege, “The Fundamental Laws of Arithmetic,” *The Monist* 25(4), 1915, pp. 481-194, “The Thought,” 1956, in P. F. Strawson, ed., *Philosophical Logic*, Oxford University Press, 1973; D. Grover, J. Camp, N. Nelnap, “The Prosentential Theory of Truth,” *Philosophical Studies* 27, 1975, pp. 73-125; A. Prior, *Object of Thought*, Oxford: Clarendon Press, 1971; M. Williams, W. V. Quine, *Philosophy of Logic*, Englewood Cliffs: Prentice-Hall, 1970; A. Tarski, “The Semantic Conception of Truth and the Foundation of Semantics,” *Philosophy and Phenomenological Research* 4, 1944, pp. 341-376; A. J. Ayer, *Language, Truth, and Logic*, 2nd ed., London: Gollancz, 1946; P. Horwich, *Truth*, 2nd ed., Oxford: Blackwell, 1999 (1990); H. Field, “The Deflationary Conception of Truth,” in G. McDonald and C. Wright (eds.), *Fact, Science, and Value*, Oxford: Blackwell, 1986, pp. 55-117 and “Deflationary Views of Meaning and Content,” in his *Truth and the Absence of Fact*, Oxford: Clarendon Press, 2001.

when the notion of truth is construed in the deflationary sense. Both Horwich and Field have advanced their own views of meaning to match their deflationary theories of truth. There are differences between their views; there seem to be more disagreements in their views of meaning than there are in their views of truth. My aim, however, isn't to take sides with one of them; it is to find some important commonalities between their theories that can help us have an overarching outlook on a deflationary theory of meaning and truth. More specifically, what I will argue in each chapter is briefly as follows:

In chapter 1, the background of the deflationary conception of truth will be discussed. Although a non-deflationary or inflationary theory of truth means any theory that doesn't endorse the deflationary conception of truth, a correspondence theory of truth is considered the major opponent of a deflationary theory of truth. The intuitive appeal of the correspondence conception of truth is that it is grounded on our basic conception of truth exemplified as follows:

'Snow is white' is true iff snow is white,

or

The proposition that snow is white is true iff snow is white.

Unlike a deflationist, a correspondence theorist attempts to justify this basic conception of truth by defining truth in terms of a correspondence relation. I will explain why this attempt hasn't succeeded by discussing the problems with correspondence theories developed by Russell, the early Wittgenstein, and Field in his inflationary phase. Field's criticism of Tarski's theory of truth and the problems with reductive theories of reference will be discussed as a way of criticizing a correspondence theory. I will argue that the problems with a correspondence theory

and the deflationist view of the role of truth constitute two motivations for a deflationary theory of truth.

In chapter 2, I will discuss Field's disquotational theory. During his deflationary phase, Field has defended a theory called 'pure disquotationalism'. It is difficult to say what this theory is in one sentence, because Field has kept changing his view of it. Initially, Field characterized it as a theory stating that the notion of truth is at bottom purely disquotational, where purely disquotational truth is characterized as a use-independent notion. It was also characterized as one that can meaningfully be ascribed only to sentences that one understands. This view, however, has confronted serious objections one of which is the modal objection, and the other is that it cannot explain logical derivations involving an ascription of truth to sentences one doesn't understand. I shall explain what these objections are, and how Field has responded to them. In particular, I will argue that his latest response is still short of resolving the problems. I will propose my own solution to the problems.

In chapter 3, the so-called success argument against a deflationary theory of truth will be examined. What I mean by 'the success argument' is one that argues for a non-deflationary theory of truth on the basis of the role played by truth in an account of the success of behavior. In the first of the chapter, I discuss Nic Damnjanovic's argument according to which a Horwich-style deflationist theory of truth shows that truth plays a causal-explanatory role in an account of the success of behavior. To argue for this unusual view, Damnjanovic invokes Jackson and Pettit's program explanation of a mental property (a sort of supervenience theory of a mental property). I argue against Damnjanovic's view on the grounds that Jackson and Pettit's program explanation cannot be applied to the deflationary conception of truth. In the second half of this chapter, I discuss Kitcher's argument according to which scientific realism requires a

correspondence theory of truth. In defending this view, Kitcher appeals to the success argument—that a deflationary theory of truth cannot make sense of the role played by truth in an account of the success of a scientific theory. On Kitcher’s view, a deflationary theory isn’t neutral between scientific realism and scientific anti-realism, because a deflationist cannot make sense of the role played by truth in an account of the success of a scientific theory. I argue against Kitcher—I argue that Kitcher’s demand for an account of systematic success is orthogonal to explaining the role of truth in an account of the success of a scientific theory. Kitcher assumes that the success of all scientific theories can and should be explained by a small number of basic principles. Kitcher then argues that a deflationist cannot explain the systematic success. I argue, however, that a deflationist doesn’t have to explain systematic success in order to make sense of the role played by truth in an account of the success of a scientific theory.

In chapter 4, I discuss deflationary theories of meaning advanced by Horwich and Field—Horwich’s use theory of meaning and Field’s linguistic view of meaning attributions. Both Horwich and Field defend theories of meaning in which truth-theoretic notions don’t play important roles. Applying what he calls ‘the deflationary spirit’ to an account of meaning, Horwich aims to show that skepticism about a theory of meaning is largely due to “pseudo-constraints” imposed on an adequate theory of meaning. I shall concentrate on two of the six pseudo-constraints discussed by Horwich—the relationality and representation constraints. Horwich’s treatment of these constraints shows a way in which a theory of meaning becomes free of misconstrued presupposition by having the deflationary concept of truth. To make the point vivid, I discuss Horwich’s solution to Kripke’s skeptical argument against Wittgenstein’s use theory of meaning. In the second part of this chapter, I compare Horwich’s deflationary theory of meaning with Field’s. I compare their theories with respect to two issues—the issues

about meaning-entities and a fact of the matter by virtue of which a meaning-attribution is true. Viewed superficially, their theories of meaning are opposed to each other with respect to these issues. I will argue, however, that important similarities are hidden under superficial differences. By arguing for the hidden similarities between Horwich's and Field's theories of meaning, I intend to find some core ideas that can characterize a deflationary theory of meaning and content. In the last part of this chapter, I examine Field's response to the success argument. Field's response to the success argument is different from Horwich's in that it is more concentrated on the role of *truth conditions* than that of truth in an account of the success of behavior. I will argue that the moral to be drawn from Field's response to the success argument is that not only truth but also truth conditions are to be construed as purely expressive devices in a deflationary theory of truth and meaning.

Chapter1. The Correspondence Theory of Truth and the Motivations for Deflationism

1.1 The Correspondence Theory of Truth

The topic of this section and the section following it is the correspondence theory of truth. The purpose of discussing the correspondence theory isn't, however, to examine the features and prospect of the theory *per se*; it is to explain the transition from the correspondence theory to the deflationary theory of truth. To do so, I will concentrate on some of the objections raised of the correspondence theory. I have a very modest ambition in carrying out this task. As with any theory of truth, the correspondence theory of truth is based on the correspondence concept of truth: the idea that truth consists in a relation to reality. Presumably, this concept of truth is the very one that ordinary people associate with 'is true'. As a result, numerous accounts have been proposed to clarify and systematize the correspondence concept of truth. Due to the lack of knowledge and space, it is impossible for me to go over all the different theories of correspondence truth. In the following, I will discuss only some of the most often-discussed versions of the correspondence theory of truth and their problems.

In order to avoid an unnecessary confusion, I will first clarify the meaning of the 'correspondence concept' of truth. To say that truth is a correspondence notion doesn't just mean that the notion of truth is closely related to the notion of reality. As Horwich points out, no one—including the deflationist—denies that the proposition that snow is white is true because snow is white.² Kunne makes a similar point; he says, "[Y]ou do not become a partisan of a

² Horwich distinguishes the correspondence intuition from the correspondence theory of truth. By 'correspondence intuition' he means the sentiment that statements owe their truth to the nature of reality. The correspondence theory of truth is based on this correspondence intuition, but it must go far beyond merely accommodating this intuition. Most of all, the correspondence theory of truth—if it is to be called a 'theory'—must *define* truth by invoking the correspondence intuition. The difference is huge. When we say that the ordinary notion of truth accommodates the correspondence intuition, it doesn't necessarily imply that the correspondence intuition is built into the definition of

correspondence conception of truth simply by assenting to the slogan that what somebody thought or said is true if and only if it agrees with reality.”³ According to Kunne, one becomes a partisan of a correspondence conception by taking the truth predicate ‘is true’ as expressing a *relational* property; i.e., a seriously *dyadic* relation.⁴ Put more precisely, to say that truth is to be understood as a correspondence notion is to say that a sentence of the form ‘*x* is true’ is to be understood in the way a sentence such as ‘Mary loves a musician’ is understood. From ‘Mary loves a musician’, we can infer ‘There is someone in the world whom Mary loves’. Similarly, to accept the correspondence theory of truth—to construe truth as a correspondence notion—implies that we should be able to infer, from a truth sentence of the form ‘*x* is true’, a sentence of the form ‘There is something in the world which *x* corresponds to’.

There is a wide variety of truth-bearers; the truth predicate can be applied to sentences, utterances, statements, propositions, beliefs, assertions, and so forth. Therefore, in discussing a theory of truth, it is important to specify what the truth predicate is primarily ascribed to. In this thesis, I will assume that the truth predicate is primarily ascribed to a sentence or sentence-like entity. The reason is heuristic. First, the version of the correspondence theory that will be discussed below is mainly characterized as a theory of truth for sentences. Second, two deflationists whose theories I will concentrate on—Horwich and Field—disagree on this very issue.⁵ While Horwich takes propositions as primary truth-bearers, Field objects to characterizing the deflationary theory of truth as a theory of truth for propositions. Given that there is such a

truth. It only means that there is room for making sense of the correspondence intuition. To say that truth is defined in terms of correspondence notion is to say that truth is a correspondence property. There will be more discussion on this point later in this chapter. See P. Horwich, *Truth*, 2nd ed., Oxford University Press, 1998, Ch. 1 and Ch. 7.

³ W. Kunene, *Conceptions of Truth*, Oxford University Press, 2003, Ch. 3.

⁴ By ‘seriously dyadic relation’, Kunene means a property that obtains two objects or entities in the world. Kunne says that from ‘Bell fell into oblivion’, nobody would seriously conclude ‘There is something into which Bell fell’. But from ‘Mary loves a psychopath’, we conclude ‘There is someone whom Mary loves’. So, ‘love’ expresses a seriously dyadic relation. See W. Kunne, *ibid.*

⁵ See H. Field, “Critical Notice: Paul Horwich’s *Truth*,” *Philosophy of Science*, Vol. 59, No. 2, 1992.

disagreement, it is better to characterize a theory of truth for an entity that both deflationists can accept. Moreover, ‘proposition’ is a term of art—it is a term that is introduced to do a specific job.⁶ As such, there is a disagreement among philosophers about how ‘proposition’ is to be understood.⁷ To avoid being tangled into such a complicated issue from the beginning, I decided to take sentences or sentence-like entities as the primary truth-bearers. So, in the following, I will apply the correspondence theory to sentences, not to propositions.

Since there are many varieties of the correspondence theory, let’s begin with the idea that all the correspondence theories are committed to.⁸ As was mentioned above, the correspondence theorist construes ‘is true’ as expressing a seriously dyadic relation. In this sense, in the correspondence theory, the logical form of a truth-statement of the form 1)

1) x is a true sentence

is 2) below:

2) True (x, y).

‘True’ here expresses a two-place, correspondence relation. So, 2) should be in turn understood in terms of 3) in the following:

3) Corresponds (x, y).

⁶ See S. Schiffer, *The Things We Mean*, Oxford University Press, 2003, Ch.1.

⁷ See S. Schiffer, *ibid.*

⁸ See W. Kunne, *Conceptions of Truth*, Oxford University Press, 2003, Ch. 3 (“Varieties of Correspondence”).

Two questions arise immediately. First, what substitutes for ‘y’ in 3)? Put differently, what is the *relatum* of the correspondence relation? According to the correspondence theory, to say that a sentence is true is to say that there is something that the sentence corresponds to. What is the thing that a true sentence corresponds to? Second, what does it mean to say that a true sentence *corresponds* to something? What is the nature of the correspondence relation? An acceptable correspondence theory must address both of these questions. In the following I will examine how and whether these questions can be answered by the correspondence theory.

Let’s first deal with the question of the *relatum* of the correspondence relation; what does a true sentence correspond to? Broadly, correspondence theorists have been trying to answer this question in one of the following two ways: one is that the *relatum* of the correspondence relation is a fact (or fact-like entity), and the other is it is an object.⁹ Following Kunne, I will call the first ‘the fact-based correspondence theory’ and the latter ‘the object-based correspondence theory’.¹⁰ According to Kunne, the object-based correspondence theory has a much longer history; basically, most correspondence theories developed before the 20 century should be considered object-based correspondence theories.¹¹ In the following, I will ignore the traditional, object-based correspondence theory, and focus on the fact-based correspondence theory developed by the 20 century philosophers. That said, it is not necessary that a correspondence theory invoke fact-like entities. As will be explained later, Tarski’s definition of truth provides a way of developing a correspondence theory of truth without invoking fact-like entities.

⁹ This categorization is not mine; I referred to Kunne’s *Conceptions of Truth*. I found this categorization very helpful to have a better understanding of a correspondence theory of truth. Marian David also makes this distinction in “The Correspondence Theory of Truth,” *Stanford Encyclopedia of Philosophy*, 2002/2009. But David didn’t make this distinction in his *Correspondence and Disquotation*, Oxford University Press, 1994.

¹⁰ W. Kunne, *ibid.*

¹¹ See W. Kunne, *ibid.* Also see Marian David, “The Correspondence Theory of Truth,” *Stanford Encyclopedia of Philosophy*, 2002/2009. Some exceptions are (according to Kunne and David) correspondence conceptions of truth entertained by Aristotle, Hume, and Mill. See Aristotle, *Metaphysics*, D. Hume, *A Treatise of Human Nature* (1749), and J. S. Mill, *System of Logic* (1843).

1.1 Fact-Based Correspondence Theories: Russell and Wittgenstein

According to the fact-based correspondence theory, to say that a sentence is true is to say that it corresponds to a fact. This theory was proposed by some of the most prominent philosophers in the early 20th century; e.g. Moore, Russell, and Wittgenstein.¹² In particular, Russell and Wittgenstein developed the fact-based correspondence theory in carrying out their bigger philosophical project: logical atomism.¹³ Put roughly, logical atomism is the idea that the world at its most fundamental level is constituted by simple entities (logical atoms), and these simple entities form complex entities.¹⁴ The notion of a fact figures prominently in the formulation of this idea. Although Russell's and Wittgenstein's theories have a lot in common, there are some important differences between them. One of the differences is that while Russell's theory mainly appeals to the notion of a fact, Wittgenstein's appeals to the notion of a state of affairs as well as the notion of a fact. So, in the following I will consider Russell's theory is best captured by the formulation (CF) while Wittgenstein's by the formulation (CS):

¹² B. Russell, "The Philosophy of Logical Atomism," *The Monist*, 1918, *Problems of Philosophy*, Oxford University Press, 1918, "On Propositions: What They Are and How They Mean," *Proceedings from Aristotelian Society*, Supplementary Volumes, Vol. 2, 1919. pp. 1-43, "On the Nature of Truth", *Proceedings of the Aristotelian Society*, 7: 28-49, 1906; G. E. Moore, "Truth and Falsity," in *the Dictionary of Philosophy and Psychology*, 1901-2, reprinted in: *Selected Writings*, edited by T. Baldwin. London and New York: Routledge 1993, 20-22; L. Wittgenstein, *Tractatus Logic-Philosophicus*, trans., by Pears and McGuinness, side-by-side-by-side edition, version 0.24 (August 1, 2011).

¹³ G. E. Moore subscribed to the identity theory (in "Truth and Falsity", although it is said that he later repudiated it). The identity theory of truth is considered an opponent of the correspondence theory of truth (Moore writes it that way in "Truth and Falsity"). Contemporary advocates of the identity theory of truth include J. Hornsby and J. Dodd. See J. Hornsby, 'Truth: The Identity Theory', *Proceedings of the Aristotelian Society*, XCVII: 1-24, 1997; reprinted in M. Lynch, ed., *The Nature of Truth: Classic and Contemporary Perspectives*, Cambridge, MA and London: MIT Press, 2001, pp. 663-81.

¹⁴ Russell says that the goal of philosophical analysis is to arrive at logical, not physical, atoms. This makes it difficult to understand the metaphysics of logical atomism; what does it mean to say that the world is constituted of logical, not physical, atoms? What underlies this peculiar metaphysics of logical atomism is Russell's (and Wittgenstein's) naïve representationalism (as I call it). Russell believed that the logical structure of an ideal language represents the logical structure of the world. This is a suspicious idea for many reasons. First, there is no reason to believe that a linguistic analysis can reveal the true, metaphysical, nature of the world. Second, it is not clear what is meant by 'the logical structure of the world'. See B. Russell, "The Philosophy of Logical Atomism," *The Monist*, 1918.

(CF) x is a true sentence \equiv_{Df} there is some fact y such that x corresponds to y ;

x is a false sentence \equiv_{Df} there is no fact y such that x corresponds to y .

(CS) x is a true sentence \equiv_{Df} there is some state of affairs y such that x corresponds to y and y obtains.

x is a false sentence \equiv_{Df} there is some state of affairs y such that x corresponds to y and y doesn't obtain.

I will discuss Russell's theory first, and then turn to Wittgenstein's.

1.1.1 Russell's Correspondence Theory of Truth

According to Russell, the relatum of the correspondence relation is a fact; that is to say, a true sentence corresponds to a fact, but a false sentence doesn't. Russell also says that a fact is what *makes* a sentence (proposition) true or false; in other words, a fact is a truth-maker.¹⁵ But what is a fact? What is the ontological or metaphysical nature of a fact? How many different kinds of fact are there? On what grounds can we distinguish one fact from another? Russell never provided satisfactory answers to these questions. So, one of the major problems with Russell's fact-based correspondence theory is that the nature of a fact is entirely obscure. To see why let's first examine what Russell himself says of the nature of a fact. According to Russell, a fact is a complex entity constituted of simpler entities; Russell says, "I mean by a "fact" anything that is

¹⁵ See B. Russell, *Problems of Philosophy*, Oxford University Press, 1912, Ch. 12 "Truth and Falsehood". As is well known, Russell kept changing his philosophical views throughout his career. His views of a proposition often shift as well. Indeed, it seems that Russell is simply careless in his use of this term. He sometimes identifies a proposition with a fact (B. Russell, "On Propositions: What They Are and How They Mean," *Proceedings from Aristotelian Society*, Supplementary Volumes, Vol. 2, 1919, pp. 1-43). At other times ("The Philosophy of Logical Atomism," *The Monist*, 1918) he simply means an indicative sentence by a 'proposition'. I here follow the second meaning of Russell's notion of a proposition.

complex.”¹⁶ But of course this explanation is far too thin to individuate facts; i.e., to distinguish one fact from another. Without having identity criteria for facts, it is entirely vacuous to say that a sentence is true or false by virtue of corresponding to a fact. In order for the notion of a fact is to be of explanatory value, we should be able to identify the fact that a given sentence is supposed to correspond to. As was mentioned earlier, if truth is a correspondence relation, a truth-statement of the form ‘*x* is true’ should be understood on the par with a statement such as ‘Mary loves Tom’. Suppose we have no idea how to distinguish Tom from other people. We would not be able to determine whether Mary loves Tom or not; in other words, we cannot say whether the loving-relation holds between Mary and Tom. Similarly, if the notion of a fact is to be of an explanatory value, we need know how to distinguish one fact from another; we need identity criteria for facts.

Unfortunately, Russell never provides a proper account of the identity criteria for facts. Indeed, Russell’s explanation of facts is so confusing that it makes us doubt whether it is beneficial to invoke the notion of facts in defining the truth and falsity of a sentence. A place to see the confusion is Russell’s explanation of a *negative fact*. According to Russell, facts of the rudimentary kind are *atomic* facts, where atomic facts are defined as the ones constituted of an *n*-place relation and *n* particulars. Russell says that atomic facts are those “contain only one verb and neither generality nor its denial.”¹⁷ Atomic facts, however, are of two varieties: *positive* and *negative*. It is relatively easy to understand what positive atomic facts are, but it is by no means easy to understand what negative atomic facts are.¹⁸ He sometimes says that a negative fact is

¹⁶ B. Russell, “On Propositions: What They Are and How They Mean,” p. 1.

¹⁷ B. Russell, *ibid.*, p.3.

¹⁸ There are, however, unsettled questions about positive atomic facts as well. First, according to Russell, ordinary proper names are not logically proper names; they are disguised definite descriptions. But if ordinary proper names are not names, what should count logically proper names? Only expressions that cannot further be analyzed into other expressions—for example, ‘this’ or ‘that’—may count logically proper names. But in such a case how are to

what a true negative sentence corresponds; for example, ‘Socrates isn’t alive’ corresponds to the negative fact that Socrates isn’t alive. But what is this negative fact? What is it constituted of? Since Russell denies that there is something in the world that corresponds to ‘not’, this negative fact is constituted of Socrates and the property being alive. But then, how does this negative fact differ from the positive fact that Socrates is alive? Is the negative fact that Socrates isn’t alive the same as the fact that Socrates is dead?¹⁹ At other times, Russell invokes a negative fact in order to explain the falsity of an atomic sentence; for example, ‘Socrates is alive’ is false because it corresponds to the negative fact that there isn’t such a fact that Socrates is alive.²⁰

When things are this complicated, we may need to suspect the plausibility of the basic assumption; in this case, the assumption is that a fact is what makes a sentence true or false. Does the notion of a fact do any explanatory work in an explanation of the truth and falsity of a sentence? Aren’t we (as Quine once complained) just projecting a fact from a true sentence merely for the sake of correspondence?²¹ The murkiness surrounding Russell’s notion of a negative fact supports this suspicion. Another factor contributing to this suspicion is that Russell invokes a general fact to explain the truth of a general sentence.²² Russell’s argument for invoking a general fact is that the truth of ‘All men are mortal’ cannot be inferred from the truth of each and every sentence such as ‘Socrates is mortal’, ‘Plato is mortal’, and so on. This claim

determine the fact that an atomic sentence corresponds to? If we have to appeal to “direct acquaintance”, most facts would be beyond our reach. Another problem is that it is not clear how far in space and time we should extend the range of positive atomic facts. Do facts include facts in the past and future? Do facts include those in a possible or counterfactual world?

¹⁹ Russell says that the sentence ‘Socrates is dead’ is the conjunction of ‘Socrates was alive’ and ‘Socrates isn’t alive now.’ But this answer isn’t satisfactory, since we have another negative sentence ‘Socrates isn’t alive’. See B. Russell, “The Philosophy of Logical Atomism.”

²⁰ Russell says, “[T]he absence of a fact is itself a negative fact.” See B. Russell, “On Propositions: What They Are and How They Mean,” p.5.

²¹ W. V. Quine, *Quiddities: An Intermittently Philosophical Dictionary*, Cambridge, Mass.: Harvard University Press, 1987, p. 213.

²² See B. Russell, “The Philosophy of Logical Atomism”. It seems, however, that Russell changed his view of negative and general facts later. There are also disagreements among Russell’s commentators. Some say that Russell accepted the existence of molecular facts as well (see “B. Russell”, *The Stanford Encyclopedia of Philosophy*).

is true in itself. But what reasons do we have to infer from this the conclusion that there is a general fact such as the fact that all men are mortal? We have no reason other than the question-begging one that a true sentence corresponds to a fact. If we accept Russell's fact-based correspondence theory, we have to invoke facts of many different kinds to explain the truth of sentences. For example, we need to invoke mathematical facts in order to explain the truth of mathematical sentences; we need to invoke moral facts in order to explain the truth of moral sentences; we have to invoke aesthetic facts in order to explain the truth of aesthetic sentences; and so on.²³ Indeed, we have to keep invoking a fact of a new kind whenever we confront a non-truth-functional compound sentence; i.e., a sentence of which the truth isn't the function of the truth of its constituent sentences. Russell denies that there is a molecular fact (such as a disjunctive or conjunctive fact) because the truth of a disjunctive or conjunctive sentence can be explained in terms of the truth of its constituents. But there are many molecular (compound) sentences that are not truth-functional; for example, modal sentences, counterfactual sentences, belief-sentences, and so forth. Does that mean we have to invoke modal facts, counterfactual facts, belief-facts, and so forth? We may do so, but what explanatory gain is accrued if we do so? Does invoking facts *explain* the truth of sentences at all?

The answer seems no, because we have no clear idea of how these facts *make* relevant sentences true. While we don't have any clear idea of what facts are, it is impossible to comprehend how facts *make* sentences true. If facts come in such a wide variety of ways—negative, general, mathematical, moral, counterfactual, etc.—then facts cannot be spatio-temporal entities. As

²³ Russell says that there are mathematical facts (See B. Russell, "The Philosophy of Logical Atomism," Lecture 1, *The Monist*, 1918). I don't know whether Russell acknowledged the existence of ethical and aesthetic facts.

Kunne says, facts, if they exist, should be abstract entities.²⁴ But if facts are abstract entities, in what sense do facts make sentences true? In what sense do mathematical facts make mathematical sentences true? Similar questions should be raised about facts of other kinds. In this way, the problem concerned with the nature of facts gives rise to another major problem with the correspondence theory; how sentences and facts are related to each other. Before discussing this problem, however, I would like to examine another type of the fact-based correspondence theory: the one that defines truth along the lines of (CS).^{25,26}

1.1.2 Wittgenstein's Correspondence Theory of Truth

According to Wittgenstein—who may be the first defining truth along the lines of (CS)—an atomic sentence represents a state of affairs, and the atomic sentence is true if and only if the state of affairs actually obtains in the world.²⁷ A state of affairs that actually obtains (or exists) in the world is defined as a fact. So, like (CF), (CS) defines truth in terms of correspondence to a fact. Unlike (CF) however, (CS) invokes the notion of a state of affairs. By invoking this additional notion, (CS) earns some advantages over (CF). One of the advantages is that (CS) has a better grip on falsehood than (CF); (CS) provides a simpler way of defining falsehood.²⁸

According to (CF), falsehood is a lack of correspondence to a fact, and this, according to Kunne,

²⁴ W. Kunne, *ibid.*, p. 144.

²⁵ For readers' convenience, I here reiterate (CS):

(CS) x is true \equiv_{Df} there is some state of affairs y such that x corresponds to y and y obtains. x is false \equiv_{Df} there is some state of affairs y such that x corresponds to y and y doesn't obtain.

²⁶ It may be better to call (CS) 'the state-affairs-based correspondence theory' in order to distinguish it from (CF). I may use the label in discussing Wittgenstein's version of the correspondence theory of truth.

²⁷ Wittgenstein doesn't explain what he means by 'a state of affairs that obtains.' The only account he gives is that a fact, by definition, is a state of affairs that *exists*. See L. Wittgenstein, *Tractatus Logico-Philosophicus*, first published by Kegan Paul (London) 1922. I referred to Pears/McGuinness English translation; Side-by-side-by-side edition, version 0.24 (August 1, 2011).

²⁸ I referred to Marian David (*Correspondence and Disquotation*) on this point. But the argument presented below isn't from David. His reason is that (CS), unlike (CF), can define truth as well as falsehood as relations to reality, where truth is a relation to those parts of reality that we refer to as facts. See Marian David, *Correspondence and Disquotation*, Oxford University Press, 1994, p.34.

means that a false sentence is in a discordance-relation with the same fact that a true sentence corresponds to:

Russell maintains that not only true sentences, but also false ones correspond to facts. In the case of ‘Socrates is dead’ the correspondence is accordance with the fact that Socrates is dead, in the case of ‘Socrates is not dead’ it is discordance with the same fact. (For Russell in 1918 the difference between a true belief and a false belief isn’t like that between a wife and a spinster but like that between friend and foe—of the same person.)²⁹

But then, we would have to map even a false sentence onto a fact; even a false sentence is related to a fact, where the nature of the relation is explained in terms of “discordance”. An obvious problem with this definition of falsehood is that it cries for an explanation of the nature of a discordance relation. What does it mean to say that a sentence is in discordance relation with a fact? While we don’t have a clear understanding of what is meant by “correspondence relation”, it is even harder to understand what is meant by “discordance relation”. By defining falsehood in this way, Russell added another obscurity to an already obscure situation. To see how hopeless the situation is, let’s consider a sentence such as ‘Madonna is 7 feet tall’. We know that this sentence is false. But what fact does this sentence is in discordance with? We cannot say that this sentence is in discordance with the fact that Madonna is 7 feet tall, because there is no such fact that Madonna is 7 feet tall. Presumably, the most plausible answer is that the sentence ‘Madonna is 7 feet tall’ is false because it is in discordance with the fact that Madonna is ___ feet tall, where the blank is to be filled by her exact height. But we may not know here exact height whereas we know that the sentence is false.

Wittgenstein’s correspondence theory averts this problem by defining truth and falsehood in the manner of (CS):

²⁹ W. Kunne, *Conceptions of Truth*, pp.118-119.

(CS) x is a true sentence \equiv_{Df} there is some state of affairs y such that x corresponds to y and y obtains.

x is a false sentence \equiv_{Df} there is some state of affairs y such that x corresponds to y and y doesn't obtain.

According to Wittgenstein's correspondence theory, even a false sentence represents a state of affairs, but the state of affairs represented by a false sentence doesn't obtain (or exist) in this world. Hence, if we define falsehood along the lines of (CS), we don't have to map a false sentence onto a fact; we don't have to invoke a fact in order to define a false sentence. Instead, we can define a false sentence as a sentence that represents a possible condition that isn't actualized in the world.³⁰ Besides, (CS) explains how a false sentence can still be meaningful although it doesn't correspond to a fact. Consider the following remark made by Russell:

[T]he components of the fact which makes a proposition true or false, as the case may be, are the *meanings* of the symbols which we must understand in order to understand the proposition.³¹

What Russell is saying here is that the meaning of an atomic sentence is the fact that the sentence corresponds to.³² So, if an atomic sentence doesn't correspond to a fact, the atomic sentence is

³⁰ It seems that Wittgenstein states of affairs are possible situations in which the basic constituents of this world are combined in ways that are different from the way they are combined in this world. Facts are just small parts of all possible situations. But then, Wittgenstein's states of affairs may be construed as possible worlds of a certain kind; possible worlds that are composed of the same basic ingredients constituting the actual world. The actual world differs from possible worlds not in terms of its basic ingredients but in terms of the ways in which those basic ingredients are combined.

³¹ B. Russell, "The Philosophy of Logical Atomism: Lecture 2," *The Monist*, 1918. Russell says that the only exception is a logical constant.

³² Russell's theory of meaning and reference is based on the principle of compositionality; the meaning and reference of a complex expression is determined by the meanings and references of the symbols that constitute the complex expression. The compositionality principle holds not only at the level of a language but also at the level of

meaningless. But atomic sentences can be false as well as true. Therefore, in order to make sense of the meaningfulness of a false atomic sentence, Russell had to relate such a sentence with a fact of some sort; e.g. a negative fact.

A major advancement of Wittgenstein's correspondence theory over Russell's theory is that Wittgenstein doesn't invoke a fact to explain the meaningfulness of an atomic sentence.

What makes an atomic (simple) sentence meaningful isn't a fact but a state of affairs that is represented by it. Indeed, Wittgenstein defines the notion of a state of affairs as the *meaning* of an atomic sentence. Wittgenstein says:

2.201 A picture contains the possibility of the situation that it represents.

2.202 A picture represents a possible situation in logical space.

2.221 What a picture represents is its sense.³³

A picture here is an atomic sentence, and the possibility of the situation is the existence or non-existence of the state of affairs represented by an atomic sentence. According to Wittgenstein, an atomic sentence, like a picture, comes to have a meaning by representing a state of affairs. So, the state of affairs represented by an atomic sentence is the meaning of the atomic sentence.

What constitutes a state of affairs, however, isn't something like a Fregean sense; i.e., an intensional entity such as a mode of presentation or way of thinking of an object. Rather, it is an extensional entity such as the referent of a name and the property (or class) determined by a predicate. For example, an atomic sentence such as 'Madonna is a dog' represents a state of affairs that is constituted of Madonna and the property doghood (or the class of dogs). So, if a

reality. In other words, a fact is a complex entity, and it is constituted of entities to which simple symbols correspond. From these, we can infer that a fact is the meaning of an atomic sentence to which it corresponds.

³³ What Wittgenstein means by 'picture', of course, is an atomic sentence. See L. Wittgenstein, *ibid.*

state of affairs is the meaning of an atomic sentence, it should be a meaning of a particular kind; the meaning that constitutes the *truth condition* of an atomic sentence. According to Wittgenstein, an atomic sentence represents a possible situation—state of affairs—and the atomic sentence is true if and only if the possible situation obtains. So, the state of affairs represented by an atomic sentence is the condition to be met in order for the sentence to be true; i.e., a truth-condition. If we put this together with what was stated in the previous paragraph—that a state of affairs is the meaning of an atomic sentence—we arrive at the conclusion that a state of affairs in Wittgenstein in *Tractatus* is both the meaning and the truth condition of an atomic sentence.

One may consider it an advantage that Wittgenstein's notion of a state of affairs has this double function. The fact of the matter, however, is just the opposite; it makes identity criteria for states of affairs impossible. To see why let's consider a pair of sentences involving co-referential singular terms such as 'Superman' and 'Clark Kent'.³⁴ Do 'Superman flies' and 'Clark Kent flies' represent the same state of affairs or not? On the one hand, it seems they should represent the same state of affairs, since states of affairs, according to Wittgenstein, are the truth conditions of atomic sentences. 'Superman flies' and 'Clark Kent flies' must have the same truth condition, since 'Superman' and 'Clark Kent' are co-referential singular terms.

A state of affairs, however, is also the meaning of an atomic sentence. Do 'Superman flies' and 'Clark Kent flies' have the same meaning? By almost any standard, the correct answer should be 'No.' I said "almost" because some philosophers insist that the meaning of a singular term is simply its reference and the meaning of a predicate the extension it is true of.³⁵ But as is argued

³⁴ Let's here assume, for the sake of argument, that 'Superman' and 'Clark Kent' are not fictional but real names.

³⁵ The advocates of the so-called 'direct-reference view of proper names' (e.g. Soames and Nathan Salmon) are of this view. According to the direct-reference view of proper names, the semantic content of a proper name is simply its referent, and the semantic content of a sentence containing a proper name is simply a Russellian proposition—a proposition composed of *n*-objects and an *n*-place property. Russell advocated this view, hence the name 'Russellian

by Devitt, the meaning of an expression should be a property that is concerned with causing behavior.³⁶ ‘Superman flies’ and ‘Clark Kent flies’ can cause vastly different behavior; for example, Lois Lane assents to the former, but dissents to the latter. So, if a state of affairs is the meaning of an atomic sentence, these sentences must represent distinct states of affairs.

Thus, we are split between two mutually incompatible identity criteria for states of affairs; one is that a state of affairs is the truth condition of an atomic sentence, and the other is that it is the meaning of an atomic sentence. It seems that we cannot keep both. To be sure, the truth-condition of a sentence is part of the meaning of the sentence. The problem with Wittgenstein’s notion of a state of affairs, however, is that a state of affairs is identified with the meaning of a sentence; it is as a whole the meaning of a sentence. This makes it impossible to sustain the notion of a state of affairs as it is characterized by Wittgenstein. The notion of a state of affairs in Wittgenstein’s *Tractatus* is supposed to be both an extensional and an intensional entity at the same time, but this doesn’t seem a viable idea. Something has to be given up, and what needs to be given up is the idea that the state of affairs represented by an indicative sentence is as a whole the meaning of the sentence. Instead, a state of affairs needs to be characterized as an entity that encapsulates only the truth condition of an indicative sentence, and a representational relation

proposition’. Wittgenstein also advocated this view. It should be noted, however, that Russell and Wittgenstein didn’t consider most proper names in ordinary language proper names in a strict, logical, sense. See M. Devitt, “Still Against Direct Reference,” in *Prospects for Meaning*, ed., R. Schantz, Berlin” Walter de Gruyter, 2010.

³⁶ According to Devitt, a semantic property—meaning—should be a property of an expression that plays the following two roles: explaining behavior and guiding us to reality. Devitt defends himself a truth-referential or representational view of semantic content—the view that the semantic content of an expression is entirely constituted by representational properties. But Devitt’s view is different from the direct-reference view of a proper name. According to Devitt, the semantic content of a name is a property of referring to an object under a specific mode. See M. Devitt, *ibid.*

needs to be characterized in a purely truth-conditional way as well. Contemporary advocates of a state of affairs seem to accept this modified Wittgensteinian notion of a state of affairs.³⁷

So, let's recast Wittgenstein's correspondence theory of truth using this modified notion of a state of affairs. On this modified theory, a true sentence is still defined as a sentence that represents a state of affairs that exists in the world. But the state of affairs that is represented by a sentence doesn't exhaust the meaning of the sentence; it is simply an entity that encapsulates the truth condition of the sentence. But even in this modified sense, Wittgenstein's theory has problems. First, a state of affairs is still an ontologically dubious entity. Let's reconsider the sentence 'Madonna is 7 feet tall'. According to Wittgenstein's theory, this sentence is false because it represents the state of affairs that Madonna is 7 feet tall, and this state of affairs doesn't exist in the world. But what is the state of affairs that Madonna is 7 feet tall? It is one thing to say that the sentence 'Madonna is 7 feet tall' has as its truth condition that Madonna is 7 feet tall, and another to say that the sentence represents the state of affairs that Madonna is 7 feet tall. According to Wittgenstein's correspondence theory, we have to understand the first (the truth condition of the sentence) in terms of the second (the state of affairs represented by the sentence). But what do we gain by doing so? Instead of gaining anything, we only seem to create unnecessary problems. This raises the suspicion that a state of affairs—just like a fact—is a theoretically *vacuous* notion; i.e., a notion that doesn't do any explanatory work.

The second problem with the Wittgensteinian correspondence theory is derivative of the first problem. Given that the ontological status of a state of affairs is obscure, it is hard to explain

³⁷ For example, according to the view that defines a state of affairs as a state of a possible world, the state of affairs that Madonna is 7 feet tall is true if and only if there is a possible world in which Madonna is 7 feet tall. When a state of affairs is defined this way, it has only a minimal connection with the meaning of a sentence; that is, if two sentences have the same meaning, then they should represent the same state of affairs. On the possible world account of states of affairs, a state of affairs is a conceptual apparatus that is employed to specify the truth condition of a sentence involving modality.

the nature of the representational relation between a sentence and a state of affairs. Wittgenstein avoided addressing this question entirely by saying that language can only *show*, but cannot say, how it represents the world. This is a quite an implausible idea which is based on an incorrect analogy between a picture and a language. In order to explain the representational relation between a sentence and a state of affairs, it is essential to define it recursively in terms of the representational relations between the constituents of a sentence and the constituents of a state of affairs. But even this is not enough, since a state of affairs isn't simply a conglomeration of entities; it is supposed to have a structure. For example, the state of affairs that Tom loves Mary is supposed to be distinct from the state of affairs that Mary loves Tom. Moreover, a state of affairs isn't necessarily part of the actual world; if it were, it couldn't be distinguished from a fact. This makes it even harder to explain the nature of the representational relation between a sentence and a state of affairs. Without an account of the nature of the representational relation, the nature of the correspondence relation couldn't be explained. That means the Wittgensteinian correspondence theory would have a hard time in explaining the nature of the correspondence relation. So, for the life of the correspondence theory, it would be better to advance the theory without invoking an entity such as a fact or state of affairs. In the next section, I will explain how Tarski's definition of truth shed light on that idea.

1.2 Field's Criticism of Tarski's Theory of Truth

The purpose of this section is to motivate the motivation for the deflationary theory of truth. As will be discussed in the next section, one of the chief motivations for the deflationary theory is that there is a serious problem with the correspondence theory of truth. In explaining what this problem is, it is essential to discuss Field's criticism of Tarski's theory of truth. The reason isn't

so much that Tarski's theory of truth is in itself a correspondence theory of truth. As is well known, there have been many controversies whether Tarski really intended to give a correspondence theory of truth with his definition of truth.³⁸ The merit of Tarski's theory of truth, however, seems that it is neutral between the correspondence theory and the deflationary theory. Being neutral, it has an affinity both with the deflationary theory and the correspondence theory. In this section, I will mostly focus on its affinity with the correspondence theory of truth. More precisely, I will focus on how the technique employed in Tarski's definition of truth can be of use to develop the correspondence theory of truth that is free from the problems with the fact-based correspondence theory. Field's criticism of Tarski's theory needs to be appreciated from that perspective as well; namely, from the perspective of the correspondence theory of truth developed out of Tarski's definition of truth.³⁹

³⁸ Popper, Gila Sher, and the earlier Davidson understand Tarski's definition of truth as a version of the correspondence theory of truth. Soames, the later Davidson, Etchemendy, and many others, however, don't think that there is any interesting connection between Tarski's definition of truth and the correspondence theory of truth. In my view, Devitt's understanding of Tarski's definition of truth seems most adequate: although Tarski intended his theory to be a version of the correspondence theory, the theory that he actually presented is more like a version of the deflationary theory of truth. See K. Popper, *Conjectures and Refutations*, Routledge, London, 1963, G. Sher "On the Possibility of a Substantive Theory of Truth", *Synthese* 117, 1998, H. Putnam, "A Comparison of Something with Something Else", *New Literary History* 17: 61-79, 1985, S. Soames, "What is a Theory of Truth?", *Journal of Philosophy* 81, 411-429, 1984, S. Soames, *Understanding Truth*, Oxford University Press, New York, 1999, J. Etchemendy, "Tarski on Truth and Logical Consequence", *Journal of Symbolic Logic* 53, 51-79, 1989, M. Devitt, "The Metaphysics of Truth," in *The Nature of Truth*, ed. Michael Lynch, Cambridge, MA: MIT Press, 2001, 579-61.

³⁹ As for the references for Tarski's own work, I referred to A. Tarski, "The Semantic Conception of Truth and the Foundation of Semantics," *Philosophy and Phenomenological Research* 4, 1944; "The Concept of Truth in Formalized Languages," *Logic, Semantics, and Metamathematics*. Oxford: Clarendon Press, 1956; and "The Establishment of Scientific Semantics," *Logic, Semantics, and Metamathematics*. Oxford: Clarendon Press, 1956. I also referred to secondary sources; S. Soames, 'What is a Theory of Truth?' *The Journal Of Philosophy*, Vol. 80. No. 8. 1984: 411-429, *Understanding Truth*. Oxford University Press, 1998; J. McDowell, 'Physicalism and Primitive Denotation: Field on Tarski'. *Erkenntnis* Vol. 13. 1978: 131-152; H. Putnam, 'A Comparison of Something with Something Else'. *New Literary History* 17, 1985: 61-79; G. Sher, 'What is Tarski's Theory of Truth'. *Topoi* 18, 1999: 149-166; P. Raatikainen, "Truth, Meaning, and Translation," *New Essays on Tarski and Philosophy*. (Douglas Patterson, Ed.) Oxford University Press, 2008; W. Kunne, *Conceptions of Truth*, Oxford University Press, 2003; J. Etchemendy, "Tarski on Truth and Logical Consequence", *Journal of Symbolic Logic* 53, 1988: 51-79.

Tarski's definition of truth is intended as a definition of truth for *sentences*. A sentence in Tarski's semantic concept of truth is not an uninterpreted syntactic entity; it is a fully interpreted entity with its meaning built into it. It makes sense to talk about the meaning of a sentence only when it is relativized to a language. So, Tarski's definition of truth is relativized to a specific language. There is of course an even more important reason why Tarski's definition of truth is relativized to a specific language: to avoid a semantic paradox. One of the main objectives of Tarski's definition of truth is to define truth in such a way that when we employ the concept of truth in accordance with his definition, we can avoid a semantic paradox. According to Tarski, our ordinary notion of truth gives rise to a semantic paradox because we ascribe the truth-predicate without making a hierarchical order of languages. In order to avoid a semantic paradox, Tarski suggests that we make distinctions between an object-language and a meta-language. In Tarski's theory, an object-language is characterized as the language for which truth is defined, whereas a meta-language as the language in terms of which the truth for the object-language is defined. But then, in order for the truth of an object-language to be defined in terms of the expressions of a meta-language, the vocabulary of a meta-language should always be richer than the vocabulary of an object-language. More precisely, a meta-language should involve semantic expressions that are ascribed to expressions of an object-language; it should involve either object-language expressions themselves or their translations; it should involve names or structural descriptions of object-language sentences; etc. As a result, a truth predicate that is defined in a Tarski-style cannot be legitimately ascribed a language it belongs to. Tarski's truth-definition doesn't permit an ascription of a truth predicate to a sentence that already involves a

truth predicate unless there is a difference between the levels of the truth predicates. For example a Tarski-style truth definition doesn't allow us to ascribe ' T_1 ' (the truth predicate of the level 1) to a sentence of the form ' s is T_1 ' or ' s is not T_1 '. Since self-ascription of this kind is the ultimate source of a semantic paradox, and a failure of making a hierarchical order of languages yields a semantic paradox, Tarski's definition of truth showed a way to avoid a semantic paradox.

In addition to the requirement of making a hierarchical order of languages, Tarski imposed two further conditions that an adequate definition of truth must satisfy; namely, *material adequacy* and *formal correctness*. According to Tarski, a definition of truth for a specific language L is materially adequate if it satisfies *Convention T*; that is, if it implies an instance of the following schema for each and every indicative sentence of L :

(T) s is true in L iff p ,

where ' s ' is to be replaced by the quotation-name (structural description) of an object-language sentence s , and ' p ' is to be replaced by a meta-language sentence that means the same as ' s ' in an object-language L . The rationale for the condition of material adequacy is that an adequate definition of truth must correctly determine the extension of true sentences in an object-language L . The correct extension of true sentences in an object-language is in turn determined by ordinary speakers' intuition of truth; the class of true sentences in an object-language L is the class of sentences that the speakers in L consider to be true. So, the definition of truth for an object-language L should satisfy *Convention T* so that it can determine the correct extension of true sentences in L . In addition, *Convention T* is imposed on the definition of truth for L so that the definition squares with the ordinary intuition of truth that had by the speakers of L .

Tarski's condition of formal correctness states that an adequate definition of truth for an object-language L should not employ in its definiens a semantic concept that isn't previously defined in terms of a non-semantic term. The motivation for this condition may be understood in two different ways. First, it may be designed to make sure that the definition of truth isn't circular. Tarski characterizes his definition of truth as the 'semantic definition of truth'. Since semantic concepts are inter-connected, if an undefined semantic concept is employed in the definition of truth, the definition will be circular. Second, as Field emphasizes, it may be that Tarski was really serious about the harmony between semantic concepts and the postulates of the unity of science and physicalism.⁴⁰ In other words, the condition of formal correctness may have been motivated by the desire to incorporate semantic concepts into a physicalistic view of the world.

Now, how is Tarski's definition of truth relevant to the discussion of the correspondence theory? Its relevance is that the recursive method (or technique) Tarski employs in his definition of truth introduces a new way in which the correspondence theory can be developed. Of course, it isn't Tarski who introduced a recursive method to the definition of truth. As was explained in the previous section, Russell and Wittgenstein also employ a recursive technique in characterizing their correspondence theories of truth. Tarski's definition of truth incorporates this idea but pushes this method further—into the definition of the truth of a *simple* sentence. In Russell's and Wittgenstein's theories, only the truth of a compound sentence is recursively defined. In their theories, a simple sentence is the smallest unit of a language, and so, the truth of a simple sentence isn't defined recursively. Instead, the truth of a simple sentence is defined in

⁴⁰ Field asks what the purpose Tarski's explicit definition of truth can serve that cannot be served by a definition of truth that employs an unexplicated semantic notion. Field finds the answer in Tarski's attempt to bring semantics into harmony with the postulates of the unity of science and physicalism. See H. Field, "Tarski's Theory of Truth." Kunene, however, says that Field overreacted to Tarski's somewhat incidental remark on physicalism. See W. Kunne, *Conceptions of Truth*, Ch.4, Oxford University Press, 2003.

terms of the correspondence relation to a fact or state of affairs. What makes Tarski's definition of truth ingenious, however, is that he applies the recursive technique to the definition of a simple sentence. To be more precise, in Tarski's semantics, the smallest unit of language isn't a simple sentence but a sentential function; i.e., an open sentence. And then, in order to define the truth of a sentence in terms of a sentential function, Tarski brings in the notion of satisfaction. Satisfaction is a relation between a sentential function and an object in the world, and it is itself recursively defined as follows:

(S) An open sentence ' Px_k ' in an object-language L is satisfied by a sequence of objects $\langle a_1, \dots, a_n \rangle$ iff Pa_k (where a_k is the k -th member of the sequence of objects).⁴¹

In Tarski's semantics, satisfaction is the most fundamental semantic notion in terms of which other semantic concepts such as the concepts of truth and denotation are defined. By invoking the notion of satisfaction, Tarski's recursive definition of truth comes to have two virtues that are lacked by Russell-and-Wittgenstein's recursive definition. One is that it better handles the definition of truth for a universally quantified sentence, and the other is that Tarski's definition provides the correspondence theorist with a way of dispensing with facts or states of affairs. These two features are related, since Russell had to invoke a general fact in order to define the truth of a universally quantified sentence. Wittgenstein denied the existence of a general fact;

⁴¹ If the meta-language is not an expansion of the object-language L , we have to use an expression that translates ' P '. So, Kunne says that a definition of satisfaction is materially adequate if it meets Convention S (instead of Convention T):

Convention S: A definition of satisfaction for an object-language L is materially adequate iff it implies all sentences that can be obtained from the schema (S):

(S) For any object y , y satisfies ϕ in L iff p ,

if one replaces ' ϕ ' with a revealing designator of an open sentence in L and ' p ' with a meta-language expression that translates the open sentence. See W. Kunne, *Conceptions of Truth*, 198.

instead, he regarded a universally quantified sentence as equivalent to infinite conjunctions of simple sentences. Tarski's strategy is more in line with Wittgenstein's, but it has more generality and so is simpler. Tarski defines the truth of a universally quantified sentence in the way below:

(S^u) $\forall x_k A$ is true in L iff, for every sequence s' that is different from s at most in the k -th place, A is true.

The difference between a variable and a name is that a variable can be replaced by any name; i.e., both by a name that is included in an object language and by a name that will be included in the language. So, by taking a sentential function (open sentence) as the fundamental unit of a language, Tarski could give an explicit definition of truth for a sentence involving universal quantification. Last, the truth of a sentence is defined, and in Tarski's formalized language, the truth of a sentence is taken care of as a limiting case. Tarski says,

It turns out that for a sentence only two cases are possible: a sentence is either satisfied by all objects, or by no objects. Hence we arrive at a definition of truth and falsehood simply by saying that a sentence is true if it is satisfied by all objects, and false otherwise.⁴²

In order to see, however, the full implication of Tarski's definition of truth on the correspondence theory, we have to consider how Tarski's definition can be extended to a natural language. Tarski's theory was originally put forward as a definition of truth for a sentence in a formalized language. The basic principles of Tarski's theory, however, can be used to define the truth of a sentence in natural language. Unlike a formalized language, a natural language has names, which means that a theory of reference or denotation is required. Tarski in effect defines the notion of denotation in terms of the notion of satisfaction as follows:

⁴² A. Tarski, "The Semantic Conception of Truth and the Foundation of Semantics," p. 353.

To say that the name x denotes a given object a is the same as to stipulate that the object a . . . satisfies a sentential function of a particular type. In colloquial language it would be a function which consists of three parts in the following order: a variable, the word 'is' and the given name x .⁴³

If we add to this Tarski's characterization of satisfaction, and we get 1):

- 1) To say that the name ' N ' in English denotes a given object a is the same as to stipulate that ' N 's 'France' and a is France; ' N ' is 'Germany' and a is Germany; or . . . or ' N ' is 'Korea' and a is Korea.

A statement such as 1) is Tarski's theory of denotation for names in one's own language. For foreign names, Tarski's theory of denotation goes in the manner of 2):

- 2) To say that the name ' N ' in German denotes a given object a is the same as to stipulate that ' N ' is 'Frankreich' and a is France; ' N ' is 'Deutschland' and a is Germany; or . . . or ' N ' is 'Korea' and a is Korea.

Once the notion of denotation is defined in the manners of 1) and 2), the truth of a simple sentence in a natural language L is recursively defined. In other words, a T-sentence of the form such as

- 3) ' $P(e)$ ' in L is true iff $P(e)$

or

⁴³ This is Field's quote of Tarski. See H. Field "Tarski's Theory of Truth," in his *Truth and the Absence of Fact*, p.15. It is originally from A. Tarski, "The Concept of Truth in Formalized Languages," in *Logic, Semantics, and Metamathematics*, p. 194.

- 4) ‘ $P^*(e)$ ’ in L^* is true iff $P(e)$

is derived from the definitions of denotation (such as 1) or 2) and of satisfaction.⁴⁴

Presented this way, it is clear that Tarski’s theory of truth isn’t *per se* a version of the correspondence theory of truth. Although Tarski called his theory of truth ‘the semantic definition of truth,’ there is nothing so much semantic about his definition of truth.⁴⁵ The reason has mainly due to the way he characterizes the notion of satisfaction. As was mentioned earlier, in Tarski’s theory, satisfaction is the most fundamental semantic notion in terms of which all the other semantic notions are defined. The way Tarski characterizes the notion of satisfaction, however, makes his theory almost (if not entirely) circular. To see why let’s consider (S) that was illustrated above: (S) An open sentence ‘ Px_k ’ in an object-language L is satisfied by a sequence of objects $\langle a_1, \dots, a_n \rangle$ iff Pa_k (where a_k is the k -th member of the sequence of objects). This characterization of satisfaction assumes that the predicate ‘ P ’ in the object-language L stands for the property being P . Tarski makes this assumption but not in these semantic terms; he

⁴⁴ 3) is a case in which the object-language L is the same as the meta-language; that is, L is the truth-ascriber’s own language. 4) is a case in which the object-language L^* is a foreign language.

⁴⁵ This is a well-known criticism of Tarski’s theory of truth. For example, Putnam states that Tarski’s truth predicates have nothing to do with semantics or the ordinary conception of truth. (H. Putnam, “A Comparison of Something with Something Else”, *New Literary History* 17: 61-79, 1985). Davidson agrees with this criticism in his later years (in “The Structure and the Content of Truth,” *The Journal of Philosophy*, Vol. 87, No. 6, 1990, pp. 279-328). The point of the criticism is that Tarski’s theory of truth doesn’t make any systematic connection between meaning- or use-properties with the concept of truth. Putnam’s criticism is that Tarski’s theory cannot *explain* why ‘Snow is white’ would have had been true iff snow is black if ‘white’ had meant the same as ‘black’ (or if ‘white’ had been used in the way we use ‘black’). As I stated at the outset, however, Tarski’s sentences are fully interpreted ones (Soames makes this point in his “What is a Theory of Truth?,” *The Journal of Philosophy* 81, 411-429, 1984). Tarski’s theory of truth assumes that the semantic properties of sentences are fully determined and explained by something other than the definition of truth. Indeed, Tarski’s theory cannot even explain why a certain sentence in an object-language L has the truth condition that it has. In that sense, Putnam’s criticism is correct; Tarski’s theory doesn’t provide any answer to usual semantic questions. But then, why did he call his theory ‘the semantic conception of truth’? I am not sure; it seems that he didn’t think hard enough when he used the term ‘semantic’. That said, nobody can deny Tarski’s contribution to semantics. Tarski’s theory seems to provide a logical tool for the development of semantic theories. If the logical tool (his logical-recursive analysis of language) is not of use to a theory of truth, it is still of use to a theory of meaning.

rather appeals to the notion of translation.⁴⁶ The fact that he appeals to the notion of translation doesn't make his definition of truth circular as long as the notion of translation isn't defined in terms of semantic notions (such as reference, denotation, or truth). But if Tarski really meant his theory of truth to be a *semantic definition of truth*, he eventually had to deal with a semantic question; for example, what makes the predicate '*P*' in *L* stand for the property being *P*? Tarski, however, never addressed semantic questions of this kind, and this makes his theory vulnerable to Field's criticism that I will introduce shortly.

Of course, Tarski's theory may be defended in the way Soames or Etchemendy does; in other words, by construing the goal of Tarski's theory in a deflationary way. Or as Devitt says, Tarski intended to give a correspondence theory of truth but ended up presenting a deflationary theory. But if so, what's the point of talking about Tarski's theory in discussing the correspondence theory of truth? The point is that the reductive technique found in Tarski's theory enables the correspondence theorist to develop the correspondence theory without invoking cumbersome entities such as facts or states of affairs. Russell and Wittgenstein had to invoke these entities as the relata of simple sentences mainly because they didn't have Tarski's ingenious recursive technique. Since Tarski defines the truth of a simple sentence in terms of the satisfaction of a predicate by an object, it isn't necessary to hypostasize entities such as facts or states of affairs. Instead, the correspondence theory of truth of a *Tarski-style* needs to invoke only the objects that are referred to by names and the properties that are stood for by predicates. This is a huge advantage because these entities—compared to entities such as facts or states of affairs—are both ontologically and conceptually less troublesome. So, the correspondence theory

⁴⁶ Tarski assumes that an object-language sentence that replaces '*s*' and a meta-language sentence that replaces '*p*' in an instance of T-sentence of the form '*s* is true in *L* iff *p*' have the same meaning. Tarski, however, didn't build the notion of meaning or translation into the definition of truth. Nor did he attempt to give an account of the meaning of an object-language sentence. He rather assumed that the meaning of an object-language sentence is already well defined, and used it in his definition of truth.

of a Tarski-style doesn't need to define the truth of a sentence in terms of correspondence to a fact or state of affairs that exists. Instead, it can define the truth of a sentence along the lines of 5) as follows:

- 5) 'Pa' is true in $L \equiv_{\text{Df}}$ the object referred to by 'a' in L has the property that the predicate 'P' in L stands for.

A definition of truth along the lines of 5), however, cannot by itself be considered a correspondence theory of truth, because it doesn't say anything about the nature of the semantic relations between objects and names and predicates and the properties that the predicates stand for. In effect, 5) is itself compatible with both the correspondence and the deflationary conceptions of truth.⁴⁷ In order to develop 5) into a correspondence theory of truth, we need a theory that explains the nature of the semantic relation between objects and names and predicates and properties in a correspondence way. Field's criticism of Tarski's theory can be appreciated from this perspective. According to Field, Tarski failed to meet his own stated policy: that he will not make use of any semantic concept if he is not able previously to reduce to other concepts."⁴⁸ This shouldn't be taken as implying that Tarski's definition of truth *explicitly* violated this policy. A definition of truth along the lines 5) explicitly violates Tarski's stated policy, since 5) employs semantic terms in the definition of truth.⁴⁹ Unlike 5), Tarski's definition

⁴⁷ It depends on how the notions of reference and standing for are further defined. If these notions are defined in a minimalistic way—in terms of implicit definitions involving infinite lists—then 5) is compatible with the deflationary notions of truth and reference. If the notions of reference and standing for are explicitly defined, 5) isn't compatible with the deflationary theory.

⁴⁸ H. Field, "Tarski's Theory of Truth," in his *Truth and the Absence of Fact*, Oxford: Clarendon Press, 2001, p.7.

⁴⁹ A definition of truth along the lines of 5) is not Tarski's definition of truth. Indeed, a definition of truth along the lines of 5) is what Field calls 'Tarski*s theory of truth'—the truth definition that Field calls 'T1'. In "Tarski's Theory of Truth," Field compares two truth definitions: T1 and T2. T2 is Tarski's theory of truth, and T1 is similar to Tarski's theory, but it explicitly employs semantic concepts such as 'denotes' and 'applies to' in the definition of

eliminates all the semantic terms in the definition of truth by replacing them with *extensionally equivalent* concepts. More precisely, Tarski succeeds in eliminating the notion of truth by defining it in terms of satisfaction and denotation, and then he eliminates the notion of denotation by giving extensionally correct definitions such as 1) and 2). Definitions such as 1) and 2) satisfy Tarski's conditions of material adequacy and formal correctness; they determine the correct extension of '*N* refers to *a*' in an object-language *L*; they are also formally correct since no semantic term is explicitly employed in their definiens. So, it appears that Tarski lived up to his stated policy; he eliminated all the semantic concepts by replacing them with extensionally equivalent concepts that don't explicitly involve semantic terms. So, why does Field say that Tarski failed to meet his own standard?

The reason, in short, is that to replace a concept with an extensionally equivalent concept isn't the same as *reducing* the concept into another concept. To argue for this point, Field appeals to an analogy to the concept of valence. Consider the following definition of the concept of valence:

- 6) $(\forall E) (\forall n) (E \text{ has a valence } n = C \text{ is potassium and } n \text{ is } +1, \text{ or } \dots \text{ or } E \text{ is sulphur and } n \text{ is } -2).$ ⁵⁰

This definition is extensionally correct, and thus it enables us to replace the concept of valence with its extensional equivalents. This definition, however, doesn't reduce a chemical fact into a non-chemical fact; i.e., a physicalistic fact of a more fundamental kind. But why is it necessary to reduce a chemical fact into a non-chemical/physicalistic fact? The rationale is derived from the idea of physicalism: the idea that chemical facts, biological facts, psychological facts, and

truth. Interestingly, Field says that Tarski should have offered the truth definition T1 instead of T2. Field's point is that T1 plus a physicalistic reduction of primitive denotation would constitute an ideal correspondence theory of truth.

semantical facts are all explicable in terms of physical facts.⁵¹ Field says that if 6) had been the only definition available for the concept of valence, scientists would have eventually had to give up the concept of valence.⁵² Fortunately, chemists could establish the reduction of the concept of valence to physical properties, and so the concept of valence was saved from being thrown away.

So, there are two reasons why Field demands a physicalistic reduction of semantic concepts, especially, the concept of denotation (or reference). First, Tarski said himself that he would not employ any previously *unreduced* semantic concept in his definition of truth. More importantly, *if* the notion of truth plays an explanatory role that is similar to that of the notion of valence, the notion of truth should be explicated in terms of physicalistic concepts. The notion of valence, being a chemical concept, enters into causal explanations of chemical elements. So, to say that the notion of truth plays the kind of explanatory role played by the notion of valence is to say that truth plays a causal-explanatory role. But if truth plays a causal-explanatory role, then there has to be an explanation of *how* it plays the role. Such an explanation requires a physicalistic reduction of the property of truth.

So, Field's criticism of Tarski's theory of truth may be best appreciated not as a criticism of Tarski's theory *per se* but as a criticism of a correspondence theory of truth of a Tarskian style *without* a theory of primitive denotation. Whether or not Tarski aimed to give a correspondence theory of truth, Tarski's theory isn't of itself a correspondence theory. The recursive technique that Tarski uses in his theory, however, offers an invaluable tool for the development of a correspondence theory. Field recognized this value, but he also recognized that a correspondence theory of a Tarskian style without a theory of primitive denotation cannot satisfy the demands of a correspondence theory of truth.

⁵¹ See H. Field, "Tarski's Theory of Truth."

⁵² H. Field, *ibid.*

1.3. The Motivations for the Deflationary Theory of Truth

Motivations for a theory are not the same as the arguments for it; the former usually alludes to the problems that the theory aims to solve, while the latter alludes to the reasons that support the theory. They sometimes, however, overlap with each other, and this is the case with one of the motivations for the deflationary theory.

There are at least two major motivations for the deflationary theory. One is that the prospect of constructing its major rivalry theory—the correspondence theory—is very dim. The other is that the role of truth, as Horwich says, isn't as it appears to be; the role of truth isn't what Field thought it was when he demanded a reductive/correspondence theory of truth in his 'Tarski's Theory of Truth.' According to deflationists, the role of truth is purely logical, and therefore, the theory of truth has only a minimal theoretical duty to satisfy—to explain what makes the concept of truth serve relevant logical functions. This means that if deflationists' view of the role of truth is correct, all the inflationary theories of truth—including the correspondence theory—are inherently wrongheaded.

The two motivations for deflationism are complementary. Anomalies accumulated in searching for an adequate correspondence theory may have inspired a new paradigm of the concept of truth. That said, they are concerned with significantly different problems with the correspondence theory of truth. The first motivation addresses a problem that arises within the correspondence theory of truth. In other words, it addresses a problem that arises when we accept the idea that truth is a correspondence notion. The second motivation, by contrast, challenges the very idea that truth is a correspondence notion. So, the second motivation stands even without the first motivation, although not vice versa. Despite so, I will discuss both motivations in the following.

Deflationism, as I understand, isn't just about the concept of truth but about a host of related concepts. The deflationary theory of truth can properly be appreciated only against the rich history of theories of meaning and content that goes back to Frege, Russell, and Wittgenstein. A discussion of the problems with the correspondence theory will hopefully shed some light on understanding the historical background of the deflationary theory of truth.

1.3.1. Motivation 1—Problems with Reductive Theories of Reference

Let's say that we accept the correspondence conception of truth. Let's also say that an adequate correspondence theory of truth should be developed in a Tarskian style for the reason that was explained above; that is, to eschew mysterious entities such as facts or states of affairs. In addition, let's suppose that Field is right in saying that a correspondence theory of truth needs a reductive theory of primitive denotation. Then, the next question to be answered is what it takes to give a reductive theory of reference. In this subsection, I will discuss the problems that arise in answering this question.

According to Devitt, there are at least three reductive/physicalistic theories of reference advanced in the 20th century; the first is Kripke-Putnam's causal-historical theory, the second is the informational theory of representation, and the third is the teleological theory of representation.⁵³ All these theories appeal to the notion of a causal relation between an expression (either a linguistic or mental expression) and an object in the world. Besides, they are chronologically and theoretically related to each other; in particular, the teleological theory is a

⁵³ M. Devitt, "Theories of Reference," *Encyclopedia of Cognitive Science*, Vol. 3. L. Nagel, ed. London: Nature Publishing Group (2003): 904-07.

development of the informational theory of representation. Each of these theories has many variants and has been subject to wild discussions. Since it is beyond capacity to examine the details of these theories, I will concentrate on one question that all these theories ultimately have to deal with; namely, the question whether any of these theories succeeds in justifying the truth of a disquotational reference-statement such as 1) below:

- 1) 'cat' (in English) refers to a cat.

Like Tarski's theory of truth, a reductive theory of reference aims to give an explicit definition of reference; it aims to define the notion of reference in terms of non-semantic terms in the style of (R^C) :

(R^C) x refers to $y =_{\text{Df}}$ x bears a non-semantic/physical relation R to y .

Given that a reductive theory of reference aims to offer an explicit definition of reference, we may speak of some conditions that these theories must satisfy. These conditions are parallel to the two conditions that Tarski imposes on a definition of truth; *formal correctness* and *material adequacy*. The formal correctness of a theory of reference is already characterized by Field; namely, a reductive theory of reference should not employ undefined semantic terms in its definiens. Hence, we have (R^C) above.

A reductive theory of reference, however, should also be materially adequate if it is to correctly determine the extension of the reference.⁵⁴ Tarski characterizes material adequacy for a theory of truth in terms of Contention T: that is, an adequate theory of truth for an object-language L should entail all the instances of (T) below:

(T) s is true in L iff p ,

where ' s ' is to be replaced by the quotation-name (structural description) of an object-language sentence s , and ' p ' is to be replaced by a meta-language sentence that means the same as ' s ' in an object-language L . What this implies, among others, is that a theory of truth for one's own language is materially adequate if all the instances of the schema (DS) below are derived from the theory:

(DS) ' p ' is true iff p ,

where ' p ' is replaced by an indicative sentence in my own language.⁵⁵ More precisely, a theory of truth for English satisfies Tarski's Contention T if it entails truth-statements such as "'Snow is white' is true in English iff snow is white", "'Grass is green' is true in English iff grass is green", and so forth.

⁵⁴ According to Tarski, a definition of truth for a language L is materially adequate if and only if it correctly determines the extension of true sentences. Put differently, a definition of truth of the form below

For all x , $True(x)$ if and only if $\phi(x)$

is materially adequate if and only if objects satisfying ϕ are the objects that we would intuitively count as being true sentences in L .

⁵⁵ Let's here assume that the sentence that replaces ' p ' doesn't involve an ambiguity, vagueness, indexical, or demonstrative.

Parallel to Tarski's Contention T, we may say that a reductive theory of reference for an object-language L is materially adequate if all the instances of the schema (DR) below are derived from the theory:

(DR) (If r exists) ' r ' refers to r and nothing else.⁵⁶

So, Contention R, parallel to Convention T, may be characterized as follows:

Convention R: All the legitimate instances of the schema (DR) are to be derived from the definition of reference of the form below:

For all x , $Refers(x)$ iff $\phi(x)$.

If a reductive theory of reference is to be adequate, it has to satisfy Contention R. The question is whether there is any reductive theory that satisfies Contention R.

I used to think, among those three variants of the causal theory of reference, that Kripke-Putnam's theory has the least chance of satisfying Convention R. But now I think the opposite; Kripke-Putnam's causal theory seems to have the best chance of satisfying Convention R *if the theory is true*. In order for a reductive theory of reference to satisfy Contention R, it has to cope with Fodor's disjunction problem. As I will argue below, it is very unlikely that either the causal-information theory or the teleological theory of reference will ever be able to cope with Fodor's disjunction problem. Kripke-Putnam's theory, if it turns out to be true, can solve Fodor's disjunction problem. The theory, however, isn't true. So, none of the three variants of the causal

⁵⁶ An assumption here is that an expression substituting for ' r ' is an expression as I understand it.

theory of truth can satisfy Contention R. Therefore, none of the causal theories of reference can be considered an adequate theory of reference. This is what I will argue in the following.

Fodor's Disjunction Problem, and the Causal-Informational and Teleological Theories⁵⁷

Fodor initially raised the disjunction problem for the causal-informational theory of representation. So, to explain what the problem is, I will briefly explain what the causal-informational theory of representation is.

The causal-information theory is a physicalistic theory of mental representation that reduces the semantic/intentional content of a mental representation to an external condition that is reliably correlated with the mental representation. This theory is based on the perfectly plausible idea that *A* carries information about *B* if *A* is reliably caused by *B*; for example, smoke carries the information that there is a fire because smoke is reliably caused by a fire. The causal-informational theory is the result of applying this idea to an account of the semantic/intentional content of mental representations. Thus, the core idea of the causal-informational theory is that the semantic or intentional content of a mental representation *R* reduces to the type of external conditions that reliably cause the tokens of the mental representation *R*.

According to Fodor, however, this theory faces a serious problem—the “disjunction problem”.⁵⁸ This problem arises because a mental representation of the same type can be reliably caused by external conditions of distinct types. For example, tokens of ‘cat’ are caused not only by cats but also by cat-like dog. But if the semantic or intentional content of a mental representation reduces to external conditions that reliably cause the tokens of the mental representation, on what ground

⁵⁷ References for this discussion are: J. Fodor, “Semantics, Wisconsin Style,” “A Theory of Content I: The Problem,” and “A Theory of Content II: The Theory,” in J. Fodor, *A Theory of Content and Other Essays*, Cambridge, Mass: MIT Press, 1990.

⁵⁸ See J. Fodor, *ibid.*

should we say that a token of 'cat' is about a cat and not about a cat-like dog? Given that the tokens of 'cat' are reliably caused both by cats and cat-like dogs, a token of 'cat' carries both the information that there is a cat around and the information that there is a catlike dog around. In order to say that a token of 'cat' is about or represents a cat, and not a cat-like dog, we should be able to select the causal relation between a token of 'cat' and a cat as *the* semantically and intentionally relevant relation. The causal-informational theory, however, is helpless in this respect. The theory cannot explain what makes the causal relation between the tokens of 'cat' and cats *the* semantically and intentionally relevant relation.

One may ask here why we have to select only one of the causal relations as the semantically and intentionally relevant relation. Why can't we say that a token of 'cat' is about a cat or a cat-like dog? Why can't we allow a mental or linguistic representation have disjunctive content? As was mentioned earlier, the core idea of the causal informational theory is derived from the relationship between the information carried by a natural sign and external conditions that reliably cause the occurrence of the natural signs. With respect to natural signs, however, we have no problem with the multiplicity of informational content. For example, smoke indicates not only that there is a fire but that there is oxygen. But here, unlike in the case of 'cat', we don't have to choose one of them as *the* information carried by the natural sign smoke. So, why is the multiplicity of information a problem only when it occurs with regard to a mental or linguistic representation? The most convincing answer that I can think of is a *pragmatic* one; that is, it is simply too inconvenient to allow disjunctive content. Suppose my friend visited a remote island and called me saying 'There is no cat in this island'. Suppose I think what he is saying is true. If we allow an expression to have disjunctive content, then the information that I can obtain from what he uttered would be as follows: Either there is no cat or cat-like dog in the island. To make

matters worse, let's say that my friend also said 'But there is a dog in this island'. But then, the information that I infer from what my friend said would be: Either there is no cat or catlike dog, but there is a dog in the island. There is no logical contradiction here, and we might even give a more accurate characterization of the semantic content of my friend's utterance by allowing disjunctive content. But to allow disjunctive content isn't a good idea for the pragmatic purposes of ordinary conversation.

At this point, the deflationist will immediately step in claiming that this is the reason why it is a wrong idea to reduce the notions of truth and reference into physical notions. Those who try to reduce the notions of truth and reference, says the deflationist, got the role of these notions entirely wrong. But not everyone agrees with the deflationist's analysis of the role of truth and reference. So, those who try to explicate the notions of truth and reference in terms of physicalistic notions developed the causal-informational theory into a more sophisticated theory—the teleological theory of intentionality. Fodor was himself drawn to this approach, but he later abandons it.⁵⁹ In this thesis, I will not discuss the details of the teleological theory.⁶⁰ I will point out, though, one crucial problem with the theory.

According to the teleological theory, we can isolate the causal relation between 'cat' and cats from the causal relation between 'cat' and catlike-dogs by appealing to the teleological function served by the causal relation between 'cat' and cats. So, the core idea of the teleological theory is that there is a special function—biological function—served only by a semantically relevant

⁵⁹ See J. Fodor, "A Theory of Content II: The Theory," in J. Fodor, *A Theory of Content and Other Essays*, Cambridge, Mass: The MIT Press, 1990.

⁶⁰ There are many variants of the teleological theory of representation, and it is beyond the scope of this thesis to discuss the problems with each of these variants. Teleological theories that I have studied are confined to those of Dretske and Papineau. With such limited knowledge of the theory, I cannot give an overall discussion of the teleological theory. See F. Dretske, *Explaining Behavior: Reasons in the World of Causes*, The MIT Press, 1991. D. Papineau, *Reality and Representation*, Blackwell, 1987.

causal relation, and we can reduce the truth-referential (intentional or representational) relation to the causal relation that has this special biological function. As Godfrey-Smith has points out, however, this theory has a serious problem when it is applied to a particular state of a mechanism.⁶¹ It makes sense to say that a mechanism has a function; for example, our belief- and desire-mechanism must have a specific biological function. It doesn't make sense, however, to say that a belief- or desire-state of each type has a distinct biological function or purpose. Let's assume, for arguments' sake, a sort of language of thought hypothesis: believing that p is believing* a mental sentence δ means that p .⁶² According to the teleological theory, the semantic or intentional relation between δ and the external condition that p reduces to a causal relation that has a special biological function. More specifically, on this theory, what makes δ have the representational/intentional/semantic/truth-conditional content that p is that an organism has a better chance of surviving when the tokens of its mental sentence δ are caused by the external condition that p than any other external conditions. In this sense, the external condition that p is defined as the "normal cause" of the mental sentence δ , and the normal cause of a mental symbol determines its semantic/intentional content.

So, according to this theory, a token of 'cat' refers to a cat, not a catlike-dog, because the causal relation between the tokens of 'cat' and cats has a special biological function that is lacked by other causal relations such as the one between the tokens of 'cat' and catlike-dogs. But what is this special biological function that can be served only by a semantically or intentionally relevant causal relation? If the nature of this special function is vaguely characterized such as 'contributing to the survival of an organism', then the teleological theorist cannot isolate the

⁶¹ P. Godfrey-Smith, "Misinformation," *Canadian Journal of Philosophy*, Vol.19, No. 4, 1989, pp. 533-550.

⁶² Here, a believing*-relation is a relation to a neural sentence, and it should be distinguished from an ordinary believing-relation. The idea is to explain an ordinary believing-relation in terms of a physicalistic believing*-relation. See J. Fodor, *The Language of Thought*, Cambridge, Mass: Harvard University Press, 1975; H. Field, "Mental Representation," (1978) in his *Truth and the Absence of Fact*, Oxford; Clarendon Press, 2001.

causal correlation between ‘cat’ and cats from other causal correlations. There is simply no reason to believe that for each mental symbol there is a causal relation of a certain type such that the occurrences only of that type contribute to the survival of an organism. For example, a frog’s mental symbol δ can be tokened either by the presence of a fly or a flylike mosquito. According to the teleological theory, however, the presence of a fly is the normal cause of the frog’s mental symbol δ because the causal relation between the tokens of δ and flies has a special biological function that cannot be served by the causal relation between the tokens of δ and fly-like mosquitoes. There is, however, little evidence to think this way, because the frog’s biological purpose can be equally well served by the frog’s actions that are performed when the tokens of δ are caused by the presence of fly-like mosquitoes.

The problem here is that an advocate of the teleological theory should spell out the nature of the special function without appealing to a semantic or intentional notion. To see why this is important, consider Papineau’s solution to the disjunction problem. According to Papineau, the disjunction problem can be solved only when the content of a belief-state is considered in liaison with a desire-state. Papineau says,

[B]iological functions are always a matter of effects; functions are *effects* in virtue of which traits are selected. Beliefs, however, do not have any effect to call their own. Rather, their biological purpose is to produce whichever result will fulfill the purposes of the desires they are acting in concert with. And then it will then explain the contents of beliefs derivatively, by saying that the purpose of beliefs is to generate actions that will produce desired results in such-and-such conditions, which conditions will therefore counts as the truth conditions of those beliefs.⁶³

Papineau is right in saying that the content of a belief-state and the content of a desire-state are interdependent. He is wrong, however, in saying that the content of a belief-state can be *derivatively* determined by the content of a desire-state. The moment we determine the content of

⁶³ D. Papineau, “Teleosemantics and Indeterminacy,” *Australian Journal of Philosophy*, Vol. 76, Issue 1, 1998. P.8.

a desire-state so do we determine the content of a belief-state. That is to say, to determine the content of a belief *is* of itself to determine the content of a desire-state..

So, Papineau's solution to the disjunction problem assumes the very thing that the solution is supposed to answer—on what ground should we say that the frog *desires that it wants to eat a fly* instead of wanting to eat a fly-like mosquito? Why can't we say, as Neander claims, that the frog desires that it wants to eat a small, dark, and moving insect?⁶⁴ Papineau claims that Neander begs the question by characterizing the frog's desire-state in this way. But the burden of proof is on Papineau's claim, not on Neander's claim. It is Papineau who claims that the frog's desire state has a determinate content, and Neander's job is just to offer a reason to doubt it. Her reason is that we can characterize the frog's desire-state in a way that doesn't invoke a fly. A small, dark, and moving insect can be a fly or flylike mosquito; therefore, the content of the frog's desire-state is indeterminate. Now, it is Papineau's turn to explain why it is necessary that we invoke a fly in characterizing the frog's desire-state. Papineau, however, doesn't seem to have an explanation. Instead, he makes a dubious claim; that content becomes determinate if we move away from frogs toward beings with more complicated psychologies.⁶⁵ But this is a naïve view; there is no fundamental difference between a frog's psychology and a human's psychology. If there were, the teleological theory wouldn't have got off the ground in the first place.

Fodor's Asymmetric Dependence

Before I discuss Kripke-Putnam's theory, I will discuss briefly Fodor's own solution to the disjunction problem: the theory of asymmetric dependence. Interestingly indeed, there are some similarities between Fodor's theory of asymmetric dependence and Kripke-Putnam's causal

⁶⁴ See D. Papineau, *ibid.* K. Neander, "Misrepresenting and Malfunctioning," *Philosophical Studies* 79 (2), 1995, pp. 109-141.

⁶⁵ D. Papineau, *ibid.*

theory of reference. Both theories heavily rely on the conceptual fundamentality of the instances of the disquotational reference (DR):

(DR) (If r exists) ‘ r ’ refers to r and nothing else.

No physicalist including the advocates of the causal-informational theory and teleological theory would deny that the instances of (DR) are true. But the causal-informational theory or the teleological theory wasn’t designed to explain the instances of (DR). The motivations of these theories come from somewhere else, and to accommodate the instances of (DR) is just one of the desiderata of these theories. Fodor’s theory and Kripke-Putnam’s theory, however, *are* designed to accommodate the instances of (DR). So, if these theories are true, the disjunction problem will be solved. The problem, however, is that it is doubtful that any of these theories is *actually* true.⁶⁶

First, according to Fodor’s theory of asymmetric dependence, although tokens of ‘cat’ can be caused by non-cats, a token of ‘cat’ refers to a cat, not a catlike dog, because the causal connection between the tokens of ‘cat’ and the property of being a cat is metaphysically fundamental. That is to say, the causal connection between the tokens of ‘cat’ and the property of being a catlike dog is asymmetrically dependent on the causal connection between the tokens of ‘cat’ and the property of being a cat.⁶⁷ The crucial notion here is ‘asymmetric dependence’, and Fodor explicates this notion in terms of a subjunctive conditional: if the nomic (lawlike) relation between $P1$ and $P2$ is asymmetrically dependent on the nomic relation between $P3$ and $P4$, then

⁶⁶ In this sense, Devitt says that Fodor’s theory of asymmetric dependence is a *possible* explanation, but there is a question whether this is the *actual* explanation. See M. Devitt, “Review: Naturalistic Explanation,” *The British Journal for the Philosophy of Science*, 1991, pp. 425-443.

⁶⁷ J. Fodor, *A Theory of Content and Other Essays*, MIT Press, 1992.

ceteris paribus, breaking the relation between *P3* and *P4* would break the relation between *P1* and *P2*.⁶⁸ To elaborate it as Block does, ‘cat’ refers to a cat if and only if although there are nomologically possible worlds in which cats cause ‘cat’ but catlike dogs don’t, there are no nomologically possible worlds in which catlike dogs cause ‘cat but cats don’t.’⁶⁹ There are many complicated questions to be discussed about this theory, but the most important question seems the following: why do we have to believe that these counterfactuals are true? Fodor asks this question himself, and his own answer, as Devitt says, is strange:

Don’t forget, this stuff is supposed to be philosophy. In particular, it’s an attempt to solve Brentano’s problem by showing that there are naturalistically specifiable, and atomistic, and sufficient conditions for a physical state to have an intentional content. In *that* context, I get to stipulate the counterfactuals. . . . That is, solving Brentano’s problem requires giving sufficient conditions for intentionality, not sufficient and necessary conditions.⁷⁰

Philosophy, however, is concerned not only with the validity of an argument but also with the soundness of an argument. In order for Fodor’s argument to be sound, the antecedent of his argument should be true. The antecedent of the argument is Fodor’s theory of asymmetric dependence. So, the question re-arises: why do we have to believe that the counterfactuals being invoked are true? To say that the relevant counterfactuals are true entails that there is no nomologically possible world in which ‘cat’-tokens are caused by catlike dogs while ‘cat’-tokens are not caused by cats. There seems, however, that there *is* such a possible world. All we need is a possible world in which the property that causes the tokens of ‘cat’ is nomologically connected to the property of being a non-cat but isn’t nomologically connected to the property of being a cat. A good example is Putnam’s Twin-Earth.

⁶⁸ J. Fodor, *ibid.*, p. 95.

⁶⁹ J. Fodor, *ibid.*, p. 111.

⁷⁰ J. Fodor, *ibid.*, 96.

In Putnam's Twin-Earth example, my doppelganger lives in the twin-earth in which the stuff that looks like water is of a different chemical structure—XYZ. By assumption, my doppelganger is in the brain-states that are type-identical to my brain-states. So, she acts and talks like me; she even uses the (phonetically and typographically) same word 'water' to pick out the stuff that looks like water. Despite so, it is obvious that the reference of my doppelganger's word 'water' is different from the reference of my word 'water'; any sane ordinary person would agree with this. So, the reference isn't determined by what is in the head. Meaning and reference, says Putnam, aren't just in the head no matter how we cut the pie.⁷¹

I will discuss Putnam's theory of reference shortly. But let me first explain why this example repudiates Fodor's theory of asymmetric dependence. In this example, my doppelganger is in the same brain-state that I am in; we are molecule-by-molecule duplicates. So, the property that causes the tokens of my doppelganger's 'water' in the twin-earth is the same as the property that causes the tokens of my word 'water' in the earth.⁷² In the twin-earth, however, there is no H₂O. So, in this world, the same brain-state (or brain-property) is caused by XYZ although it isn't caused by H₂O. But according to Fodor's theory, for each brain-state *B*, there has to be one and only one type of an external condition *E* to which *B* bears a special causal connection. What makes this causal relation special is that the brain-state *B* wouldn't be caused by other external conditions unless it is caused by the external condition *E*. As was stated above, what this means is that there is no possible world in which *B* is caused by an external condition *E** but isn't caused by the external condition *E* (where $E \neq E^*$), although there is a possible world in which *B*

⁷¹ H. Putnam, "The Meaning of 'Meaning'," *Minnesota Studies in the Philosophy of Science* 7: 131-193, 1975; H. Putnam, *Mind, Language, and Reality*: Philosophical Papers Vol. 2, Cambridge: Cambridge University Press, 1975.

⁷² According to Fodor, the property that causes the tokens of an expression-type is the property of a brain-state. So, the causal relation between the tokens of an expression-type and external objects is the relation between properties—the relation between the property of a brain-state that causes the tokens of an expression-type and the property of being a certain entity; e.g. being a horse, being water, being a cat, etc.

is caused by E but not caused by E^* . Putnam's twin-earth argument, however, offers a counter-example to this theory. Putnam's twin-earth example shows that there is simply no special external condition E to which a brain-state B bears a special causal connection. In this world, the brain state that causes the tokens of 'water' is caused by H_2O but it isn't caused by XYZ . In the twin-earth, the same brain-state is caused by XYZ while it isn't caused by H_2O . If there is a third possible world, it may be caused by another external property. This seems enough to show that Fodor's thesis is incorrect; there is no metaphysically special causal relation on which other causal connections are dependent. The causal relation between the tokens of 'water' and XYZ isn't metaphysically dependent on the causal relation between the tokens of 'water' and H_2O . Fodor's theory doesn't give any reason to think that the causal relation between the tokens of 'cat' and cats is metaphysically fundamental in such a way that other causal relations are dependent on it. But if we cannot single out the causal relation between the tokens of 'cat' and cats as metaphysically fundamental, Fodor's theory cannot solve the disjunction problem.

Kripke-Putnam's Causal Theory of Reference

Now, let's consider whether Kripke-Putnam's causal historical theory of reference can help solving the disjunction problem. As I stated above, if Kripke-Putnam's causal historical theory is correct, it can. What I mean by this conditional statement is that if proper names and natural-kind terms are *rigid designators* as they are claimed by Kripke-Putnam's theory, the disjunction problem will be solved. But then, the question is whether proper names and natural-kind terms are rigid designators.

First, let's examine what it means to say that an expression is a rigid designator. According to Kripke, to say that an expression is a rigid designator is to say that it has the same reference or

extension in every (relevant) possible world. More specifically, according to this theory, a proper name such as ‘Aristotle’ is a rigid designator because it refers to one and the same person in all possible worlds in which Aristotle exists. A natural-kind term such as ‘water’ is also construed as a rigid designator because ‘water’, according to this theory, determines the same extension in every possible world in which water exists. Some definite descriptions are also construed as rigid designators although most definite descriptions are not; for example, the number succeeding 2’ refers to 3 in every possible world, and so, it is a rigid designator.

This simple definition of a rigid designator, however, raises many complicated questions. First, what is exactly referred to by an expression that is characterized as a ‘rigid designator’?

According to this theory, a proper name and a natural-kind term are rigid designators, and therefore, they refer to the same entity in every possible world. But what is exactly referred to by a proper name or natural-kind term? Second, what is the nature of the semantic relation between a rigid designator and its reference? What makes a proper name or a natural-kind term refer to whatever entity that it refers to? These two questions are related, since we cannot determine the reference of a rigid designator unless we can characterize the nature of the semantic relation between a rigid designator and its reference. Notice that the conception of reference advertised by Kripke-Putnam’s theory is by no means deflationary; in Kripke-Putnam’s theory, an instance of (DR)

(DR) (If r exists) ‘ r ’ refers to r

isn’t considered true a priori. The conception of reference advertised by Kripke-Putnam’s theory is inflationary, and so, it doesn’t consider the trivialities of the instances of (DR)—‘Aristotle’

refers to Aristotle, ‘water’ refers to water, etc.—as constituting the whole story of a theory of reference.⁷³ Moreover, the idea of reference in Kripke-Putnam’s theory is very specific; namely, a proper name or a natural-kind term is considered as referring to *that very specific thing in the actual world*. As will be discussed, however, Kripke-Putnam’s theory never offers a clear explanation of what that specific thing is. Considering what is advertised by this theory, what this theory really offers is very meager. I will explain why.

First, according to Kripke-Putnam’s theory, the entity that is referred to by a rigid designator is what is related to the rigid designator in terms of an actual causal history. An actual causal history is composed of two steps: the reference-fixing step and the reference-transferring step. What is crucial in the reference-fixing step is a non-descriptive way of referring to an entity. A typical example of a non-descriptive way of referring to an entity is direct or indirect (deferred) ostension.⁷⁴ According to Kripke-Putnam’s theory, the reference of a proper name is fixed by direct ostension; the reference of a proper name such as ‘Aristotle’ is fixed when someone directly points at a certain baby dubbing him ‘Aristotle’. Once the reference of a proper name is fixed by direct ostension, the semantic relationship between the name and its reference has to be properly transferred to the other members of a linguistic community. Indirect or deferred ostension is involved in fixing the reference of a natural-kind term. For example, when someone

⁷³ According to Horwich, any theory of reference that doesn’t consider the trivialities of the instances of (DR) as constituting the whole story of a theory of reference is to be characterized as inflationary. See P. Horwich, *Meaning*, p. 118.

⁷⁴ The difference between direct ostension and indirect (deferred) ostension is that whereas the reference of an expression is included in the activity of pointing in direct ostension, the reference of an expression isn’t included in the activity of pointing in indirect (deferred) ostension. The kind of ostension involved in the determination of the reference of a natural-kind term in Kripke-Putnam’s theory is indirect (deferred) ostension. For example, the reference of ‘water’ isn’t directly contained in the activity of pointing. What is directly pointed at is a particular instance of water, and the natural kind *water* or H_2O is indirectly picked out via the activity of pointing at the particular instance. See W. V. Quine, “Ontological Relativity”, in his *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1969.

first introduced the term 'water' by pointing at a particular instance, the reference of 'water' was fixed by deferred ostension; 'water' refers to the natural kind that the particular instance belongs to. In explaining the reference of a natural-kind term Kripke and Putnam also appeal to the view called 'scientific essentialism'; the view that that the essential properties of natural kinds are to be discovered a posteriori by scientific investigations. For example, we couldn't have known a priori that water is H₂O. So, if the current theory of water is true, then an identity-statement such as 'Water is H₂O' is metaphysically necessary but epistemologically a posteriori.

It should be noted, however, that although both Kripke and Putnam endorse scientific essentialism, scientific essentialism is orthogonal to their theory of reference. There are two reasons for this; first, as Kripke admits himself, any scientific theory may turn out to be false. For example, water may not turn out to be H₂O. But then, we cannot rely on scientific theories to connect a natural-kind term to its reference. If we did, we could no longer consider a natural-kind term a rigid designator; the reference of a natural-kind term would keep changing with the vicissitudes of scientific theories. Another reason that scientific essentialism isn't central to Kripke-Putnam's theory of reference is that this theory was motivated to refute Frege-Russell's description theory of reference. One of Kripke's major complaints of the description theory is that if the references of a proper name and a natural-kind term were determined via satisfying definite descriptions, then the references of these expressions would vary from world to world. But if we turn to an essential property of a natural kind in picking out the reference of a natural-kind term, we would have to treat the natural-kind term as an abbreviation of a definite description. Suppose we pick out the reference of 'water' by investigating its chemical structure. In such a case, we would have to consider anything that satisfies 'the stuff that is H₂O' to be the reference of 'water'. There is no guarantee, however, that the stuff that satisfies this definite

description in a possible world is the same stuff that satisfies the definite description in the actual world. Therefore, scientific essentialism isn't central to Kripke-Putnam's theory of reference; to the contrary, it may result in a consequence that is contradictory to the idea that proper names and natural-kind terms are rigid designators.

What is central instead to Kripke-Putnam's theory of reference is the idea that the reference of a rigid designator is fixed by direct or indirect (deferred) ostension. Kripke makes this point very clear with the following consideration. Suppose a cat doesn't turn out to be an animal but turns out to be an automaton. In such a case, says Kripke, the correct thing to say isn't that a cat doesn't exist but that a cat isn't an animal as we originally thought it was. Kripke says,

The original concept of cat is: *that kind of thing*, where the kind can be identified by paradigmatic instances. It is not something picked out by any qualitative dictionary definitions.⁷⁵

But how can direct or indirect (deferred) ostension fix the references of a proper name and a natural-kind term in a way that can make them rigid designators? Kripke-Putnam's idea is that a proper name and a natural-kind term rigidly designate the entities that they refer to in the *actual* world. But if the reference-relation in the actual world is to be explicated in terms of direct or indirect ostension, the reference of a proper name or a natural-kind term would be highly indeterminate. One can refer to many different things by direct or indirect ostension. The only way to solve this problem, as Kripke admits himself, is to appeal to a speaker's intention. But by appealing to a speaker's intention, we would give up much of Kripke-Putnam's theory of reference. There is simply no way of appealing to a speaker's intention without associating a proper name or natural-kind term with definite descriptions; we cannot articulate a speaker's intention merely by invoking direct or indirect ostension. In the passage quoted above, Kripke

⁷⁵ S. Kripke, *Naming and Necessity*, Lecture III, Cambridge, Mass: Harvard University Press, 1972.

says that the original concept of cat is *that kind of thing*. But *that kind of thing* is almost an empty concept. What is the reference of ‘water’? According to Kripke-Putnam’s theory, it is also *that kind of thing*, since reference cannot be picked out by any “qualitative dictionary definitions”. The semantic machinery provided by Kripke-Putnam’s theory is so barren that it seems almost meaningless to say that a proper name or natural-kind term is a rigid designator. With such a skimpy explanation, we have little idea of what an expression refers to even in the actual world. But then, to say that an expression refers to the same entity in every possible world seems of little semantic importance.

The initial question that drove us here is whether a causal theory of reference can provide a reductive account of the reference relation; the relation between ‘Aristotle’ and Aristotle, ‘water’ and water, ‘cat’ and the class of cats, etc. None of the causal theories of reference that have been examined—the causal-informational theory, the teleological theory, Fodor’s asymmetric dependence theory, and Kripke-Putnam’s causal-historical theory—fills the bill. This raises the doubt for the sustainability of the correspondence theory of truth. According to Marian David, one of the chief motivations for the deflationary theory is “that no physicalistic explanation of semantic and intentional notions—a fortiori no physicalistic explanation of representation and similar relations—is forthcoming.”⁷⁶ But why is it such a big problem? Because without a reductive theory of reference the correspondence theory would fail *internally*. The correspondence theory of truth for sentences stems from the idea that truth is a dyadic relation, and as such the theory is intended to provide an account of the nature of the semantic relation between indicative sentences and external conditions. Since the number of indicative sentences is infinite, such an account should be given in a systematic/recursive way—via an

⁷⁶ M. David, *Correspondence and Disquotation: An Essay on the Nature of Truth*, Oxford University Press, 1994. P. 58.

account of the nature of the semantic relation between the constituents of indicative sentences and external conditions. This is the reason that a physicalistic explanation of the notion of reference is required by the correspondence theory of truth. Unfortunately, no physicalistic explanation of reference, and this motivates a search for an alternative theory of truth; i.e., the deflationary theory of truth.

This internal problem of the correspondence theory, however, stirred even a greater suspicion about the correspondence theory of truth; that this theory may be based on a wrong conception of truth. This is the second motivation of the deflationary theory of truth.

1.3.2. The Motivation II—The Role of Truth

Devitt says that one of the major differences between the correspondence theory and the deflationary theory of truth is that whereas the focus of the correspondence theory is on the nature and the role of *truth*, the focus of the deflationary theory is on the nature and the role of *the truth term*; for example, of ‘true’.⁷⁷ What Devitt says here tells us something important about the deflationary theory of truth. To see why let’s take a look at what Horwich says about the problems with the inflationary theory of truth including the correspondence theory. Horwich states that the inflationary theory of truth is based on two related misconceptions: first, truth has some hidden structure that needs to be discovered; and secondly, we can explain central principles employing the term ‘true’, and thereby solve a host of problems in logic, semantics, and epistemology.⁷⁸

⁷⁷ M. Devitt, “The Metaphysics of Truth,” in *The Nature of Truth*, ed. Michael Lynch, Cambridge, MA: MIT Press, 2001, 579-611.

⁷⁸ P. Horwich, *Truth*, 2nd edition, Oxford: Blackwell, 1999. P.2.

According to Horwich, however, such an idea is based on the misconception of the truth term, and this misconception is caused by an incorrect application of a *linguistic analogy*. Predicates such as ‘is magnetic’ and ‘is diabetic’ ascribe substantive properties to their bearers in the sense that the properties ascribed by these predicate are real ingredients of the world. Since ‘is true’ is linguistically similar to these substantive predicates, it has traditionally thought thought that the truth predicate also ascribes a substantive property to its bearers.

If we go beneath the surface, however, we find that the truth predicate serves a very different function. Despite its surface structure, the function of the truth predicate is not to ascribe a substantive property to its bearers. The function of the truth predicate is simply “to permit the formulation of generalizations (including laws of logic) that would otherwise call for substitutional quantification, or some alternative device of ‘infinite conjunction.’”⁷⁹ This idea—or some ideas similar to it—was already defended by those philosophers whose views of truth sowed the seeds for contemporary deflationism. For example, Ramsey’s redundancy theory of truth (or “no-theory” of truth), Strawson’s performative theory of truth, the prosentential theory of truth developed by Belnap, Camp, and Grover, and last but not least Quine’s thesis of truth as a device of semantic ascent offered dramatically deflated analyses of the function of truth.⁸⁰ As Devitt points out, they all focused on *the truth term*, not *truth* itself; more precisely, their attitudes are that the nature of truth can only be revealed through the analysis of the nature and the role of *the truth term*. This attitude is also adopted by contemporary deflationists such as Horwich, Field, Williams, Leeds, Hill, and so forth. In this sense, it may be said that the

⁷⁹ P. Horwich, “The Essence of Expressivism,” *Analysis*, Vol. 54. No 1.1994. P.19.

⁸⁰ See F. Ramsey, “Facts and Propositions,”(1927) in *Philosophical Paper*, ed. D. H. Mellor, Cambridge University Press, 1990; P. F. Strawson, “Truth,” *Aristotelian Society Supplementary*, 1950, Vo. 24; J. Camp, D. Grover, and N. Belnap, “A Prosentential Theory of Truth,” *Philosophical Studies* 27, 1975, pp. 73-125; w. v. Quine, W. V. Quine, *Philosophy of Logic*, Eaglewood Cliffs: Prentice-Hall, 1970.

deflationary theory of truth is the result of applying the time-honored philosophical method that has been employed by analytically-minded philosophers to truth terms.⁸¹

So, the deflationist focuses on the question of how we use the truth term or what we use the truth term for. The deflationist's answer is that we use the truth term simply as "ersatz" for substitutional quantifiers or infinite conjunctions (or disjunctions). In other words, the purpose for which we employ the truth predicate is "purely logical". For example, when we want to endorse, agree, or affirm what someone stated, but we don't remember exactly the sentence he uttered, we employ the truth predicate: e.g. 'What Oscar said is true'. We also employ the truth predicate when we want to generalize an infinite (or a finite) of sentences or propositions. For example, instead of stating that if John is tall, then John is tall, if Mark is a man, then Mark is a man, and so forth, we state that any statement of the form "If P, then P" is true. Or we may employ the truth predicate when we want to say that a theory is true when we cannot state each and every statement that constitutes the theory. For example, when we state that Einstein's theory is true while we don't know each and every statement constituting his theory, we employ the truth predicate as an ersatz for substitutional quantification or any alternative device enabling us to express the generalization.

In each of these cases, we don't employ the truth predicate to ascribe a substantive property to a statement, proposition, or a sentence. For example, when we state that what Oscar said is true, we don't ascribe to the sentence that Oscar uttered a property that is part of the world. This is the difference between the truth predicate and other predicates such as 'is magnetic' and 'is diabetic', since when we employ the latter predicates, we ascribe properties that are part of the world to the bearers of these predicates. For example, when we state a sentence of the form

⁸¹ Horwich says that an expression might have a meaning that is somewhat disguised by its superficial form. I suppose a similar idea was behind Russell and Wittgenstein's search for the "logical" form of language.

1) X is magnetic,

we ascribe to X the property of being magnetic that is part of the physical world. On the other hand, when we state a sentence of the form

2) X is true,

we don't ascribe to X (a sentence, statement, or proposition) a property that is part of the physical world. Superficially, both 2) and 3) have the linguistic form 3) as follows:

3) X is F .

The property of truth, however, is an entirely different kind from the property of being magnetic; whereas being magnetic is a physical property, being true or truth isn't. Truth is a purely logical property in the sense that it is a *purely inferential property*. The only function served by 'is true' in a statement of the form 2) is that it enables us to infer X itself. So, if we know that

4) $X =$ 'Snow is white' (or $X =$ the proposition that snow is white)

we can infer 5) from 2):

5) Snow is white.⁸²

Of course, the term ‘is magnetic’ also has an inferential function; we can infer from 1) a statement of the form 6):

6) *X* is not a plastic.

The question, however, is what makes these inferences possible. In order to infer 6) from 1), we need to make an empirical investigation of the nature of the property of being magnetic; we need empirical tests and observations to connect the property of being magnetic to the property of not being a plastic. That’s not the case, however, with the inferential relation between 2) and 5). In order to infer 5) from 2), we only need to understand English and the non-theoretical identification 4). None of them requires an empirical investigation or a theoretical justification. Anyone who understands English would agree that ‘Snow is white’ is true iff snow is white; ‘Grass is green’ is true iff grass is green; and so forth. Put simply, whereas the inference of 6) from 1) is made possible by a physical theory of magnetism, the inference of 5) from 2) is made possible by infinite conjunctions such as 7) and 8) that we trivially accept:

7) ‘Snow is white’ is true iff snow is white; ‘Grass is green’ is true iff grass is green; and so forth.

8) The proposition that snow is white is true iff snow is white; the proposition that grass is green is true iff grass is green; and so forth.

⁸² As was stated earlier, the deflationary theory of truth can be characterized for a sentence and a proposition. I here take account of these two possible versions of the deflationary theory of truth.

So, one of the two core theses of the deflationary theory is that what we mean by a truth-statement of the form 1)— X is true—is exhausted by 1*) in the following:

1*) If $X =$ ‘Snow is white’ (or the proposition that snow is white), then snow is white; if $X =$ ‘Grass is green’ (or the proposition that grass is green), then grass is green; and so on.

We cannot generalize 1*) in the usual way by invoking objectual quantification, since objectual quantification doesn’t permit quantifying into expressions that occur inside quotation marks. We may appeal to substitutional quantification, but even if we did, we would still want to express the generalization in ordinary language. The truth predicate serves this purpose with the help of some ancillary devices; e.g. quotation marks or a nominalizing phrase (‘the proposition that’). As Quine says, to express the generalization of 1*), we engage in semantic ascent by using ancillary devices while we use the truth predicate to cancel off the effect of semantic ascent. Moreover, what enables the truth predicate to serve this purpose is our willingness to accept the instances of the disquotational schema (DS) and the equivalence schema (E):

(DS) ‘ p ’ is true iff p .

(E) $\langle P \rangle$ is true iff P .

The instances of these schemas are epistemologically (or conceptually) fundamental; we accept them trivially or a priori. Besides, truth has no other role to play; the sole purpose of truth consists in facilitating generalizations. So, no inflated theory—a theory of truth that says more

than what is stated by an instance of (DS) or (E)—is needed. This is the second core thesis of the deflationary theory of truth.

Presented this way, what is stated by the deflationary theory is so thin that one may wonder how such an impoverished theory can produce a fruitful discussion. When it comes to deflationism, however, a fruitful discussion isn't so much generated by what is *stated* by the theory as what is *denied* by the theory. Since the deflationary theory of truth isn't of one kind, what is denied by the deflationary varies from theory to theory. That said, there are some claims commonly denied by deflationary theories of almost any kind. First, deflationism denies the idea that truth has a serious *explanatory* role to play. To say that truth has a serious explanatory role to play is to say that the role of truth goes beyond expressing generalizations. So, in order to defend deflationism, it must be shown that the truth predicate, in all the contexts in which it is used, is employed merely as a device of expressing generalizations. In particular, the deflationist must make sense of all the major philosophical principles involving the notion of truth without inflating the role of truth beyond what is acknowledged by the deflationary theory. This is a difficult task, because the use of the truth predicate is paramount. Consider some examples of philosophical principles involving the notion of truth: 'The meaning of a sentence involves its truth condition'; 'True beliefs facilitate successful behavior'; 'The success of a scientific theory is best explained by its truth'; 'Psychological laws essentially involve truth and reference'; and so forth. The deflationary theory must accommodate all these principles without assigning any additional role to the notion of truth.

What makes the deflationist project even more difficult is that many philosophically important notions—such as the notions of representational or semantic content, intentionality, meaning, realism, and anti-realism—have been deployed in a very close connection with the notion of

truth. In particular, the notions of representational content and intentionality are so tightly connected with the notion of truth that it would be inconsistent if one adopts a deflationary attitude toward the notion of truth while one adopts an inflationary attitude towards the notions of representational content and intentionality. The question is then what it is to give a deflationary account of the notions of representational content and intentionality. The two deflationists that I will discuss in this thesis—Horwich and Field—offer slightly different answers to this question. I will discuss it in Chapter 4.

The deflationist adopts a similar attitude toward the notion of reference as she does toward truth. For example, Horwich sees that the role of reference terms (such as ‘refers to’, ‘is about’, etc.) consists in enabling us to obtain valuable information from other people’s beliefs. Suppose someone said ‘John is the inventor of Viagra’. By knowing that ‘John’ refers to the guy living next door, I can learn that the guy living next door invented Viagra.⁸³ The whole point of *de re*, as opposed to *de dicto*, attribution of content is simply to make such an inference possible. But if this is the only function served by reference terms, then all we need is an instance of the reference schema (R) below:

(R) (x)(Tokens of $*n*$ refer to x iff $n = x$).⁸⁴

⁸³ I will present a more detailed explanation of Horwich’s account of the concept of reference in Ch. 2 (2.2). See P. Horwich, *Meaning*, Ch. 5 (“Reference”), Oxford: Clarendon Press, 1999.

⁸⁴ Horwich uses ‘ $n*$ ’, not ‘ n ’, in order to accommodate the ambiguity of an expression. So, ‘ $n*$ ’ indicate that ‘ n ’ is individuated not just phonologically but also by meaning. See P. Horwich, *ibid.* As will be discussed in Ch.2., Field characterizes the deflationary theory of reference in the same manner.

Just as the function or utility of the truth predicate determines the kind of a theory of truth that is needed, so does the function or utility of the reference predicate determine the kind of a theory of reference that is needed.

To sum up, the gist of the second motivation for the deflationary theory of truth is that the roles of truth terms and reference terms are not such that they require substantive/deductive theories. Given the roles that truth terms and reference terms play, theories of truth and reference need nothing more than what is stated by the instances of the disquotational or equivalence schemas for truth and reference. This is also the reason why Field eventually got to renounce the view he defended in “Tarski’s Theory of Truth.” In “Tarski’s Theory of Truth,” Field assumed that the role of truth is similar to the role of valence; that is, just like the notion of valence, the notion of truth has a causal-explanatory role. Construing the role of truth on a par with the role of valence, Field argued for the need of a reductive theory of reference. Field no longer defends the view; he now construes the role of truth as a purely logical device. In the next chapter, we will see how radically Field changed his view of truth from the days of “Tarski’s Theory of Truth.” He says that he’d like to think that by advocating radically opposed views he’d have gotten it right once. Let’s see whether his wish can come true.

Chapter 2. Field's Pure Disquotationalism

2.1. Introduction

In this chapter, I will discuss Field's version of deflationism—the theory he calls 'pure disquotationalism'. I do not intend to give a full characterization of Field's theory from the beginning; I intend it to be unfolded as the discussion of the theory evolves. So, at the outset, I will only give a brief account of Field's pure disquotationalism.⁸⁵

Field defines pure disquotationalism as the view that the notion of truth is *at bottom* purely disquotational. To understand what Field means by this claim, we first need to understand what he means by the 'purely disquotational' notion of truth. According to Field, to say that the notion of truth is purely disquotational is to say that the notion of truth has the following two features:

- i) 'true' as it is understood by a given person applies only to an utterance (or a sentence) that that person understands, and

⁸⁵ For the presentation of Field's pure disquotationalism, the following articles are referred to: H. Field, "The Deflationary Conception of Truth," in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 55-117; "Deflationist Views of Meaning and Content," (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001; "Disquotational Truth and Factually Defective Discourse," (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001; "Attributions of Meaning and Content," in his *Truth and the Absence of Fact*, Oxford University Press, 2001; "Apriority as an Evaluative Notion," in his *Truth and the Absence of Fact*, Oxford University Press, 2001; "Critical Notice: Paul Horwich's *Truth*," *Philosophy of Science*, Vol. 59, No. 2, 1992; "Indeterminacy, Degree of Belief, Excluded Middle," his *Truth and the Absence of Fact*, Oxford University Press, 2001.

- ii) for any utterance (or a sentence) ‘*p*’ that a person *X* understands, a truth-ascription of the form ‘‘*p* is true’ is cognitively equivalent for *X* to the utterance ‘*p*’ itself; for *X*, to say that ‘*p* is true is to say that *p*.⁸⁶

As will be explained later, Field changes his view on the feature i) in response to the criticism raised of it. I will also argue later in this chapter that pure disquotationalism doesn’t need the feature i) and will be better off without it. For these reasons, a more detailed explanation of the feature i) will be saved for a later discussion.

Unlike the feature i), the feature ii) will be with Field’s disquotational theory of truth as far as this theory goes—it is the core feature of the purely disquotational notion of truth. As a version of deflationism about truth, Field’s pure disquotationalism defends certain ideas of truth that are shared by other types of deflationism—truth is a purely logical notion in the sense that its role is expressive, and not explanatory; truth is not a notion of which the nature has to be explained by a substantive.⁸⁷ Here, a ‘substantive’ theory can be understood as any theory of

⁸⁶ Surely, to say that two sentences are cognitively equivalent implies a much stronger claim that they merely have the same truth value. Field defines the notion of cognitive equivalence in terms of the conceptual and computational roles of sentences; to say that the truth-claim ‘‘Snow is white’ is true’ is cognitively equivalent to the truth-less claim ‘Snow is white’ is to say that they have the same inferential patterns. See H. Field, ‘‘Disquotational Truth and Factually Defective Discourse,’’ in his *Truth and the Absence of Fact*, p. 405.

⁸⁷ Most of all, this view of truth accords with Horwich’s deflationary conception of truth; Horwich also construes truth as a purely logical notion serving as a device of endorsing propositions and expressing generalizations (see P. Horwich, *Truth*. 2nd ed., Oxford: Blackwell, 1999). Field’s view of truth, however, also accords with Stephen Leeds’s and Michael Williams’s views of truth (see S. Leeds, ‘‘Theories of Reference and Truth,’’ *Erkenntnis* 13, 1978, pp.111-129 and ‘‘Truth, Correspondence, and Success,’’ *Philosophical Studies*79, 1995, pp.1-36; M. Williams, ‘‘Do We (Epistemologists) Need a Theory of Truth?’’ *Philosophical Topics* 14, 1986, pp. 223-242, ‘‘Meaning and Deflationary Truth,’’ *Journal of Philosophy* 96, 1999, pp. 545-564, and ‘‘Context, Meaning, and Truth,’’ *Philosophical Studies* 117, 2004, pp. 107-130). In particular, Leeds’s influence on Field’s deflationism seems important; I got the impression that Leeds woke Field from an inflationary slumber (see Field’s ‘‘The Deflationary Conception of Truth,’’ in G. McDonald and C. Wright (eds.), *Fact, Science and Value*, Oxford: Blackwell, 1986, pp. 55-117). Field’s deflationary view of truth is in line with Grover, Camp, and Belnap’s prosententialist view of truth, according to which ‘is true’ isn’t a predicate but a prosentence-forming operator (see D. Grover, J. Camp and N. Belnap, ‘‘A Prosentential Theory of Truth,’’ *Philosophical Studies* 27, 1975, pp. 73-125). Although it is hard to

truth that gives a more inflated account of truth than what is stated by instances of the pure disquotational schema (DS) in the following:

(DS) ‘ p ’ is true iff p .⁸⁸

By stating that a truth-claim—for example, ‘‘Snow is white’ is true’—is *as a whole* cognitively equivalent to a truth-less claim— ‘Snow is white’—the pure disquotationalist makes it clear that a substantive theory of truth is not called for when the notion of truth is construed in the purely disquotational sense. If the truth-claim ‘‘Snow is white’ is true’ is as a whole cognitively equivalent to ‘Snow is white’, then the following biconditional is true as a matter of simple identity:

‘Snow is white’ is true iff snow is white.

accept the claim that ‘is true’ isn’t a predicate, the prosententialist analysis of the function of truth terms sheds light on the understanding of the deflationary conception of truth; in particular, on the understanding of deflationist view of the role of truth. The prosententialist view of truth seems to have been inspired by Strawson’s view of truth (see P. Strawson, “Truth,” *Analysis*, Vo. 9, 1949, pp. 83-97, and P. Strawson, “Truth,” *Aristotelian Society Supplementary Volume*, 1950). Given the close connection between Grover, Camp, and Belnap’s prosententialist view of truth and Brandon’s view’s, it may be said that Field’s view of truth is in line with Brandon’s view as well, although Field rarely mentions Brandon (see R. Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, Mass: Harvard University Press, 1998, *Articulating Reasons: An Introduction to Inferentialism*, Cambridge, Mass: Harvard University Press, 2000).

⁸⁸ A ‘substantive theory of truth’ can also be understood as any theory of truth in which an instance of the schema (DS) ‘ p ’ is true iff p isn’t considered an analytic truth (see H. Field, “Disquotational Truth and Factually Defective Discourse,” (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001).

So, on the purely disquotational conception of truth, an instance of (DS) is a logical truth; it is tautologous to say that ‘Snow is white’ is true if and only if snow is white. But then, no further explanation is needed for why ‘Snow is white’ is true iff snow is white. In this sense, the concept of truth is exhausted by what is implied by the instances of (DS) when truth is construed in the purely disquotational sense. Also in this sense, no substantive theory of truth is needed that analyzes the nature of truth in terms of other concepts.

In Field’s pure disquotationalism, the feature ii) is also at the heart of the claim that truth is *nothing but* a logical device. This can be explained as follows. In order to make sense of the logical function served by the notion of truth, the cognitive equivalence thesis isn’t required. To say that the notion of truth has a logical function is to say that we appeal to the notion of truth as a device of expressing generalizations. For example, we appeal to the truth predicate when we want to express our agreement with an utterance, but we cannot articulate it. Suppose I want to express my agreement with what John said, but I don’t remember exactly the sentence he uttered. In such a case, the truth predicate is indispensable—by saying ‘What John said is true’, I can express my agreement, which I couldn’t do if I didn’t have the truth predicate.⁸⁹ Taking another example, suppose we want to generalize an infinite conjunction such as ‘If $1+2=3$, then $1+2=3$; if Fido is an animal, then Fido is an animal; and . . .’ We cannot generalize this infinite conjunction by appealing to usual, objectual quantification (e.g. $(p)(\text{if } p, \text{ then } p)$), because we cannot quantify into sentential positions—sentences are not names. The truth predicate is handy in these cases, because the truth predicate enables us to talk about sentences while it also cancels

⁸⁹ Here, the notion of truth serves as a device of generalization in the following sense—the truth-claim ‘What John said is true’ is analyzed into ‘‘Snow is white’ is the sentence that John uttered, and snow is white; ‘Grass is green’ is the sentence that John uttered, and grass is green; or . . .’ We cannot generalize this infinite disjunction into $(E p)(‘p’ \text{ is the sentence that John uttered, and } p)$, because, first, the variable occurs both inside and outside quotations, and second, we cannot (within the realm of objectual quantification) quantify into sentential positions (sentences are not names).

out the effect of talking about sentences. As Quine says, we can talk about sentences by using quotations—quotations are devices for naming sentences.⁹⁰ But the truth predicate has the force of canceling the effect of talking about sentences—by saying that the sentence ‘Snow is white’ is true, we can say that snow is white. In this sense, the truth predicate is a device for disquotation. That the truth predicate has this feature is made explicit by the disquotational truth schema (DS): ‘*p*’ is true iff *p*. As long as we accept the instances of this schema, we can explain the logical function served by the notion of truth. But an instance of (DS) doesn’t by itself imply that a truth-claim of the form ‘‘*p*’ is true’ is *cognitively equivalent* to a truth-less claim ‘*p*’; it only implies that they have the same truth-value.⁹¹

So, in order to make sense of the logical function served by the notion of truth, we don’t need the cognitive equivalence thesis. The pure disquotationalist’s claim, however, isn’t simply that the notion of truth has a logical function—his claim is that there is no other function served by the notion of truth. Here, the cognitive equivalence thesis is required—more precisely, it is required because truth isn’t construed as an *explanatory* notion in Field’s purely disquotational theory of truth. According to Field, truth doesn’t play a causal-explanatory role in explanations of meaning and the success of behavior.⁹² If the notion of truth had a causal-explanatory role, the truth predicate would express a causal-explanatory property. But then, the notion of truth would be construed on the par with the notions of being magnetic, being H₂O, and so forth. The result would be totally inconsistent with the core thesis of pure disquotationalism—no theory of truth other than what is implied by the instances of the pure disquotational schema (DS) is needed. If

⁹⁰ W. V. Quine, *Philosophy of Logic*, Eaglewood Cliffs: Prentice-Hall, 1970.

⁹¹ Quine also says that (in *Philosophy of Logic*, p. 11) the disquotational feature of the truth predicate is made explicit by Tarski’s T-sentences; for example, ‘Snow is white’ if and only if snow is white. But Tarski said nothing like the cognitive equivalence between the left-hand side and the right-hand side of ‘iff’ in a T-sentence—he only says that T-sentences are true.

⁹² This will be discussed in Ch. 4.

the notion of truth is construed on the par with the notions of being magnetic, being H₂O, etc., then the nature of truth couldn't be exhausted by the instances of the disquotational schema. So, from the purely disquotationalist point of view, the cognitive equivalence schema is needed to make sure that truth is nothing but a logical notion; that it isn't an explanatory notion.

The cognitive equivalence thesis, however, confronts several difficult problems. One of them is that it conflicts with the modal properties of truth. If a truth-claim such as ‘‘Snow is white’ is true’ is cognitively equivalent to ‘Snow is white’, then an instance of (DS) is a necessary truth; it is a logical truth. This feature of the purely disquotational truth, however, doesn't go well with the modal properties of the ordinary notion of truth. This problem will be discussed in the next section.

Another problem with the cognitive equivalence thesis is related to the feature i) mentioned above—that ‘true’ as it is understood by a given person applies only to an utterance (or a sentence) that that person understands. Field *needed* this feature to defend the cognitive equivalence thesis; that is, the feature i) is subordinate to the feature ii).⁹³ In order to say that the left-hand side and the right-hand side of ‘iff’ in

‘*p*’ is true iff *p*

are cognitively equivalent, it is essential that the sentence replacing ‘*p*’ at the left side means the same as the sentence replacing ‘*p*’ at the right side. There is more than one way to have this condition met; for example, the pure disquotationalist may say that the sentence replacing ‘*p*’ at the left side belongs to the same language that the sentence replacing ‘*p*’ at the right side does, and there is interpersonal synonymy within a language. In other words, if the pure

⁹³ The reason the past tense is used here is that Field later changes his view on the feature i). The change of Field's view on i) will be discussed in this chapter.

disquotationalist assumes that the same sentence-type is understood in the same way by different speakers in a language, he can easily have the condition met.

This road is blocked to Field—he skeptical of interpersonal synonymy.⁹⁴ As a skeptic of interpersonal synonymy, Field looked to another way of having the condition met. Thus, Field characterizes the purely disquotational truth predicate as one that is ascribed to a sentence as it is understood by a speaker—he says, “As a rough heuristic, we could say that for a person to call an utterance true in this pure disquotational sense is to say that it is true-as-he-understands-it.”⁹⁵ So, Field needed the feature i) as a way of making sense of the feature ii).

The feature i), however, causes a serious problem—probably an even more serious problem than the problem with not accommodating the modal properties of truth. If we accept Field’s characterization of the purely disquotational notion of truth, then we wouldn’t be able to ascribe the truth predicate to a sentence in a language that we don’t understand. Initially, Field wasn’t so much worried about this problem; his initial attitude was that if the purely disquotational notion of truth diverges from the ordinary notion of truth in this respect, then it is not a problem with pure disquotationalism but a problem with how ordinary people use the truth predicate.⁹⁶ This attitude had to be changed when he confronted the objection raised by Shapiro. This will be discussed in sections 2.2.1 and 2.2.2.

⁹⁴ See H. Field, “Deflationist Views of Meaning and Content” and “Attributions of Meaning and Content” in his *Truth and the Absence of Fact*, Oxford University Press, 2001. Field’s skepticism about interpersonal synonymy will be discussed in Ch. 4.

⁹⁵ H. Field, “Deflationist Views of Meaning and Content,” in his *Truth and the Absence of Fact*, Oxford University Press, 2001, p. 106.

⁹⁶ Field says that his initial attitude was simply to bite the bullet, and say that the disquotational truth predicate cannot be applied to a sentence that one doesn’t understand. See H. Field, “Postscript: Deflationist View of Meaning and Content,” in his *Truth and the Absence of Fact*, p. 147.

In responding to the objection raised by Shapiro, Field makes a considerable change to his initial thesis of disquotationalism.⁹⁷ Field doesn't state explicitly how his ideas have changed, but I will try to make it clear in the section 2.2.3. Although the changes Field makes to his initial thesis is helpful to solve the objection raised by Shapiro, they are not enough to entirely fix the problem. So, in the last section of this chapter, I will propose a different solution to the problem from Field's. This new solution proposes a disquotational theory of truth that doesn't have the feature i). I will argue that the disquotational notion of truth isn't immanent to a truth-ascriber's understanding of a language. In pure disquotationalism, it is the notion of truth-conditions—not the notion of truth itself—that has to be construed as being immanent to a language that a truth-attributor understands.⁹⁸ By making the distinction between the notion of truth and the notion of a truth condition, the disquotationalist may resolve some of the most vexing problems with the theory.

2.2 The Modal Objection

2.2.1 The Problem and a Misunderstanding

⁹⁷ See H. Field, "Postscript: Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*. Also see S. Shapiro, "The Guru, The Logician, and The Deflationist: Truth and Logical Consequence," *Nous*, 37 (1), 2003: 113-132.

⁹⁸ Field says:

I am inclined to think that a minimalist needs to take the more radical position that talk of truth makes sense primarily in connection with our own language, and makes sense for other languages only insofar as they are translatable into our own. ("Truth is an immanent notion", as Quine puts it.)

H. Field, "Critical Notice: Paul Horwich's *Truth*," *Philosophy of Science*, Vol. 59, 1992, pp. 321-330, p. 325.

The modal objection to Field's pure disquotationalism is that Field's theory of truth gets the modal properties of the truth predicate wrong. The objection was raised by various people; for example, Marian David, Vann McGee, W. Kunne, Ray Buchanan, and, most of all, Field himself.⁹⁹ Interestingly, Putnam raised the same objection to Tarski's definition of truth—that in Tarski's definition of truth, a biconditional such as

- 1) 'Snow is white' is true (in English) iff snow is white

becomes a logical, thereby, necessary truth.¹⁰⁰ The modal objection raised to Field's disquotational theory is basically to the same effect. That is to say, in Field's theory, an instance of (DS) such as 'Snow is white' is true iff snow is white is construed as having the following modal status:

- 2) \Box ('Snow is white' is true iff snow is white).

The reason why the modal status of an instance of the disquotational schema is construed along the lines of 2) is that an instance of (DS) is considered a logical truth—it is a tautology. Given the cognitive equivalence thesis that was explained previously, the truth-claim that 'Snow is

⁹⁹ See M. David, *Correspondence and Disquotation: An Essay on the Nature of Truth*, Oxford University Press, 1994; V. McGee, "A Semantic Conception of Truth?" *Philosophical Topics* 21 (2): 83-111; W. Kunne, "Disquotationalist Conceptions of Truth," in *What is Truth?*, R. Schantz (ed.), Berlin; New York; de Gruyter, 2002; H. Field, "The Deflationary Conception of Truth," in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 55-117; H. Field, "Deflationist View of Meaning and Content," (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001.

¹⁰⁰ This is another reason why Tarski's theory of truth is often considered deflationary; that is to say, Tarski was not interested in explaining the semantic properties of an expression with his theory of truth. S. Soames offers a deflationary interpretation of Tarski's theory of truth on this very ground. See S. Soames, "What is a Theory of Truth?," *The Journal of Philosophy* 81, 411-429, 1984. See also H. Putnam, "A Comparison of Something with Something Else", *New Literary History* 17: 61-79, 1985.

white' is true is cognitively equivalent to the truth-less claim that snow is white. In Field's disquotationalism, an instance of (DS) is true as a matter of simple identity.¹⁰¹ But then, an instance of (DS) is a logical truth, and therefore, the modal status of an instance of (DS) should be understood along the lines of 2).

Field's position on the modal properties of the notion of truth has gone through some changes. At first, Field took a radical position—he said that the purely disquotational notion of truth is the one that makes counterfactual claims C_1 and C_2 come out equivalent:

C_1 : If we had used the word 'white' differently, 'Grass is white' might have been true.

C_2 : If we had used the word 'white' differently, grass might have been white.¹⁰²

To say that the counterfactual claims C_1 and C_2 are equivalent is to say that the way we use the word 'white' doesn't affect the truth condition of 'Grass is white'. To say that C_1 and C_2 are equivalent implies that it is necessary that 'Grass is white' has the truth condition that grass is white. But that is possible only if the way we use words doesn't affect the truth condition of a sentence. So, by claiming that the purely disquotational notion of truth is the one that renders C_1 and C_2 equivalent, Field wanted to make it clear that purely disquotational truth is a *use-independent* notion. More precisely, Field's point is that when the notion of truth is understood

¹⁰¹See H. Field, "The Deflationary Conception of Truth," in G. MacDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 55-117.

¹⁰² See H. Field, "The Deflationary Concept of Truth," in G. MacDonald and C. Wright (eds.) *Fact, Science and Morality*, Oxford: Blackwell, 1986, p. 58.

in the purely disquotational sense, the truth condition of a sentence is characterized independently of the semantic—meaning or use properties—of a sentence.¹⁰³

To say that C_1 and C_2 are equivalent is obviously at odds with the ordinary notion of truth. That is to say, ordinary folks would hardly take ordinary subjunctive conditionals C_1 and C_2 to be equivalent or the subjunctive statement 2) above to be true. Consider Marian David's objection to Field's disquotationalism:

So maybe (D) (' p ' is true if and only if p) has to be understood as the statement of a law-like *regularity*. Such an interpretation would perhaps permit the definition to occupy some middle ground between a necessary equivalence and a material biconditional. But this interpretation is not plausible either, for (D) supports the wrong subjunctives:

- a) If 'Snow is green' were true, then snow would be green.¹⁰⁴
- b) If 'Snow is white' were not true, then snow would not be white.

Both subjunctives are false. If the sentence 'Snow is green' were indeed true, then that would be so *either* because we would be using the word 'green' as we do now and snow would be green *or* because snow would be white and we would be using the word 'green' to mean what we now mean by 'white'. If a) were correct, we could make snow green quite easily, namely simply by choosing to use 'green' from now on as we used 'white' until now.

¹⁰³ In the same paper where Field characterizes the purely disquotational notion of truth as the one making C_1 and C_2 equivalent, he says that he understands a theory of truth as a theory of a *truth condition*. What he means by this is that we need a substantive (non-deflationary) theory of truth *if* the notion of a truth condition turns out to be a causal-explanatory notion. If a truth condition is a causal-explanatory notion, then we need a substantive theory of a truth condition, and a substantive theory of a truth condition is no other than a substantive theory of truth. So, Field's initial idea was the following: if a truth condition is a use-independent notion, then we don't need to explain the truth condition of a sentence in terms of the semantic properties of a sentence. But then, we don't need a substantive theory of a truth condition. It is a convenient way of thinking, but it has a flaw. The flaw is that he didn't see the third possibility—that a truth condition is a use-dependent notion but its use-dependency is explained in a deflationary way. It is a matter of how we explain the connection between meaning and a truth condition. Field's flaw in his initial stage of deflationism is that he only thought of two possibilities—either a truth condition is defined in terms of a semantic notion or it has no connection with a semantic notion. There is, however, a third possibility—that a truth condition is connected with a semantic notion but it is not *defined* in terms of a semantic notion. Field finally arrives at this position after more than a decade from when he first characterized purely disquotational truth as a use-independent notion in his "The Deflationary Concept of Truth," in G. MacDonald and C. Wright (eds.) *Fact, Science and Morality*, Oxford: Blackwell, 1986. See 2.2.2 and 2.2.3 in this chapter.

¹⁰⁴ Numbering is modified from the original.

Similarly for b); if it were correct, we could make snow stop becoming white simply by using 'white' to refer to, say, the color of fruit bars.¹⁰⁵

According to David's argument, ordinary folks consider subjunctive conditionals such as a) and b) false because we understand the subjunctive clauses such as 'If 'Snow is green' were true' and 'If 'Snow is white' were not true' as describing counterfactual situations in which we use the words differently. In other words, according to David, ordinary folks understand subjunctive truth-statements such as 3) and 4) as expressing 3^o) and 4^o) in the following respectively:

- 3) 'Snow is green' might have been true.
- 4) 'Snow is white' might not have been true.
- 3^o) In some possible world, we might have used the word 'snow' or 'green' in such a way that the sentence 'Snow is green' is true in the world.
- 4^o) In some possible world, we might have used the word 'snow' or 'white' in such a way that the sentence 'Snow is white' is not true in the world

To say that we ordinary folks understand 3) and 4) in terms of 3^o) and 4^o) respectively is to say that ordinary people's understanding of the modal properties of 'true' is different from the pure disquotationalist's. According to the pure disquotationalist, the modal property of 'true' is such that an instance of (DS) is necessarily true; for example, it is a necessary truth that 'Snow is white' is true iff snow is white. According to David, however, this isn't the way we, ordinary

¹⁰⁵ M. David, *Correspondence and Disquotation: An Essay on the Nature of Truth*, Oxford University Press, 1994. pp. 130-131.

folks, understand the modal properties of ‘true’. Ordinary folks understand the modal properties of ‘true’ as having to do with the meta-linguistic, semantic, or use properties that words might have. Therefore, we don’t consider subjunctive conditionals such as a) and b) to be true. To think that a) and b) are true, says David, is to have the crazy idea that if we had used words differently, snow might have been green or snow might not have been white.¹⁰⁶ In other words, to think that a) and b) are true is to think that we could change world easily by changing the way we use words.¹⁰⁷

Is David’s argument sound? Although this argument succeeds in posing a serious challenge to Field’s theory, it has its own flaw. I will discuss its flaw before discussing its merit.

According to David, a subjunctive truth-statement such as a) or b)—‘If ‘Snow is green’ were true, snow would be green’ or ‘If ‘Snow is white’ were not true, snow would not be white’—is “unambiguously” false.¹⁰⁸ The fact of the matter, however, is much more complicated than that—the meaning of a subjunctive truth-statement is *ambiguous*. David’s understanding of a subjunctive truth-statement reveals only *one* aspect of the meaning of a subjunctive truth-statement; it doesn’t exhaust the entire meaning of a subjunctive truth-statement. Granted, we sometimes understand subjunctive truth-statements such as 3) and 4) in terms of 3⁰) and 4⁰). That may also reflect ordinary folks’ immediate response to subjunctive truth-statements such as 3) and 4). But our ordinary notion of truth is a mixed bag, and this many-sidedness of the notion of truth is accentuated when it is conjoined with modality.

¹⁰⁶ M. David, *Correspondence and Disquotation: An Essay on the Nature of Truth*, Oxford University Press, 1994, Ch. 4.

¹⁰⁷ M. David, *Correspondence and Disquotation: An Essay on the Nature of Truth*, Oxford University Press, 1994, Ch. 4.

¹⁰⁸ M. David, *ibid.*

Depending upon a speaker's intention in a specific context, subjunctive truth-statements such as 3) and 4) can also be understood in terms of 3^{pd}) and 4^{pd}), not only in terms of 3^o) and 4^o):

3^{pd}) Snow might have been green.

4^{pd}) Snow might not have been white.

We sometimes mean 3^{pd}) and 4^{pd}) when utter subjunctive truth-statements such as 3) and 4). Let's suppose the Pope said 'The earth is flat' to Galileo when Galileo was excommunicated several hundred years ago. Suppose Galileo responded to the Pope by saying 'What you say might have been true in another world'. Here, Galileo doesn't use the truth predicate in the way that David understands it. In this context, Galileo isn't interested in the usage of a word. What Galileo intends to say in this context isn't that if we had used the word 'earth' differently, the Pope's utterance *e* might have been true. What he intends to say is rather that the Pope's utterance ('The earth is flat) might have been true if and only if the earth would have been flat itself. So, in a context in which a speaker isn't interested in the semantic/use properties of a sentence, a subjunctive truth-state of the form '*p*' might have been true' should be understood in terms of 'It might have been that *p*', rather than 'In some possible world, we might have used the sentence '*p*' in such a way that '*p*' is true in that world.'

It is a misconception about the modal properties of truth to think that subjunctive truth-statements such as 3) and 4) should always be understood in terms of *meta-linguistic* subjunctive statements such as 3^o) and 4^o). We sometimes intend to make a meta-linguistic statement such as

3^o) or 4^o) in uttering a subjunctive truth-statement such as 3) or 4). But this doesn't exhaust the entire purpose of a subjunctive truth-statement. In other contexts, in uttering a subjunctive truth-statement, we simply want to talk about the modal properties of the *world*. In such a context, we don't want the semantic properties of words to vary from one possible world to another possible world; namely, when we want to talk about the modal properties of the *world* rather than *words*, we need to keep the truth condition of a sentence constant from world to world. If the truth condition of a sentence changes from world to world, we cannot talk about the modal properties of the *world*. We cannot, because we would not talk about the same world when the semantic properties of words change. When we talk about the modal properties of the world, however, we talk about different ways the same world might have been. So, when we want to talk about the modal properties of the world, we need to keep the semantic properties constant from world to world. Therefore, in such a context, the modal statement 2)—□ ('Snow is white' is true iff snow is white)—is what we need.

Viewed this way, it is unfair to say that Field's pure disquotationalism is entirely wrong about the modal properties of truth. To the contrary, a certain aspect of the modal properties of 'true' requires Field's purely disquotational understanding of the modal properties of truth. Unfortunately, David misses this point. In the passage quoted above, David says

If a)—If 'Snow is green' were true, then snow would be green—were correct, we could make snow green quite easily, namely simply by choosing to use 'green' from now on as we used 'white' until now.¹⁰⁹

Of course, no disquotationalist would claim that we can make snow have the color green just by choosing to use the word 'green' in the way we use 'white' now. That is a crazy idea. David

¹⁰⁹ M. David, *ibid.*

credits this crazy idea to the disquotationalist, because he understands the subjunctive truth-statement “‘Snow is green’ were true’ in terms of a subjunctive *meta-linguistic* statement. David understands

5) if ‘Snow is green’ were true

in terms of 5*) in the following:

5^o) if we used the word ‘green’ in the way we use the word ‘white’ in the actual world.

If 5) is understood in terms of 5*), then to say that a) is true implies that we can easily make snow green by choosing to use a word differently. So, David credits this crazy idea because he thinks that 5) can be understood only in terms of 5*).

The subjunctive truth-statement 5), however, can be understood in terms of 5^{pd}) in the following:

5^{pd}) if snow were green,

if one's intention in uttering 5) is to talk about the modal properties of the world. In such context, a)— If 'Snow is green' were true, then snow would be green—is true. This is the reason why the purely disquotational conception of truth has a point. When we make a subjunctive truth-statement such as 'What the scientist said might be true', we don't make a modal statement about the use of a sentence. After all, if it makes sense to say that our purpose of stating that 'Snow is white' is true is simply to state that snow is white (at least in some contexts), then the disquotationalist view of the modal properties of truth should make sense as well. In making a truth-statement of the form '*p* is true', are we always interested in the semantic properties of '*p*'? If the answer is no, then a subjunctive truth-statement of the form '*p* might have been true' can be understood in terms of an object-level subjunctive statement 'It might have been that *p*'. David's mistake is not to see this aspect of the truth predicate.

That said, Field's pure disquotationalism still owes us an explanation of how a subjunctive truth-statement can be about modal properties of the semantic, use-dependent features of words. Although David and other critics of pure disquotationalism got one aspect of the modal properties of 'true' wrong, they are right about the other aspect. In uttering a subjunctive truth-statement of the form '*p* might have been true', we sometimes intend to make a modal statement about the semantic, use properties of the words. How can this non-disquotational—or, as Field puts, "inflationary"—reading of a subjunctive truth-statement be accommodated by Field's pure disquotationalism? Ever since Field proposed the theory, he has attempted several solutions to respond to this question. In the following, I will examine the most-often discussed solution among them—the quasi-disquotational solution.

2.2.2 The Quasi-Disquotational Solution

The term ‘quasi-disquotational’ is from the quasi-disquotational notion of truth that Field introduces in order to accommodate the non-disquotational reading of a subjunctive truth-statement.¹¹⁰ Before explaining Field’s account of the quasi-disquotational notion of truth, however, it may be helpful to have knowledge about Field’s general strategy in dealing with problems of this sort—the problems caused by the discrepancy between the ordinary notion of truth and the purely disquotational notion of truth. From the beginning, Field was fully aware of that the ordinary notion of truth often diverges from the purely disquotational notion of truth. As was stated previously, Field characterizes purely disquotational truth as a use-independent notion. The ordinary notion of truth, however, is not a use-independent notion. Put more accurately, the ordinary notion of truth has both aspects—the use-dependent aspect and the use-independent aspect. At times, we employ the truth predicate without having any intention of giving semantic information about the sentence to which it is ascribed. At other times, we employ the truth predicate *with* an intention to give semantic information about the sentence to which it is ascribed. Parallel with this duality, the ordinary truth predicate is sometimes ascribed to a purely orthographic sentence-type, and at other times it is ascribed to a sentence-type that has its semantic properties built into it.

¹¹⁰ See H. Field, “Deflationist Views of Meaning and Content,” in his *Truth and the Absence of Fact*. Field also invokes the extended and the modified disquotational notions of truth in order to deal with the modal objection (see his “The Deflationary Conception of Truth,” in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 1986, pp. 55-117 and “Deflationist Views of Meaning and Content”). I won’t discuss Field’s extended and modified disquotational approaches to the modal problem, because he considers the quasi-disquotational approach as more promising than the other two. Even the quasi-disquotational approach, however, is eventually given up for another approach—the schematic approach. See H. Field, “Deflationist Views of Meaning and Content” and “Postscript: Deflationist Views of Meaning and Content” in his *Truth and the Absence of Fact*.

To say that purely disquotational truth is a use-independent notion is to say that the purely disquotational truth predicate isn't ascribed to a purely orthographic type but to a sentence-type that has semantic properties built into it. Most of all, if the truth predicate is ascribed to a purely orthographic type, it doesn't make sense to say that the truth-claim "'Snow is white' is true" is cognitively equivalent to the truth-less claim 'Snow is white'. Only when 'Snow is white' at the left-hand side of 'iff' in the biconditional

'Snow is white' is true iff snow is white

is understood as meaning the same as 'Snow is white' at the right-hand side, it makes sense to say that the truth-statement at the left-side is as a whole cognitively equivalent to the truth-less statement at the right side. Since the cognitive equivalence thesis is the core idea characterizing the purely disquotational notion of truth, the purely disquotational truth predicate is meaningful only when it is ascribed to a sentence-type that has semantic properties built into it. In this sense, Field says that the sentence replacing '*p*' at the left-hand side of the disquotational schema (DS) should be construed as a *computational type*—a sentence-type individuated in terms of its computational properties.¹¹¹ In Field's account, the meaning of an expression is explained in terms of its computational properties; namely, the inferential/conceptual roles and the indication relations of the expression.¹¹² Moreover, since Field is skeptical of interpersonal synonymy, it doesn't make sense to talk about the interpersonal sameness of a computational property. So, in Field's theory, the sentence 'Snow is white' at the left side and the right side of 'iff' in the

¹¹¹ H. Field, "Disquotational truth and Factually Defective Discourse," in his *Truth and the Absence of Fact*, Oxford University Press, 2001.

¹¹² See H. Field, "Deflationist Views of Meaning and Content" and "Attributions of Meaning and Content" in his *Truth and the Absence of Fact*. Field's deflationist account of meaning will be discussed in detail in Ch. 4.

biconditional above is to be understood in terms of a computational sentence- type understood by individual speaker *X*.

Viewed this way, it is inevitable that the purely disquotational notion of truth conflicts with the modal properties of the ordinary notion of truth. An aspect of the ordinary notion of truth is that a subjunctive truth-statement of the form i) in the following can be understood in terms of a meta-linguistic subjunctive statement of the form ii):

- i) ‘*p*’ might have been true.
- ii) In some possible world, ‘*p*’ might have had semantic properties that make ‘*p*’ in that world true.

If we understand the sentence replacing ‘*p*’ in i) as a computational type—or a sentence-type that has its semantic properties built into it—we cannot derive ii) from i). It simply doesn’t make sense to say that sentence ‘*p*’ as I understand it now might have had different semantic properties. We can derive ii) from i) only when we construe the sentence replacing ‘*p*’ as a purely orthographic type. Only then, does it make sense to say that the sentence replacing ‘*p*’ might have had different semantic properties than it has now. In order to derive ii) from i), we should detach the semantic (use) properties from the syntactic/orthographic properties of a sentence.

Whenever the purely disquotational notion of truth diverges from the ordinary notion of truth in this way, Field tries to explain the latter in terms of the purely disquotational notion of

truth, using fairly limited additional resources.¹¹³ This is Field’s general strategy in dealing with problems of this sort.¹¹⁴ The motivation behind Field’s invoking the quasi-disquotational notion of truth should be understood from this perspective as well—the quasi-disquotational notion of truth is at bottom purely disquotational, but some additional resources are invoked in its characterization. As with almost everything else, Field offers more than one characterization of the quasi-disquotational notion of truth.¹¹⁵ I chose to present here the original version, and it defines the quasi-disquotational notion of truth as follows:

¹¹³ H. Field, “Disquotational Truth and Factually Defective Discourse,” in his *Truth and the Absence of Fact*, Oxford University Press, 2001, p. 223.

¹¹⁴ Despite many changes Field has brought into his theory of deflationism, it seems that he still holds onto this general strategy. But he weakened his view on how this general strategy has to be operated. At first, he thought that any non-disquotational notion of truth (the ordinary notion of truth) has to be *defined* by the purely disquotational notion. This view, however, changed to a weaker view that explains the connection between the quasi-disquotational and the purely disquotational notions of truth in terms of a schema. See sec. 2.3.2 of this chapter.

¹¹⁵ The original version is in H. Field, “Deflationary Views of Meaning and Content,” and the other two versions are in “Postscript to “Deflationary Views of Meaning and Content,” both in H. Field, *Truth and the Absence of Fact*, Oxford University Press, 2001. *One* of later versions, however, is very wrong:

- (*) S is true_{qd} (at possible world v) iff S is to be translated by a sentence of mind (in the actual world) that is purely disquotationally true (at v).

(See “Postscript to “Deflationary Views of Meaning and Content.”)

This version is wrong because ‘ S ’ is here a purely orthographic type, therefore, it doesn’t make sense to “translate” ‘ S ’ into a sentence of mine. We cannot translate a purely orthographic type. The other one (**) has a problem as well:

- (**) $S_{X,u}$ is true_{qd} (at possible world v) iff $S_{X,u}$ is to be translated by a sentence of mine (in the actual world) that is purely disquotationally true (at v).

About this version, Field says that it avoids using two distinct truth predicates—the purely disquotational one and the quasi-disquotational one—because the truth predicate is ascribed to a computational type ‘ $S_{X,u}$ ’. So, he says that we can use the single truth predicate by taking ‘ S ’ in (*) as a computational type, rather than a purely orthographic type. But does (**) use the single truth predicate? It still uses two distinct predicates: the quasi-disquotational and purely disquotational truth predicates. So, I am not sure why Field says that he now can use the single truth predicate with version (**). That doesn’t mean that it is wrong to consider the truth predicate ascribed to ‘ $S_{X,u}$ ’ quasi-disquotational. It should be considered quasi-disquotational in Field’s theory, because Field characterizes the purely disquotational truth predicate as being ascribed to a sentence “as-I-understand-it”. Although ‘ $S_{X,u}$ ’ is a computational type, it is a computational type understood by speaker X at the possible world v . The sentence replacing ‘ $S_{X,u}$ ’ doesn’t occur as a sentence-as-I-understand-it. So, it is outside the extension of the purely disquotational truth predicate as characterized by Field. But then, the truth predicate ascribed to ‘ $S_{X,u}$ ’ is quasi-disquotational by “Field’s own standards.” Field offers another approach (***), and that approach will be discussed in the next section.

(QD) $\square (S \text{ is true}_{\text{qd}} \text{ iff } \Sigma p[\exists m(m \text{ is the meaning of } S \text{ and } m \text{ is the meaning of 'p' in the actual world}) \& \text{'p' is true}_{\text{pd}}])$,

and analogously,

$\square [S \text{ has the quasi-disquotational truth condition that } p \text{ iff } \exists m(m \text{ is the meaning of } S \text{ and } m \text{ is the meaning of 'p' in the actual world})]$.¹¹⁶

In order to see how the quasi-disquotational notion of truth helps addressing the modal objection, let's consider a subjunctive truth-statement 1) in the following:

1) 'Snow is green' might have been true_{qd},

where 1) is understood along the lines of 1*) in the following:

1*) In some possible world, the sentence 'Snow is green' might have had semantic properties that make 'Snow is green' true in that possible world.

The reason why there is such confusion is that Field restricted the extension of the purely disquotational truth predicate to a sentence-as-I-understand-it. Once we get rid of this restriction, we can consider the truth predicate ascribed to ' $S_{x,u}$ ' in (**) as purely disquotational. Interestingly (and curiously), Field goes against his own restriction in the approach (***) that will be discussed later. The problem is that he doesn't say explicitly that he dropped the restriction.

¹¹⁶ 'true_{qd}' abbreviates 'true in the quasi-disquotational sense' and 'true_{pd}' abbreviates 'true in the purely disquotational sense'. See H. Field, "Deflationist Views of Meaning and Content," (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001, p. 131.

Given that 1) is understood in terms of 1*), the truth predicate in 1) cannot be construed as being ascribed to a computational type. So, ‘Snow is green’ in 1) should be construed as a purely orthographic type. That means the truth predicate employed in 1) isn’t purely disquotational.

To accommodate the truth predicate employed in 1), Field correlates the purely orthographic sentence-type ‘Snow is green’ with a true sentence ‘*p*’ in one’s own language, where ‘*p*’ is construed as a computational type. The purpose of correlating the purely orthographic type ‘Snow is green’ with a computational type ‘*p*’ in one’s language is to extend the purely disquotational truth predicate to the purely orthographic type ‘Snow is green’ by invoking the meaning *m* that connects the purely orthographic type with the computational type.¹¹⁷ So, strictly speaking, the truth predicate is ascribed to a sentence ‘*p*’ in one’s own language, where the truth condition of this sentence is exhaustively explained in terms of the purely disquotational schema (DS). Using the equivalence of meaning as a bridge, however, the purely disquotational truth predicate can be indirectly ascribed to the purely orthographic type ‘Snow is green’. In this sense, the truth predicate employed in 1) is *quasi-disquotational*; although it isn’t itself disquotational, it is defined in terms of the purely disquotational truth predicate using the notion of a meaning-entity *m* as an additional resource.

By invoking the notion of meaning-entity *m*, says Field, the quasi-disquotational notion of truth tries to “mimic” a non-purely disquotational notion of truth. Notice that one of the essential features of the purely disquotational notion of truth is that it cannot be ascribed to a purely orthographic type. In this sense, the ordinary notion of truth isn’t purely disquotational, since it can also be meaningfully ascribed to a purely orthographic type, thereby yielding modal

¹¹⁷ Does it make sense to say that a purely orthographic type ‘Snow is green’ has the same meaning that a sentence ‘*p*’ in my language does? It does—the idea is to think that the semantic property of a sentence is detachable from the sentence as a purely orthographic type.

properties that are opposite to the modal properties of the purely disquotational notion of truth. By introducing the quasi-disquotational notion of truth, Field tries to bridge the gap between the ordinary notion of truth and the purely disquotational notion of truth. To see why let's consider how the quasi-disquotational definition of truth (QD) analyzes the substantive truth-statement 1): 'Snow is green' might have been true. According to (QD), 1) is to be analyzed in terms of 1^{qd}) in the following:

1^{qd}) 'Snow is green' might have had the same meaning m that a sentence ' p ' in my language has in the actual world, and ' p ' is true_{pd}.

Analogously, to say that 'Snow is green' might have the truth condition that p is to say that there is a meaning m such that 'Snow is green' has m at some possible world, and m is the meaning of ' p ' in the actual world. Intuitively, this is very similar to how we understand the ordinary notion of truth as it is ascribed to a purely orthographic type. An aspect of the ordinary notion of truth is such that a sentence x has the truth condition that p because x means that p . (QD) captures this basic intuition, but it does so by invoking a computational type ' p ' in one's own language. Once we invoke a sentence in our own language, we can define the ordinary notion that diverges from the purely disquotational one using an additional resource as a bridge.

How plausible is this solution? To answer this question, let's consider some concerns that can be raised about it. One is that (QD) invokes both interpersonal synonymy and the notion of a meaning-entity. To do so, however, doesn't cohere with Field's view of meaning, since Field is skeptical about the notions of interpersonal synonymy and a meaning-entity. How serious is this

concern? The answer is “Not so much.” First, (QD) can be easily modified into a weaker version that doesn’t invoke interpersonal synonymy or a meaning-entity. Instead, the pure disquotationalist can invoke the notion of a “similarity of meaning-characteristics”, which means the similarity of conceptual/inferential roles and indication relations (or simply, the similarity of use-properties). But then, (QD) may be modified into (QD*) as follows:

(QD*) $\square (S \text{ is true}_{\text{qd}} \text{ iff } \Sigma p[\exists \phi (\phi \text{ is the meaning-characteristics of 'p' in the actual world and } S \text{ has meaning-characteristics similar to } \phi) \text{ and 'p' is true}_{\text{pd}}])$,

and analogously,

$\square [S \text{ has the truth condition that } p \text{ iff } \exists \phi (\phi \text{ is the meaning-characteristics of 'p' in the actual world and } S \text{ has meaning-characteristics similar to } \phi)]$.

As long as (QD) can be understood in terms of (QD*), it doesn’t seem a serious problem that (QD) invokes interpersonal synonymy and the notion of a meaning-entity.

There is, however, a more serious problem with the quasi-disquotational approach: the problem that the explanation provided by the quasi-disquotational approach (QD) puts things “backwards”. The crux of the modal objection is that it is of no use to say that a biconditional such as 2) in the following is a necessary truth if we construe ‘Snow is green’ at the left side as a computational type:

2) 'Snow is green' is true iff snow is green

To say this, a critic might say, is to admit that the truth of a computational type can be understood only in terms of the semantic properties of a sentence. Unless we understand the meaning of 'Snow is green' at the left side in a certain way, we could not consider 2) a necessary truth. But then, says the same critic, it is only fair to say that the truth and truth condition of a computational type should be understood in terms of the meaning of a purely orthographic type.

True, the quasi-disquotational definition (QD) almost captures this idea, but it makes an incorrigible mistake---it invokes the truth of a computational type '*p*'. The whole point of my objection, says the critic, is that it is phony or spurious to say that the truth---in particular, truth condition---of a computational type '*p*' is use-independent. How can it be so if the notion of a computational type is to be understood in terms of semantic notions? A sentence of my language, for example, 'Snow is green', being a computational type, has the truth condition that snow is green because of its semantic properties. Hence, it is putting the matter backwards to invoke the truth and truth condition of a computational type in defining the truth and truth condition of a purely orthographic type. The pure disquotationalist may invoke the truth or truth condition of a sentence in one's own language in order to accommodate the truth and truth condition of a sentence in another language (see the next section). But the modal objection is different, because the modal objection raises a question of the very ground of the truth and truth condition of a sentence in my own language.

If there is a critic arguing this way, he seems right—there is something wrong with the quasi-disquotational approach. Not surprisingly, Field later abandons it for another solution.¹¹⁸ Before discussing the new solution, however, let's consider what needs to be done. What needs to be done, as Field says, is to “mimic” a non-purely disquotational theory of truth.¹¹⁹ A ‘non-purely disquotational theory of truth’ means a theory of truth that invokes a semantic notion in the characterizations of truth and truth conditions. Field's purely disquotational theory of truth doesn't do this, since the purely disquotational truth predicate is ascribed to a computational type, not a purely orthographic type. Thus, the meaning of the purely disquotational truth predicate is exhausted by the instances of the disquotational schema (DS): ‘ p ’ is true iff p .

This aspect of the purely disquotational theory—that it doesn't invoke a semantic notion in the characterization of truth and truth conditions—is the very reason for the modal objection. To address the modal objection, therefore, the purely disquotational theory has to invoke a semantic notion in the characterization of truth and truth conditions. The question is whether and how this can be done in a way that doesn't cause a conflict with the purely disquotational conception of truth. Recall that Field characterizes purely disquotational truth as a use-independent notion. In order to avoid an inconsistency, Field tried to invoke a semantic notion in such a way that the notion of truth is still purely disquotational at the bottom level, while the semantic notion bridges the purely disquotational notion with the non-purely disquotational notion of truth. This is the idea behind the quasi-disquotational solution, but it doesn't seem to work.

¹¹⁸ He abandons in the postscript to “Deflationist Views of Meaning and Content,”; see “Postscript: Deflationist Views of Meaning and Content,” in his *Truth and the Absence of Fact*.

¹¹⁹ H. Field, “Deflationist Views of Meaning and Content,” in his *Truth and the Absence of Fact*, p. 132.

So, the pure disquotationalist has to find an alternative way in which he can invoke a semantic notion in the characterizations of truth and truth conditions without causing an inconsistency within the theory. In a sense, it seems that the pure disquotationalist has arrived at an impasse. Field's initial idea was that the notion of truth ascribed to sentences of almost any kind—including purely orthographic types—can be explained in terms of the purely disquotational notion of truth ascribed to sentences in one's own language. The modal objection raises a serious doubt about the plausibility of this view. There is, however, another—even more serious—problem that challenges the idea. That has to do with extending the purely disquotational truth predicate to a sentence in another language. So, before discussing how Field modifies his initial views, I will examine another problem challenging them.

2. 3. Attributions of Truth to Sentences in Another Language

In this section, I will discuss Field's account of attributions of truth to sentences in another language. This is by no means a simple matter, because he has kept changing his views. So, unless great care is taken, a presentation of his view will be either misleading or inconsistent. Field's initial position was that the purely disquotational truth predicate cannot be ascribed to a sentence in a language that one (the truth-attributor) doesn't understand. So, his discussion on the attribution of truth to a sentence in another language was exclusively focused on how to extend the purely disquotational truth predicate to a sentence in a language that one understands. Even here, Field offered several different accounts.¹²⁰ A major change, however, was brought about

¹²⁰ Field offers two accounts—the extended disquotational and the modified disquotational theories of truth. Both of them belong to the translational approach that will be discussed in the next section. Both of them rely on the translation of a sentence in another language to a sentence in one's own language while the purely disquotational notion of truth ascribed to one's own language serves as the rock-bottom of attributions of truth. The difference

when Shapiro mounted a powerful argument against Field's theory.¹²¹ Field finally changed his position—he now grants that the purely disquotational truth predicate can be ascribed to a sentence that one doesn't understand.¹²² In doing so, however, he ends up inducing a major change in his theory. As will be discussed shortly, the change is big enough to wonder whether Field still holds onto the purely disquotational conception of truth.¹²³

The discussion of this section will proceed as follows; first, I will explain Field's original views before they encountered Shapiro's objection. Then, I will introduce Shapiro's objection and Field's response to the objection. Third and last, I will criticize Field's response to Shapiro's objection and thereby suggest a new solution that is slightly different from Field's. This new solution, if it works, will provide a way to hold onto the purely disquotational conception of truth.

2.3.1 Attributing Truth to a Sentence in a Language that One Understands

In this section, I will discuss Field's original views about extending the purely disquotational truth predicate to a sentence in another language. Since he initially denied ascribing the truth predicate to a sentence that one doesn't understand, the extensions of the truth predicate in this section will be restricted to sentences that a speaker understands.

between them is that whereas the extended disquotational approach invokes the notion of interlinguistic/interpersonal synonymy, the modified disquotational approach doesn't. The difference between these two approaches, however, isn't so important in the discussion of the problems with Field's pure disquotationalism. So, I won't discuss them. See H. Field, "The Deflationary Conception of Truth," in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 55-117, and "Deflationary Views of Meaning and Content," in his *Truth and the Absence of Fact*, Oxford University Press, 2001.

¹²¹ S. Shapiro, "The Guru, The Logician, and The Deflationist: Truth and Logical Consequence," *Nous*, 37 (1), 2003: 113-132.

¹²² See H. Field, "Postscript: Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*.

¹²³ Field doesn't say *explicitly* that he changed his views, although it is clear that he did so. See H. Field, *ibid*.

Broadly, Field offers two different kinds of accounts for an ascription of truth to a sentence in another language: the first is the *translational approach* and the other is the *non-translational approach*.¹²⁴ According to the translational approach, the truth of a sentence in another language that one understands is defined in the manner of (Tr) as follows:

(Tr) S in another language L is true_{qd} iff $\exists p$ (S is translated into ' p ' in my language, and ' p ' is true_{pd}).

According to the non-translational approach, the truth of a sentence in another language that one understands is defined in the manner of (N-Tr) as follows:

(N-Tr) S in another language L (that one understands) is true iff S .

Of these two accounts, I will explain the non-translational approach before I explain the translational approach.

The non-translational approach (N-Tr) is the result of applying the purely disquotational schema (DS)

(DS) ' p ' is true iff p

¹²⁴ See H. Field, "Deflationist Views of Meaning and Content," (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001.

directly to a sentence in another language. Field says that it is a misconception of his theory that the purely disquotational schema is applicable only to a sentence in one's own language—the schema can *directly* be applied to a sentence in any language as long as one understands the language.¹²⁵ Suppose I understand 'La neige est blanche' in French. But then, in the purely disquotational conception of truth, to say '“La neige est blanche” is true' is to say 'La neige est blanche.' So, with this approach, the notion of truth ascribed to the French sentence 'La neige est blanche' is exhausted by the biconditional 1) in the following:

- 1) 'La neige est blanche' is true iff la neige est blanche.

The biconditional 1) looks strange because it is a mixture of expressions belonging to two different languages. Except for that, this approach seems perfect in grappling with the purely disquotational conception of truth. The most important idea of the disquotational conception of truth is that truth is a purely logical notion—the truth predicate is a sort of substitutional device whose sole function is to express generalizations over sentences. The non-translational approach (N-Tr) is faithful to this core idea of the purely disquotational conception of truth. In case one doesn't have a special need to express one's understanding of a sentence in another language using a sentence in one's own language, 1) is all that is needed. According to Field, however, this approach can be employed only for a sentence in a language that one understands—it cannot be meaningfully applied to a sentence in a language that one doesn't understand. But is this

¹²⁵ H. Field, *ibid.*

restriction necessary? Especially after Field recanted his initial position in response to Shapiro's objection, it is not clear whether this restriction is needed. I will argue later that Field's theory doesn't need this restriction, and that will be my contribution to Field's theory of truth.

Despite the fact that Field is quite content with the non-translational approach, he spends a lot of time explaining the translational approach to an attribution of truth to a sentence in another language. The reason, however, isn't so much that we must understand a truth-statement such as 'La neige est blanche' is true' in terms of 2) in the following:

2) 'La neige est blanche' is true iff snow is white.

Rather, the reason is that a translational biconditional such as 2) should be explained within the framework of the purely disquotational theory of truth. This is another misconception of Field's theory of truth—to construe Field as claiming we must understand an attribution of truth to a sentence in another language in terms of a translational biconditional such as 2) above. Given that Field grants the non-translational approach, that isn't the reason why Field spends so much time with the translational approach. His reason is rather that we sometimes express an attribution of truth to a sentence in another language in terms of a translational biconditional such as 2). A translational biconditional such as 2), however, challenges the purely disquotational conception of truth, since the truth predicate involved in 2) cannot be considered the purely disquotational truth predicate. According to 2), 'La neige est blanche' has the truth condition that snow is white, but the nature of this truth-conditional relation cannot be explained

independently of a semantic notion. In this sense, the notion of truth involved in 2) isn't use-independent, and so it is at odds with the purely disquotational concept of truth.

To accommodate the notion of truth employed in a translational biconditional such as 2), Field defines the notion of truth ascribed to 'La neige est blanche' in terms of the purely disquotational notion of truth using the notion of translation as an additional resource. So, the methodology is basically the same as the one used in addressing the modal objection; that is, the truth predicate involved in 2) is the quasi-disquotational truth predicate ('true_{qd}'), and it is to be defined in terms of the purely disquotational truth predicate ('true_{pd}'), using the notion of translation as a 'limited additional resource'. This is the idea formulated by (Tr) above. The only difference between (Tr) and (QD) is that whereas (QD) appeals to a prior notion of meaning as an additional resource, (Tr) appeals to the notion of translation.¹²⁶ So, like (QD), the plausibility of (Tr) largely depends on whether the additional resource—the notion of translation—can be explained in a suitably deflationary way. More precisely, the question is the following: can the notion of translation be explained in a way that it doesn't inflate the notion of truth beyond what is allowed by the deflationary conception of truth?

In order to answer this question, it has to be made clear what it is to explain the notion of translation in a way that doesn't inflate the notion of truth beyond what is allowed by the

¹²⁶ The idea underlying (QD) (the quasi-disquotational approach) and (Tr) (the translational approach) is the same—truth is at bottom purely disquotational, and any notion of truth that isn't purely disquotational should be explained in terms of purely disquotational truth plus some additional resources. But then, why does Field invoke a prior notion of meaning in addressing the modal objection? The main reason seems to be that he wants (as he says) to mimic an inflationary notion of truth as much as possible. The intuition behind the modal objection is that the sentences in our language might have had different meaning, and thereby they might have had different truth conditions. When we make a modal statement like this, we construe sentences as purely orthographic types—otherwise, it doesn't make sense to say that the sentences in our language might have had different meanings and truth conditions. So, it is indeed impossible to appeal to the notion of translation in dealing with the modal objection—we can translate a sentence only if it is construed as an entity that has a certain meaning. That is to say, we cannot translate a sentence construed as a purely orthographic type. On the other hand, when we ascribe truth to a sentence in a foreign language, the sentence isn't construed as a purely orthographic type. See H. Field, "Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*, p. 131.

deflationary conception of truth. The deflationary conception of truth has it that truth is a purely logical concept that doesn't have an explanatory function. So, what is essential is that the notion of translation be explained in a way that the notion of truth doesn't play an explanatory role in an account of meaning and content. There is, however, a dilemma here—that is, the notions of meaning and truth are essentially connected to each other. Consider the meaning-to-truth schema (M-to-T) as follows:

(M-to-T) If x means that p , then x is true iff p ,

An instance of (M-to-T) is true *a priori*. It is a conceptual truth that if a sentence x means that p , then x has the truth condition that p . From this it follows that the truth condition of a sentence is a necessary condition for x 's meaning that p —to have the truth condition that p is essential to mean that p . Speaking at the level of understanding language, (M-to-T) implies that we must understand the truth condition of a sentence in order to understand its meaning. So, there is an inseparable connection between the notion of meaning and the notion of truth or a truth condition. The question is how to make sense of this connection without assigning an explanatory role to the notion of truth or a truth condition.

It seems that the best way to make sense of this connection in a deflationary way is to explain the nature of the connection in the way Horwich does.¹²⁷ According to Horwich, a

¹²⁷ Horwich offers a neat and simple explanation of the connection between the notion of meaning and the notion of a truth condition. According to Horwich, the meaning of an expression x reduces the use-property of the expression, $U(x)$. The use-property $U(x)$, however, doesn't strongly determine its extension; namely its reference or truth condition. So, we cannot infer the truth condition of a sentence x *directly* from the use-property $U(x)$. For example, suppose a sentence in another language x has a certain use property $U(x)$. We cannot infer its truth condition directly

deflationist can take the following two steps: first, a deflationist has to explain the notion of meaning without employing a truth-conditional notion; and second, a deflationist has to explain the connection between the notions of meaning and truth as a matter of a purely conceptual (analytic) relation.¹²⁸ Field agrees with this idea, although he doesn't put the point as clearly as Horwich does.¹²⁹ Both Horwich and Field agree that the notion of meaning should be explicated without employing a truth-conditional notion. In Horwich's account, it is explicated in terms of the notion of a use-property, and in Field's account, it is explicated in terms of the notion of a computation role; i.e., inferential/conceptual roles and indication roles of an expression. Given that the notion of a use-property includes the notion of a computation role, there is no substantive difference between their views in this regard. Field agrees with the second step as well—according to Field, as long as a deflationist doesn't explain the meaning-to-truth connection without making a natural connection between them, he will be fine. What he means by 'not making a natural connection' is that the nature of their connection is understood in a purely conceptual way.¹³⁰

But what is it exactly to explain the connection between the notions of meaning and truth in a "purely conceptual" way? Put differently, what is it exactly to explain their connection

from this use-property unless we correlate the use-property $U(x)$ with a sentence in our own language. Since the use-property—which is the meaning-property in Horwich's account—cannot of itself determine the extension of an expression, truth-theoretic notions cannot be reduced to a use-property. So, there is no inconsistency between reducing meaning to a use-property and the irreducibility of truth-theoretic notions. Once we explain a meaning-attribution of the form ' x means that p ' in terms of the notions of a use-property and translation, we "trivially" derive the truth condition of x by appealing to the Meaning-to-Truth principle. There will be more discussion on Horwich's theory of meaning in Ch. 4 of this thesis. See P. Horwich, *Meaning*, Oxford: Clarendon Press, 1999.

¹²⁸ P. Horwich, *Meaning*, Ch. 2 & 3.

¹²⁹ The difference between Horwich and Field is that Horwich explains the connection between meaning and truth as a matter of an instance of the schema (M-to-T), whereas Field doesn't—he tries to define a non-purely disquotational truth condition in terms of a purely disquotational condition plus some limited resources. He later (in the postscript to "Deflationist Views of Meaning and Content" in his *Truth and the Absence of Fact*) says that he was wrong in trying to do so. So, it seems that Field eventually accepted Horwich's approach (which may be called 'the schematic approach').

¹³⁰ See H. Field, "Deflationary View of Meaning and Content," in his *Truth and the Absence of Fact*: Oxford University Press, 2001, p. p. 132.

without making a “natural” connection between them? If deflationism about truth requires that the connection between meaning and truth be explained in a “suitably deflationary” way, and if this in turn requires that the nature of their connection be explained in a purely conceptual way, it is important to be clear about what that implies. Field himself considers several different ways in which we can explain the nature of the connection between meaning and truth in a purely conceptual way; i.e., without making a natural connection between them. Broadly, there two options available to the deflationist:

- (I) There is a fact of the matter by virtue of which a sentence ‘*s*’ in another language *L* means that *p*; or
- (II) There is no fact of the matter by virtue of which a sentence ‘*s*’ in another language *L* means that *p*.

If we accept (I), we can also accept interlinguistic synonymy. In addition, if we accept (I), we can be equipped with a theoretical basis for accepting a meaning-entity. On the other hand, if we accept (II), then we cannot accept interlinguistic synonymy. In addition, if we accept (II), we cannot accept the notion of a meaning-entity. Field favors (II), but he doesn’t think that option (I) is incompatible with the deflationary conception of truth as long as the notion of interpersonal/interlinguistic synonymy is explained without invoking a truth-conditional notion.

Following Field, I will argue that both (I) and (II) are compatible with the deflationary conception of truth. Suppose we explicate—as Horwich does—the notion of

interpersonal/interlinguistic synonymy in terms of the notion of a use-property (or a computational role). In such a case, a fact of the matter by virtue of which ‘*s*’ in another language *L* has the truth condition that *p* is that ‘*s*’ in *L* and ‘*p*’ in my language have the same use-property *U(x)*. This doesn’t, however, result in a naturalistic reduction of the notion of a truth condition; namely, (I) doesn’t lead us to an inflationary account of truth of the form (IP) as follows:

(IP) ‘*s*’ in *L* is true iff *p* ↔ ‘*s*’ has the naturalistic property *F*.

First of all, a use-property is a potpourri of many distinct features of an expression; for example, assertability conditions, inferential/conceptual roles, indication relations, etc. So, from the statement that ‘*s*’ has the use-property *U(x)*, we cannot infer that ‘*s*’ has a certain naturalistic property *F*. Second, unless the use-property of ‘*s*’ in *L* is correlated with the use-property ‘*p*’ in our own language, we cannot derive the truth condition of ‘*s*’ directly from its use-property. It is one thing to say that ‘*s*’ has a certain use-property *U(x)*, and another to say that ‘*s*’ means that *p*. To say the latter, we should be able to correlate the use-property *U(x)* with a sentence ‘*p*’ in our language. In this sense, both truth conditions and meaning are *immanent* notions when the notion of meaning is explicated in terms of a use-property; we cannot talk about the meaning or truth condition of a sentence that we don’t understand.¹³¹ If the use-property of ‘*s*’ in *L* has no match

¹³¹ The point here isn’t we don’t know the meaning or truth condition of a sentence that we don’t understand (which is a tautology). The point is rather that neither the meaning nor the truth condition of a sentence is determined by its intrinsic properties alone. I am sure that this is the case with Field’s disquotationalist theory of truth, but I am less sure about whether it is with Horwich’s. One possible objection is that whether or not we can translate a sentence, it must have a certain use-property, and this use-property determines the meaning and truth condition of a sentence. The problem with this objection, however, is that Horwich denies that a use-property alone determines its truth-

with a sentence in our own language, we wouldn't be able to assign meaning or a truth condition to 's'. But then, even if the notion of meaning reduces to the notion of a use-property (as it is done in Horwich's account), we cannot read off the meaning or truth condition of a sentence from the use-property of a sentence alone. The fact that 's' in *L* has a certain use-property cannot of itself determine its meaning or truth condition—we should be able to correlate 's' with 'p' in our own language.

As a consequence, even if a deflationist accepts (I)—interpersonal/interlinguistic synonymy—he doesn't risk inflating the concept of truth. Given that the notion of meaning is explicated in terms of many distinct features of an expression, and, in particular, meaning is a more or less immanent notion in deflationism, there is no way to make a “natural” connection between '*x* means that *p*' and '*x* is true iff *p*'. The fact that a sentence in another language has a naturalistic property *F* cannot of itself determine either its meaning or truth condition. But then, the simplest way to explain the connection between meaning and a truth condition is to explain it as a matter of the trivial, conceptual connection expressed by the schema (M-to-T).

Field, however, chooses an even safer option—option (II). It is beyond the scope of this section to provide a full examination of Field's account of meaning and content (see Ch. 4 for more details). To give a brief account, though, there seem to be two reasons why Field is skeptical about interpersonal/interlinguistic synonymy. First, Field explicates the notion of meaning in terms of the meaning-characteristics of an expression, where this notion is further

conditional property (see Ch. 4 for a more detailed discussion. See also P. Horwich, *Meaning*, Ch. 2, 3, and 10). He says that we can read off neither the meaning nor the extension of a predicate from its use-property alone. This can only be understood as saying that unless we can match the use-property of an expression in another language with the use-property of an expression in our own language, we can neither ascribe a meaning nor an extension to a predicate. Therefore, neither Horwich nor Field seems to consider the meaning or truth condition of a sentence as being determined by its intrinsic properties alone. But then, both meaning and truth conditions should be immanent notions in their theories. The problem I try to solve, however, is how to recognize this feature of deflationism without making the notion of truth *per se* an immanent notion.

explicated in terms of the computational role of an expression; i.e., inferential/conceptual roles and indication relation.. A computational role, however, is a very fine-grained notion; in particular, the inferential/conceptual roles of an expression can vary from person to person even within a language. For example, the computational role of ‘noodle’ in Sally’s idiolect may be different from the computational role of ‘noodle’ in Tom’s idiolect even if both of them are English speakers. Between speakers of different languages, the gap may well be even bigger. For example, the Chinese word ‘麵’ is normally translated into ‘noodle’ in English. But the computational role of ‘麵’ in a Chinese speaker’s idiolect is quite different from the computational role of ‘noodle’ in an English speaker’s idiolect.¹³² But then, in Field’s theory, what we mean by a meaning-attribution such as

3) ‘麵’ in Chinese means <noodle>,

is 4) in the following:

4) The meaning-characteristics of ‘麵’ in Chinese are most *similar* to the meaning-characteristics of ‘noodle’ in English.

¹³² First of all, Chinese people (in the past and maybe even now in rural places) use noodles as glue. So, the inferential/conceptual roles of ‘麵’ are different from those of ‘noodles’ in English. Second, their noodles look very different—if we say that noodles can have the shape of shell, many Chinese people will laugh at us. There are numerous examples of this kind. The point is not merely that cultures are relative—physical circumstances are sometimes very different alos. That means what is indicated by a word *x* in another language that is translated to, for example, ‘tree’ in English can indicate things that are quite different from what is indicated by ‘tree’ in English. But we still have to translate the word into ‘tree’, since there is no alternative. Translation is a way of trying to understand another language by finding commonalities. Finding commonalities, however, is a matter of matching and tailoring that can be done only from a translator’s point of view. There is a saying that translation is treason, of which the point is just to emphasize how difficult it is to translate one language to another.

So, by denying interpersonal/interlinguistic synonymy, Field defends the idea that a meaning-attribution should be understood in terms of the “similarity”, not sameness, of meaning-characteristics.

There is, however, another reason why Field is skeptical about interpersonal/interlinguistic synonymy—he views meaning as not fully factual but partly evaluative notion.¹³³ Field defends a kind of projectivist theory of meaning and content that is found in the theories of Quine, the simulation theory of mind, and others.¹³⁴ According to this theory, when we attribute meaning and content to another person’s utterances and mental states, we project our own mental states—we understand another person’s utterances and mental states by “empathizing” with the person’s mental states. To a certain extent, the idea seems like a commonsense. It is inevitable that we project our own mental states in translation and interpretation, because we use sentences from our own language when we translate sentences in another language. A language, however, isn’t just a collection of words that contain information about the world; it embodies a culture, ways the speakers of the language look at the world, logical principles they accept, and so forth. Similarly, in interpreting other people’s mental states, we often use ourselves as a model. We infer that a person is in mental state *M* by observing behavior *B* because we are ourselves in the mental state *M* when we exhibit that behavior.

¹³³ H. Field, “Deflationist Views of Meaning and Content” and “Attributions of Meaning and Content” in his *Truth and the Absence of Fact*. There will a more detailed discussion on this distinction in Ch. 4.

¹³⁴ Field says in “Postscript: Mental Representation” (in his *Truth and the Absence of Fact*, p. 68) that he now takes the Quinean projectivist view more seriously. Field discusses Quine’s projectivist view of psychological explanations in his “The Deflationary Conception of Truth,” in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 1986, pp. 55-117. Field also mentions R. Gordon’s “Folk Psychology as Simulation,” *Mind and Language* 1, pp. 158-171. Putnam also emphasizes the role played by “empathy” (Verstehen) in semantics and social sciences in his *Meaning and the Moral Sciences*, Routledge Kegan & Paul, 1978. See also W. V O. Quine, *Word and Object*, Cambridge, Mass.: M.I.T. Press, 1960.

If it is inevitable, at least to a certain extent, that we project ourselves in translation and interpretation, in order to claim that there is a fact of the matter in meaning, it is essential to show that there is interlinguistic/interpersonal synonymy. Consider ‘La neige est blanche’ in French. We translate it into ‘Snow is white’, and thus say that it means that snow is white. But the same sentence in French is translated into many different sentences in many different languages. In order to say that there is a fact of the matter, we have to show that all the different sentences translating ‘La neige est blanche’ mean the same thing. Unless this is shown, we have to conclude that it is one thing that we translate it into ‘Snow is white’, and thereby claim that ‘La neige est blanche’ means that snow is white; it is another that Germans translate it into ‘Schnee ist weiss’, and thereby claim that it means that Schnee ist Weiss; it is yet another that Koreans translate it into ‘눈은 하얗다’, and thereby claim that it means that 눈은 하얗다; and so on.¹³⁵ Each of these translations is considered “correct”, but if we cannot show that all these sentences mean the same, we have to accept a relativistic conclusion of some sort. In other words, although all of those translations and meaning-attributions are correct, they are not about exactly one and the same fact; there isn’t “the” meaning-fact that all of these meaning-attributions are about.

It is important, however, not to exaggerate this relativistic aspect of Field’s notion of meaning. Although Field denies that there isn’t “the” correct translation, he still grants that there are standards for translation—some translations are better than others.¹³⁶ Most of all, before we

¹³⁵ Strictly speaking, the phrases such as ‘means that Schnee ist Weiss’ and ‘means that 눈은 하얗다’ should be written in German and Korean respectively.

¹³⁶ To say that some translations are better than others is to say that a translation that is more similar in respect of meaning-characteristics is better than a translation that is less similar in respect of meaning-characteristics. So, there is an objective standard. However, the standards for translation and interpretation are not exhausted by an objective standard; they also have a non-objective standard such as an interpreter’s background beliefs and values. In this

project our background beliefs, we have to examine the meaning-characteristics of the foreign expression. It is a more-or-less objective matter to examine the meaning-characteristics of an expression, since it consists in examining the inferential/conceptual roles and indication relations of an expression.¹³⁷ Second, there should be a sufficient degree of the similarity of meaning-characteristics. For example, we cannot translate ‘pomme’ in French into ‘banana’ in English because there is a word in English—‘apple’—that is more similar in meaning-characteristics. So, even if there is relativism in Field’s view of meaning, it isn’t too rampant. The kind of semantic relativism defended by Field is that a meaning-attribution has a third parameter—the background beliefs and values that are implicitly accepted by a translator or an interpreter.

Unfortunately, the relativistic side of Field’s theory of meaning is often exaggerated by some of his critics. For example, Kunne states that Field’s “idiolectic disquotationalism” (as Kunne calls the theory) cannot explain how there can be a contradiction in beliefs in the following case.¹³⁸ Suppose someone uttered ‘La neige est blanche’. Suppose further that Sally responds to it by saying ‘That’s true’, whereas Tom responds to it by saying ‘No, that’s not true’. Listening to their conversation, we think that Sally and Tom contradict each other. Kunne’s objection to Field’s idiolectic disquotationalism is that Field’s theory cannot explain how Sally and Tom contradict each other. Because Field denies interpersonal synonymy, there is no guarantee that Sally’s understanding of ‘La neige est blanche’ is the same as Tom’s understanding of this sentence. More precisely, even if both Sally and Tom translate ‘La neige

sense, there isn’t “the” correct translation or interpretation. See H. Field, “Deflationist Views of Meaning and Content,” (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001.

¹³⁷ In other words, it consists in examining the inferential patterns of an expression and the causal connections between an expression and the world. To examine these features of an expression is a more or less objective matter.

¹³⁸ W. Kunne, “Disquotationalist Conceptions of Truth,” in *What is Truth?* Ed., R. Schantz, Berlin; New York; de Gruyter, 2002.

est blanche' into 'Snow is white', there is no guarantee that 'Snow is white' in Sally's idiolect means the same as 'Snow is white' in Tom's idiolect. So, in Field's theory, when Sally said 'That's true', she said that 'Snow is white_{sally}' is true, and when Tom said 'That's not true', he said that 'Snow is white_{tom}' is not true. Without assuming interpersonal synonymy, we cannot be sure that 'Snow is white_{sally}' means the same as 'Snow is white_{tom}'. Therefore, Field's theory cannot explain how Sally and Tom contradict each other.

The problem with Kunne's argument is that he misses a crucial point about Field's skepticism about meaning—the reason why Field is skeptical about synonymy is that meaning and synonymy are “evaluative” notions. The details of Field's view on this point will be explained in Ch. 4. A core idea of this view, however, is that a meaning-attribution or synonymy-claim should be understood as it is relativized to a meaning-attributor's personal goal or interest. Surely, this is a controversial idea, but this isn't the place to discuss the controversies concerning the view. Apart from the controversies, however, Field can explain the contradiction between Sally's and Tom's claims by relativizing them to a listener. First, by listening to the conversation between Sally and Tom, we infer 5):

5) Sally said that 'Snow is white' is true, and Tom said that 'Snow is white' is not true.

Second, unlike Kunne, we take 'Snow is white' to abbreviate 'Snow is white_{me}'. So on this account, 5) has to be analyzed into 6):

6) Sally said that ‘Snow is white_{me}’ is true, and Tom said that ‘Snow is white_{me}’ is not true.

But then, we don’t need to assume interpersonal synonymy in order to make sense of the contradiction between Sally’s and Tom’s claims. As Kunne says, in Field’s disquotationalism, we have to construe Sally’s truth-claim ‘‘Snow is white’ is true’ in terms of ‘‘Snow is white_{sally}’ is true’; similarly, Tom’s truth-claim ‘‘Snow is white’ is not true’ in terms of ‘‘Snow is white_{tom}’ is not true’. When *we* hear them talking, however, we interpret ‘Snow is white’ in terms of a sentence in our own idiolect—‘Snow is white_{me}’, and ascribe the truth predicate to this sentence. So, even if the meaning of ‘Snow is white_{sally}’ \neq the meaning of ‘Snow is white_{tom}’, we can make sense of the contradiction between Sally’s and Tom’s claims.

That said, it is still perfectly possible that the meaning-characteristics of ‘Snow is white_{me}’ in each person’s idiolect are “very similar.” My understanding of ‘Snow is white’ isn’t exactly the same as your understanding of ‘Snow is white’, but they must have similar meaning-characteristics. That is inevitable, because if they didn’t, we would not be able to get what we want—for example, I say ‘I want an apple’, but you would give me a banana; I tell you that it will rain tomorrow, and you would prepare a picnic in a park; and so forth. In other words, unless the inferential patterns and indication relations of personalized/computational types (p_{tom} , p_{mary} , etc.) are sufficiently similar, linguistic communications would fall into pieces. So, even if there is no objective fact in the sense that an expression has “the” meaning, there is still some objective basis for a meaning-attribution that Field should recognize. He does, but he should have emphasized it more clearly. In order for us to have linguistic communications, our understanding of a language should have some common basis. We can grant this common basis

even if we don't accept interpersonal/interlinguistic synonymy. But then, the difference between those who accept interpersonal/interlinguistic synonymy and those who don't isn't so great.

2. 3. 2 Attributions of Truth to Sentences in a Language That One Doesn't Understand

As was discussed in the previous section, Field's account of attributions of truth to sentences in another language doesn't have a serious problem as long as we ascribe the truth predicate to sentences in a language that we understand. An attribution of truth to a sentence in another language can be accommodated by Field's theory in terms of either (N-Tr)—by applying the disquotational schema directly to a sentence in another language—or (Tr)—by invoking the notion of translation. Field's account, however, runs into a serious problem in accommodating an attribution of truth to a sentence in a language that one doesn't understand. As was mentioned earlier, Field wasn't much concerned about this problem—he declared that the purely disquotational truth predicate cannot be extended to a sentence in a language that one doesn't understand.¹³⁹ He simply didn't see a reason for which we extend the purely disquotational truth predicate to a sentence that we don't understand. If the whole purpose of employing the truth predicate is to express generalizations over sentences, we have no reason to extend the purely disquotational truth predicate to sentences that we don't understand. This idea, however, was challenged by Shapiro's thought-experiment. Shapiro's thought-experiment goes as follows.¹⁴⁰

¹³⁹ H. Field, "Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*.

¹⁴⁰ S. Shapiro, "The Guru, The Logician, and The Deflationist: Truth and Logical Consequence," *Nous*, 37 (1), 2003: 113-132. Indeed, Shapiro not only criticizes Field's disquotationalism but also Horwich's minimalism with regard to the issue of ascribing the truth predicate to a sentence that a truth-attributor doesn't understand. Shapiro says that Horwich's theory is no better than Field's in addressing this problem.

Suppose there is a guru who makes pronouncements about set theory; suppose further that there is a disciple who thinks that everything the guru says is true, but who doesn't understand set theory; last, suppose that there is a logician who distrusts the guru's set-theoretical pronouncements, but likes to draw their number-theoretic consequences, which the disciple understands. If the disciple trusts the logician's acumen about logic, and if ' p ' is a number-theoretic sentence that the logician proves to be entailed by the guru's set theory, then the disciple will infer that ' p ' is true, so p . Since the disciple believes everything that the guru says to be true, and consequence preserves truth, the disciple infers that ' p ' is true. According to Field, the problem posed by Shapiro's example is that "though the disciple understands this conclusion, the reasoning is blocked if we can't meaningfully apply 'true' to the guru's utterances."¹⁴¹ To put the problem more clearly, the explanandum here is how the disciple infers that p . Its explanans proceeds as follows:

- i) The logician says ' p ' is entailed by the guru's utterance ' s ' about set theory.
- ii) The disciple trusts the logician's acumen about logic.
- iii) Therefore, the disciple infers that ' p ' is entailed by ' s '. (from i) and ii))
- iv) The disciple believes that consequence preserves truth.
- v) Therefore, the disciple believes that if what the guru says about set theory is true—if ' s ' is true—then ' p ' is true. (from iii) and iv))
- vi) The disciple believes everything that the guru says. (assumption)
- vii) Therefore, the disciple believes that what the guru says about set theory—' s '—is true. (from vi))
- viii) The disciple infers that ' p ' is true. (from v) and vii))
- ix) Therefore, the disciple infers that p . (from viii) plus the disciple's knowledge of the disquotational schema)

The question here is how vii) can be accommodated by the pure disquotationalist theory.

¹⁴¹ H. Field, "Postscript: Deflationary Views of Meaning and Content," in *Truth and the Absence of Fact*, p. 147.

To respond to Shapiro's objection, Field attempts several solutions.¹⁴² The first is to construe 's' as a sentence in a "potential" extension of a truth-attributor's language. That is to say, 's' is a sentence that might exist in a truth-attributor's future language if his understanding were to improve.¹⁴³ The problem with this solution, however, is that one's language will never improve to the point that he can understand all the true sentences in the world. For example, what about a sentence that has not yet been uttered by anyone but can be potentially uttered by someone in the future? Our ordinary understanding of the notion of truth allows us to ascribe the truth predicate even to such a sentence. That is to say, we grant that even if we don't understand a sentence, it is either true or not true. Field seems to think that Shapiro's objection is concerned only with making sense of logical derivations, because at least here he recognizes the need to ascribe to a sentence that one doesn't understand. But Shapiro makes it clear that the problem with Field's disquotationalism (or deflationism in general) is that it "does not exhaust the legitimate uses of the notion of truth."¹⁴⁴ Shapiro is right about this, since even if I don't understand quantum mechanics, I sometimes need to acknowledge that a sentence belonging to quantum mechanics is true simply because most experts in the field hold that it is true. When we accept experts' opinions, we don't necessarily understand what they say; we believe what they say is true, because they are experts and we have no reason to undermine their credibility. Knowledge by testimony provides another example. Imagine that I am in a Russian law court where a witness is testifying in Russian. If I have no reason to doubt the witness's reliability, I would say that the witness's testimony is true (even if I understand no word in Russian). Field's solution doesn't work in these cases, since when we say that these sentence are true, we are not

¹⁴² H. Field, *ibid.*

¹⁴³ H. Field, *ibid.*

¹⁴⁴ S. Shapiro, "The Guru, The Logician, and The Deflationist: Truth and Logical Consequence," *Nous*, 37 (1), 2003: 113-132.

saying that these sentences will be true in the future when our understanding improves. Field here changes the subject—the question is not what we *will* mean by a truth-attribution of the form ‘*s* is true’ when our understanding improves. The question is what we mean by a truth-ascription *now* when we don’t understand a sentence.

Recognizing the problem with the solution appealing to a “potential expansion” of understanding, Field offers a new solution.¹⁴⁵ His new solution is formulated in (***) as follows:

(***) If $S_{X,u}$ is translatable as ‘ p ’ then $\Box(S_{X,u}$ is true iff p)
 [And if $S_{X,u}$ is translatable as ‘ p ’ then $\Box(S_{X,u}$ has the truth condition that p).],

where ‘ $S_{X,u}$ ’ abbreviates ‘a sentence S as it is understood by an individual X in a possible world u ’, and ‘ \Box ’ means ‘necessarily’. This new solution differs from Field’s previous solutions in two respects. First, (***) is a *schema*, like the purely disquotational schema (DS) (‘ p ’ is true iff p). So, according to this new proposal, we should treat a conditional statement such as 1)

- 1) If ‘La neige est blanche’ is translatable as ‘Snow is white’, then \Box (‘La neige est blanche’ is true iff snow is white).

as an *axiom*.¹⁴⁶ By treating a truth-statement such as 1) as an axiom, Field gives up the idea that he held for a long time—that a non-purely disquotational truth predicate should be *defined* in terms of the purely disquotational truth predicate (using fairly limited resources). Thus he says,

¹⁴⁵ H. Field, “Postscript: Deflationary Views of Meaning and Content,” in *Truth and the Absence of Fact*.

I must have thought that the motivation for demanding a substantive theory of truth and truth conditions and reference, like that advocated in Chapters 1 and 2, could only be undercut if these notions are explicitly defined in terms of more-or-less logical notions plus translation. But this seems a mistake: the argument for a substantive theory of truth conditions and reference depends on taking truth conditions and reference as having a certain kind of ‘causal-explanatory’ role, as noted in the postscript to Chapter 1 ; and introducing a notion of truth conditions by means of the schema (***) does nothing to make the notion ‘causal explanatory’,¹⁴⁷

This is quite a change, since if we accept this schematic solution, we don’t need two disquotational truth predicates—the purely disquotational and quasi-disquotational. That means all the previous solutions invoking the quasi-disquotational truth predicate are no longer needed. More precisely, if we accept this schematic solution, we don’t need the quasi-disquotational solution (QD) to the modal objection (section 2.2.2) or the translational approach (Tr) that Field offered to explain a truth-ascription to a sentence in another language that one understands. All of these approaches attempt to *define* the ordinary truth predicate in terms of the purely disquotational truth predicate using some additional resources. According to Field now, however, all these approaches were based on a wrong idea—the idea that the notion of a substantive theory of a truth condition would be needed unless a non-purely disquotational truth predicate is defined in terms of the purely disquotational predicate using some additional resources. This idea is mistaken because what really matters for the deflationary theory of truth is that the notions of a truth condition and reference don’t have a certain kind of causal-explanatory role.

So, by proposing the schematic solution appealing to (***), Field induces a big change in his purely disquotational theory of truth. Originally, Field characterized purely disquotational truth as a use-independent notion; it is a notion that should be understood in a total separation

¹⁴⁶ Here, ‘La neige est blanche’ should be understood as a computational type, not a purely orthographic type.

¹⁴⁷ H. Field, “Postscript: Deflationary Views of Meaning and Content,” in *Truth and the Absence of Fact*, p. 153.

from the semantic properties of a word. It was indeed this idea that resulted in the quasi-disquotational approach (QD) and the translational approach (Tr). Field's original idea was the following. First, the notion of truth ascribed to a sentence in one's own language can be understood independently of a semantic notion, and in this sense, the notion of truth ascribed to a sentence in one's own language is purely disquotational. Second, the ordinary notion of truth isn't purely disquotational, since the notion of truth ascribed to a purely orthographic type or a sentence in another language is not use-independent; it cannot be understood in a total separation from a semantic notion. But this isn't a problem as long as the ordinary notion of truth can be defined in terms of the purely disquotational notion of truth using fairly limited resources. The underlying idea here is to separate a truth-part from a meaning-part involved in a truth-claim; the truth-part is exhaustively explained by the purely disquotational notion of truth ascribed to a sentence in one's own language, and the meaning-part is explained by invoking additional limited resources. As was discussed throughout this chapter, however, this idea turned out to have too many problems. It falls into infinite regress or circular reasoning when it is applied to the modal objection, and it cannot accommodate a truth-ascription to a sentence in a language that one doesn't understand.¹⁴⁸ So, at last, Field gives up the whole idea—the idea of defining the ordinary notion of truth in terms of the purely disquotational notion of truth using limited resources. But what does it mean to give up this idea? It means that Field gives up the use-independent feature of the purely disquotational notion of truth. If he hadn't characterized disquotational truth as a use-independent notion—a notion that is entirely separated from a

¹⁴⁸ In short, the quasi-disquotational approach falls into infinite regress or circular reasoning because of the very nature of the modal objection—the modal objection questions whether we can understand the instances of (DS) without invoking the meaning of sentences. For example, it questions whether we can make sense of ‘‘Snow is white’ is true iff snow is white’ without assuming that ‘Snow is white’ means that snow is white. The problem with the quasi-disquotational approach is that it assumes the very thing that needs to be explained—that the instances of (DS) are true.

semantic notion—he wouldn't have taken such a tortuous road. Although he doesn't say it explicitly, he makes it quite clear that he now gives up the use-independent feature of disquotational truth. He says "I now prefer to do things in the more natural way."¹⁴⁹ The more natural way is just to acknowledge that the notion of a truth condition cannot be understood independently of a semantic notion. This new idea is exhibited by the schema (***)

Given that the purely disquotational notion of truth was originally characterized as a use-independent notion, does this have to be taken as giving up the purely disquotational theory as a whole? In a sense, 'yes'—Field now has no reason to keep the "purely" disquotational notion of truth. From the broader perspective, however, this isn't a disaster at all. It just means that a version of deflationism—Field's pure disquotationalism—has to be rejected. As Field says himself above, even if the deflationary theory of truth will be sustained as long as the notions of a truth condition and reference don't have a certain kind of causal-explanatory role. So, even if a deflationist has to acknowledge the inseparable connection between the notion of a truth-condition and the notion of meaning, the deflationary conception of truth can still be sustained.

But what if the schematic solution appealing to (***) is not the best solution that can be offered by the pure disquotationalist? Although it is by no means clear whether there is a better solution than (***), it is quite clear that the solution appealing to (***) has a problem. The problem is that it still cannot accommodate an ascription of truth to a sentence in a language that one doesn't understand. According to Field, the schematic solution (***) better handles the problem of ascribing the truth predicate to a sentence that one doesn't understand, because it is noncommittal when ' $S_{X,u}$ ' is untranslatable. According to (***), if ' $S_{X,u}$ ' is translated to ' p ' which one understands, then it is necessary that ' $S_{X,u}$ ' is true iff p . But if ' $S_{X,u}$ ' is untranslatable,

¹⁴⁹ H. Field, *ibid.*

its truth condition is *indeterminate*—we cannot fill out the right-hand side of ‘iff’ in ‘ $\Box(S_{X,u}$ is true iff ___’.

It seems reasonable to consider the truth condition of an untranslatable sentence indeterminate, since even an inflationist cannot do much better than that. An inflationist—for example, a correspondence theorist or coherent theorist—may insist that the truth condition of a sentence is determined by its *intrinsic* properties; for example, the causal properties of a sentence or a sentence’s relations to other sentences. The problem with this claim is that the inflationist has to prove that the truth condition of a sentence is determined by the intrinsic properties of a sentence. In order to do better than the deflationist about the truth condition of an untranslatable sentence, the inflationist should be able to show that the notions of truth and truth conditions are *reducible* to the intrinsic properties of a sentence. As was discussed in Ch. 1, however, it is extremely difficult to show that the notions of truth and truth conditions are reducible to the intrinsic properties of a sentence. If such a reductive theory had been available, we wouldn’t now be discussing the prospect of the deflationary theory. So, even an inflationist cannot help but admitting that the truth of an untranslatable sentence is indeterminate.

When it comes to a deflationist, she has no other option but to accept the indeterminacy of the truth condition of an untranslatable sentence. A deflationist cannot explain the truth condition of a sentence by defining its truth condition of a sentence in terms of its intrinsic properties. A deflationist cannot accept an explanation of the form as follows:

(IT) x is true iff $p \leftrightarrow x$ has the property F .

To accept an explanation of this form is inconsistent with the core ideas of a deflationary theory of truth. Neither does it seem a reasonable idea to separate the notion of a truth condition from the notion of meaning as Field did before he turned to the schematic solution (***)). As was discussed, the modal objection is raised because Field characterized the notion of truth ascribed to a sentence in one's own language as a use-independent notion. Besides, it would be inconsistent to say that the truth condition of an untranslatable sentence is indeterminate if the notion of a truth condition were not essentially connected to one's understanding of the meaning of a sentence. Last and the most important, even a deflationist has to offer some account of the truth condition of a sentence. The only option available to a deflationist is to derive the truth condition of a sentence from its meaning, while he explains the relationship between them in terms of a purely conceptual connection. The idea is to accept an instance of the meaning-to-truth schema (M-to-T)

(M-to-T) If x means that p , then x is true iff p ,

as a conceptual truth. Both the meaning-to-truth schema and the schema (***) exhibit the idea that the truth condition of a sentence is derived from its meaning as a matter of a simple conceptual truth. In this way, the deflationist can avoid inflating a theory of truth while he explains the minimal, conceptual connection between the notions of meaning and a truth condition.

But if a deflationist has no other way of explaining the truth condition of a sentence except by making use of the conceptual connection between the notion of meaning and the

notion of a truth condition, then when the meaning of a sentence is indeterminate, he cannot but say that its truth condition is indeterminate as well. One may here object that it is wrong to say that the truth condition of a sentence is indeterminate just because the truth-attributor doesn't it. A deflationist about truth, however, has no other option. To say that meaning or a meaning-ascription is not immanent to one's understanding of an expression is to say that the meaning of an expression is determined by some intrinsic properties of an expression alone. To say this, however, is highly dangerous to the core ideas of deflationism. Suppose the meaning of a sentence x is determined by its intrinsic property F in the manner of (IM):

(IM) x means that $p \leftrightarrow x$ is F .

Let's add to (IM) the meaning-to-truth principle: x means that $p \rightarrow x$ is true iff p . The result would be a justification of the inflationary principle of truth (IT) mentioned above: x is true $\leftrightarrow x$ is F . Although a deflationist recognizes meaning-constituting properties (e.g. computational roles, indication relations, and use-properties), she doesn't say that we can read off the meaning of an expression from these meaning-constituting properties alone.¹⁵⁰ That is to say, even if an expression x in another language has its use-properties intrinsically, if we cannot translate it into an expression in our own language, we wouldn't be able to attribute a meaning to it.

So, it seems that the notion of meaning is essentially immanent in a deflationary theory of truth and meaning. This implies the truth condition of an untranslatable sentence is essentially indeterminate in a deflationary theory as well. So, if the notion of meaning is essentially

¹⁵⁰ See Ch. 4 on this issue.

immanent, so is the notion of a truth condition. Therefore, if we accept the deflationary notion of truth, we should treat the truth condition of a sentence that we don't understand as indeterminate.

According to Field's schema (***), however, not only the truth condition but also the truth of an untranslatable sentence is indeterminate—according to this schema, we cannot say anything about the truth of an untranslatable sentence. But then, the schema (***) cannot address Shapiro's counterexample; we cannot still make sense of the disciple's derivation of the truth of a number-theoretic sentence, which depends on the disciple's believing the truth of a set-theoretic sentence which he cannot understand. Besides, Field's solution doesn't square with the ordinary intuition of truth; we're inclined to think that a sentence can be true or not true even if we don't understand it. For these reasons, I will propose a solution that is different from Field's.

2.3.3 A Different Approach

What I will argue for in this section is that the purely disquotational notion of truth is not an immanent notion. The disquotational notion of truth is usually construed as an immanent notion—the notion of truth that can be meaningfully ascribed only to a sentence that one understands. That is also how Shapiro understands the disquotational notion of truth, and that's the whole motivation for his thought-experiment—to show that there is a serious problem with this feature of disquotational truth.¹⁵¹ In this section, however, I will argue that it isn't essential that disquotational truth be understood as an immanent notion.

¹⁵¹ As was mentioned earlier, however, Shapiro thinks that it is a problem with a deflationary notion of truth in general. See S. Shapiro, *ibid.*

Assuming that there is no way to solve Shapiro's problem while we keep the schema (***) , I propose a different schema—a schema that allows us to apply the disquotational schema to a sentence that we don't understand. The schema is (****) in the following:

(****) $\Box('S_{X,u}$ ' is true iff $S_{X,u}$).

The major difference between this schema and Field's schema (***)

(***) If $S_{X,u}$ is translatable as ' p ' then $\Box(S_{X,u}$ is true iff p)
 [And if $S_{X,u}$ is translatable as ' p ' then $\Box(S_{X,u}$ has the truth condition that p .)]

is that it doesn't invoke the notion of translation in characterizing the truth of a sentence. So, if we accept the schema (****), then we can make sense of the truth of a sentence in another language or another person's idiolect without invoking a sentence in one's own language. This difference is important in addressing Shapiro's objection. Shapiro's objection is that the disquotationalist cannot make sense of the disciple's ascription of the truth predicate to the guru's set-theoretic sentence that the disciple doesn't understand. Now, the disquotationalist can explain it as following. First, notice that the truth predicate in (****) is ascribed a sentence understood by a speaker X at a possible world u . Surely, this is also the case in Field's schema (***) , which is an important improvement from Field's initial position. Initially, Field took the disquotational schema (DS)

(DS) ' p ' is true iff p

to be the basis of the disquotational notion of truth. This idea led him to define a non-purely disquotational notion of truth in terms of the purely disquotational notion of truth plus limited

resources. By proposing the schema (***) , however, Field gave up the idea—what is at the bottom of the disquotational notion of truth is no longer the disquotational schema (DS) but the schema (***) . A big difference between (DS) and (***) is that the truth predicate in (***) is ascribed to a sentence understood by speaker X at a possible world u —the truth predicate isn't ascribed to a sentence understood by *me* at the actual world. So, (DS) is a special case of (***)—it is a case in which a speaker and an attributor are the same person.

The schema (****) respects this feature of (***)—that the truth predicate ascribed to a sentence understood by a speaker is logically prior to the truth predicate ascribed to a sentence understood by a truth-attributor. Applying this idea to Shapiro's guru-example, we first ascribe the truth predicate to a set-theoretic sentence understood by the guru as follows:

1) \square (' S_{guru} ' is true iff S_{guru}).

Second, if ' S ' is a sentence in a foreign language, and ' S_{guru} ' is translatable into ' p ' in my language, we can derive 2) from 1):

2) \square (' S_{guru} ' is true iff p).

But step 2) isn't necessary to make sense of the truth of the guru's set-theoretic sentence; it can be fully made sense of by 1). Suppose we don't understand ' S_{guru} '. We can still make sense of the truth of the guru's sentence, since the truth predicate in 1) is ascribed to a sentence understood by

the guru, and not by me. Initially, Field ascribed the disquotational truth predicate primarily to a sentence understood by me, and this generated Shapiro's objection. As a response, Field proposes the schema (***) in which the disquotational truth predicate is primarily ascribed to a sentence understood by a speaker. So, Field's schema (***) is close enough to solve Shapiro's objection *except that* it still invokes the notion of translation. By invoking the notion of translation, Field in effect nullifies the effect of ascribing the disquotational truth predicate to a sentence understood by a speaker. It is mysterious why Field still invokes the notion of translation after he liberates disquotational truth from the narrow confinement of sentences understood by a truth-attributor. It seems that Field is still in the grip of his initial idea that the disquotational truth predicate is primarily ascribed to a sentence understood by me.

The schema (****) corrects this problem with the schema (***). As a consequence, it can finally address Shapiro's objection. A drawback with the schema (****), however, is that it cannot characterize the *truth condition* of a sentence. As was stated earlier, a truth condition is an essentially immanent notion when we construe it in the deflationary sense. In his "The Deflationary Conception of Truth," Field says,

Note that as I understand a theory of truth, it could equally be called a theory of *truth-conditions*; the question that forms the subject of a theory of truth is substantially equivalent to the question of what it is for an utterance or state of thinking to have a given set of truth-conditions.¹⁵²

¹⁵² H. Field, "The Deflationary Conception of Truth," in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 1986, pp. 55-117, p. 55.

Ironically, Field's disquotational theory of truth denies the need of a theory of truth conditions that explains what it is for an utterance or state of thinking to have a given set of truth condition. No theory is needed—once we understand the meaning of a sentence, we derive its truth condition trivially by applying the meaning-to-truth schema (M-to-T): if x means that p , then x is true iff p . In Field's theory of meaning, however, the notion of meaning—or the property ' x means that p '—is itself an immanent notion. The meaning of a sentence isn't determined by its intrinsic properties alone—we should be able to interpret (or translate) it into a sentence in our own language. But then, we cannot talk about the truth condition of an untranslatable sentence; for example, we cannot make sense of the truth condition of the guru's set-theoretic sentence. In order to explain the disciple's logical derivation, however, we don't need to make sense of the truth condition of the guru's sentence; we only need to make sense of its truth. This can be done by appealing to the schema (****). Field's schema (***) is derived from (****) when we can understand ' $S_{X,u}$ '. If we cannot, we cannot derive (***) and the truth condition of the sentence should be considered indeterminate. So, by making the distinction between an ascription of truth and an ascription of a truth condition, the disquotationalist can address Shapiro's objection.

Chapter 3. Deflationism and the Success Argument¹⁵³

3.1 Introduction

A core idea of deflationism is that truth is a purely logical notion or property. One important implication of this idea is that truth shouldn't be assimilated into naturalistic properties such as being a certain valence number, being turquoise, and being sodium chloride.¹⁵⁴ Unlike these naturalistic notions, a purely logical notion cannot play a causal-explanatory role. The so-called 'success-argument' against the deflationary theory of truth is purported to challenge this very idea of the deflationary theory of truth—the idea that truth is a purely logical notion that doesn't play a causal-explanatory role.¹⁵⁵ The gist of this argument proceeds as follows. First, the notion of truth plays an important role in an account of the success of behavior. For example, if I desire to drink water, and my belief that there is a fountain across the street is true, then I will succeed in drinking water by crossing the street. Second, according to the deflationary theory of truth, truth is a purely logical notion; the utility of the notion of truth is exhausted by its role as a device of expressing generalizations. Therefore, third, the explanatory role played by the notion of truth in an account of the success of behavior cannot be accommodated by the deflationary theory of truth. So, according to the success-argument, the explanatory role played by truth in an account of the success of behavior doesn't square well with the deflationary conception of truth.

The purpose of this chapter is to examine whether and how the deflationist can accommodate the role of truth in an account of the success of behavior. For this purpose, two

¹⁵³ I use the term '*the* success argument' as meaning the argument claiming that the role played by truth in an account of the success of behavior cannot be accommodated by the deflationary theory of truth. Both Damjanovic and Kitcher use the term 'the success argument' in this sense.

¹⁵⁴ See P. Horwich, *Truth*, 2nd ed., Oxford: Blackwell, 1999.

¹⁵⁵ The expression 'the success argument' has different meanings in different contexts. In this thesis, it means what is stated above—the argument is purported to show that deflationism is false on the grounds that the role played by truth in an account of the success of behavior cannot be accommodated by the deflationary theory of truth.

quite different criticisms drawing on the success-argument will be discussed: Nic Damnjanovic's and Kitcher's criticisms of the deflationist response to the success-argument.¹⁵⁶ Although they both find the deflationist response to the success-argument unsatisfactory, they base their criticisms on very different reasons. Damnjanovic's reason is that the deflationist's response is inconsistent—the deflationist's response to the success-argument shows that truth *is* a causal-explanatory property. To support this view, Damnjanovic brings in Jackson and Pettit's program explanation; when the standard deflationist response to the success-argument is conjoined with Jackson and Pettit's theory, it shows that the notion of truth plays a causal-explanatory role in an account of the success of behavior.

Philip Kitcher has an entirely different agenda when he criticizes the deflationist response to the success-argument. As an ardent advocate of scientific realism, Kitcher's main interest is to decide whether the deflationary theory of truth is compatible with the doctrine of scientific realism. On Kitcher's view, they are not compatible—more precisely, scientific realism needs a non-deflationary notion of truth; in particular, the correspondence notion of truth. To support this view, Kitcher focuses on the role of truth in the so-called 'success-to-truth rule'. Two features of the success-to-truth rule make relevant to the discussion of deflationism and realism. First, viewed broadly, the success-to-truth rule is also an account of the success of behavior. Second, it

¹⁵⁶ Damnjanovic alludes to Field's and Horwich's responses to the success-argument in his criticism of deflationism. Kitcher is mainly concerned with Horwich's response to the success-argument; he also mentions Michael Levin and Michael Williams as those who are sympathetic to Horwich's response to the success-argument. See N. Damnjanovic, "Deflationism and the Success Argument," *Philosophical Quarterly* Vol. 155, 2005, pp. 53-67; P. Kitcher, "On the Explanatory Role of Correspondence Truth," *Philosophy and Phenomenological Research*, Vol. 62 (2) 2002, pp. 346-364; P. Horwich, *Truth*, 2nd ed., Oxford: Blackwell, 1999; H. Field, "The Deflationary Conception of Truth," in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, pp. 55-117, and "Deflationist Views of Meaning and Content," (1994), in his *Truth and the Absence of Fact*, Oxford University Press, 2001.

embodies one of the most important arguments for scientific realism—the idea defended by the no-miracle argument and the abductive argument for scientific realism. The main thesis of Kitcher’s argument is that the deflationist cannot accommodate the role of truth in this success-to-truth rule. From this, Kitcher aims to infer the conclusion that scientific realism needs a non-deflationary notion of truth.

In the following, I will criticize Damnjanovic’s and Kitcher’s arguments in turn. Unlike what Damnjanovic contends, the deflationist’s response to the success-argument doesn’t show that truth is a causal-explanatory property. There is simply no room for applying Jackson and Pettit’s theory to the deflationary theory of truth. To think differently indicates that one has misunderstood either the deflationary theory or Jackson and Pettit’s theory (or both). Kitcher’s criticism of deflationism also fails because he is confused about the goal of the success-to-truth rule. Kitcher argues that if we construe the notion of truth in the deflationary sense, we cannot explain systematic successes. To explain systematic successes, however, is not what the success-to-truth rule is supposed to do—it is supposed to defend the doctrine of scientific realism. Once we put aside Kitcher’s demand of explaining systematic successes, we will find that the role of truth in the success-to-truth rule can be accommodated by the deflationary theory of truth. Indeed, the deflationary notion of truth—contrary to what Kitcher contends—is ideally suited for the purpose for which the success-to-truth rule is put forward.

3.2 Damnjanovic’s Criticism of Deflationism

3.2.1 The Standard Deflationist Response to the Success Argument

Nic Damnjanovic, in his “Deflationism and the Success Argument,” tries to show that deflationists are wrong in denying that truth is a causal-explanatory property.¹⁵⁷ Damnjanovic claims that the standard deflationist response to the success argument is *inconsistent* because Jackson and Pettit’s program explanation shows that even logical properties can be causally explanatory.¹⁵⁸ The first half of this chapter will be devoted to examine the plausibility of this view. As a first step of this task, I will in this section explain the standard deflationist response to the success argument that Damnjanovic accuses of being inconsistent. Before embarking on this task, I will clarify a small terminological issue. Damnjanovic uses the expression ‘*causal-explanatory*’ in discussing the deflationist response to the success argument. He doesn’t explain, however, what is exactly meant by “causal” here; neither does anyone who uses this expression in the discussion of the success argument against deflationism.¹⁵⁹ The concept of causation, however, is extremely complicated, and I am disinclined to getting entangled into the discussion of this difficult concept. So, in using the notion of a causal-explanatory role in the following, I will assume only one thing: a concept has a causal-explanatory role if and only if its role isn’t purely logical. In other words, a causal-explanatory property isn’t a purely logical property; a causal-explanatory property is a non-logical/natural property such as being turquoise or being made of tin.¹⁶⁰

¹⁵⁷ N. Damnjanovic, “Deflationism and the Success Argument,” *Philosophical Quarterly* Vol. 155, 2005, pp. 53-67.

¹⁵⁸ There will be a brief account of Jackson and Pettit’s program explanation in the next section.

¹⁵⁹ As Damnjanovic says, it is Field who explains the success argument using this terminology—according to Field, the success argument challenges the deflationary theory of truth by arguing for the causal-explanatory role of truth in an account of the success of behavior. Field doesn’t, however, explain what he means by ‘being causal’ here. Ironically, he later expressed skepticism about causation. See “The Deflationary Conception of Truth,” in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 55-117, H. Field, “Causation in a Physical World”, in Michael J. Loux and Dean Zimmerman (eds.), *Oxford Handbook of Metaphysics* (Oxford University Press, Oxford, 2003), pp. 435–60.

¹⁶⁰ The only way for truth to be a causal-explanatory property is that it is a natural or physical property, since only in such a case we can say that the truth of a belief *causes* the success of behavior. It is by no means clear, however, what it is for the truth of a belief to cause the success of behavior. Two problems arise: first, we have no idea how to define the notion of truth in terms of a physicalistic notion; second, even if we assume that the notion of truth can be

According to Damnjanovic, the success argument argues from the role of truth in an account of an agent's success of behavior to the conclusion that truth is a causal-explanatory property. Deflationists have typically responded to this argument in two stages as follows: first, they try to show that the truth predicate appears only as a device of generalization in an account of the success of behavior; and therefore, second, the notion of truth plays a purely logical—not causal-explanatory—role in a success-explanation. Damnjanovic finds this standard deflationist response inconsistent. To make his points more effectively, Damnjanovic recasts the success argument against deflationism in terms of five steps as follows:

- i) If *A* has true beliefs about how to get what *A* wants, *A* is more likely to get what *A* wants [platitude].
- ii) Therefore if *A* has beliefs about how to get what *A* wants which have the property of being true, *A* is more likely to get what *A* wants [(i) and pleonastic transformation].
- iii) Therefore, the property of being true is appealed to in a causal-explanatory generalization [(ii) and the definition of causal-explanatory generalization].
- iv) Therefore truth is a causal-explanatory property [(iii)].
- v) Therefore deflationism is false [(iv) and the definition of deflationism].

It is hardly imaginable that someone should raise an objection to the inference from (i) to (iii).

After all, we do appeal to the property of truth in order to explain the success of an agent's

behavior, and neither does any deflationist as far as I know deny the intuition underlying the

inference from (i) to (iii).¹⁶¹ So, the potentially questionable step is the inference from (iii) to (iv).

defined in terms of some sort of a physicalistic notion, we would still be left in dark about the nature of the alleged causal connection. What does it mean to say that the correspondence/causal connection between a belief state and the external world causes the success of behavior?

¹⁶¹ Some deflationists such as prosententialists hold that truth is not a property. For them, 'is true' is not even an ordinary predicate: it is a prosentence-forming operator. Horwich, however, doesn't deny that truth is a property, since 'true' should be formalized in logic as a predicate. To Horwich whether truth is a property or not is a less interesting question than the question of what *sort* of property truth is. What should be emphasized, thus, is that truth

The standard deflationist response blocks the inference from (iii) to (iv); it accepts (iii), but doesn't accept (iv). According to this response, the property of truth is appealed to as a device of generalization, so to accept (iii) doesn't lead to accepting (iv). Damnjanovic disagrees; he claims that the deflationist can accept the inference from (iii) to (iv). Surprisingly, Damnjanovic doesn't think the claim (iv)—that truth is a causal-explanatory property—to be incompatible with the deflationary theory of truth. This is a very peculiar idea, since most leading deflationists think that truth is a purely logical notion, and they also think that if truth is a purely logical notion, truth cannot be a causal-explanatory property. Peculiar or not, Damnjanovic's very purpose is to show this seemingly impossible combination is possible. To be more precise, Damnjanovic wants to show that there is no inconsistency between the following two claims a) and b):

- a) The deflationary theory of truth is correct.
- b) Truth is a causal-explanatory property.

I am too conservative to agree with Damnjanovic—I don't think a) and b) both can be sustained. There are, however, things to be done before I mount criticisms of Damnjanovic's view; first, there should be a more detailed explanation of the standard deflationist response to the success argument; and second, there should be a scrutiny of Damnjanovic's arguments.

As was mentioned, the standard deflationist response that Damnjanovic has in mind is Horwich's response to the success argument. The upshot of Horwich's response to the success argument is that although the property of truth is appealed to in a causal-explanation of the

is not a *complex* or a *naturalistic* property, unlike properties such as being a tree, being a dog, having a certain valence, and etc.

success of behavior, all the work done by the property of truth can be exhaustively explained by assuming no more than what is implied by an instance of the equivalence schema (E) below:

(E) The proposition that P is true iff P .

Suppose that we endorse a statement which goes as follows¹⁶²:

- 1) If Bill desires to have a beer, and he believes that merely by nodding he will get one, then if his belief is true, he will get what he wants.

According to Horwich, the following derivation suffices as an account of the role of truth in the success-explanation 1):

- a) Bill wants <Bill has a beer>¹⁶³.
- b) Bill believes <Bill nods \rightarrow Bill has a beer>.
- c) [Bill wants <Bill has a beer> & Bill believes <Bill nods \rightarrow Bill has a beer>] \rightarrow Bill nods.
- d) Bill nods.
- e) Bill's belief is true.
- f) <Bill nods \rightarrow Bill has a beer> is true.
- g) Bill nods \rightarrow Bill has a beer.
- h) Bill has a beer.
- i) <Bill has a beer> is true iff Bill has a beer
- j) <Bill has a beer> is true.

a), b), and c) are assumptions. Given these assumptions, d) is derived. e) is another assumption; in addition to these assumptions, let's further assume that Bill's belief b) is true. Assuming that Bill's belief b) is true and the instances of the equivalence schema (E) are true, we can derive g).

¹⁶² P. Horwich, *Truth*, 2nd ed. pp. 22-23

¹⁶³ The brackets are used to express 'the proposition that ...'.

From g) and d), we can derive h). From a) and i) + j), we can derive the conclusion that Bill achieves his desire. This exhausts Horwich's explanation of the success of Bill's behavior.

What makes this explanation deflationary is that this explanation appeals to no other theory of truth than the deflationary theory in deriving the success of the agent's behavior. The explanation from (a) to (j) appeals to the truth of the agent's belief in explaining the success of his behavior (that is the step (e) above). As we can see in the steps (f) and (g), its role is exhaustively explained in terms of what is implied by an instance of the equivalence schema (E) in the following:

(E) $\langle P \rangle$ is true iff P .

Hence, Horwich concludes that the success argument doesn't pose a real threat to the deflationary theory of truth.

But in what sense does the deflationist hold that the truth predicate appears as a purely logical device in an explanation of the success of behavior? The deflationist response explains why we don't need to assume a theory of truth other than the deflationary one in order to explain the role of truth in a success-explanation. It doesn't explain, however, why the deflationist objects to the inference from (iii) to (iv) of the success argument against deflationism. The reason has to do with role of truth in a generalized success-explanation such as 2) in the following

- 2) An agent's behavior is more likely to succeed in satisfying the agent's goal when the agent has a true belief.

When the truth predicate is ascribed to a single belief that is articulated as it is in the explanation 1) above, the deflationist says that the truth predicate is eliminable. We can see this in the derivation (e) to (g) above. But when the truth predicate appears in a generalized success-explanation such as 2), the truth predicate is not eliminable. According to the deflationist, the sole utility of the notion of truth lies in enabling us to express a generalization such as 2) above. The deflationist explains the role of truth in a generalization such as 2) in terms of the following steps. First, 2) abbreviates an infinite conjunction such as 3) in the following¹⁶⁴:

- 3) If one believes <one nods \rightarrow one gets a beer> and one nods \rightarrow one gets a beer, then one is more likely to succeed in achieving one's goal by nodding; if one believes <one exercises

¹⁶⁴ Gupta objects to this very idea that the notion of truth enables us to express infinite conjunctions and disjunctions (see his "A Critique of Deflationism" *Philosophical Topics*, Vol. 21, 1993). Gupta's criticism, however, seems to be based on a misunderstanding of deflationism. According to Gupta, in order for the deflationist to make this claim, she needs to say that a generalization such as 2) *means the same as* an infinite conjunction such as 3). The deflationist, however, doesn't have to make such a strong claim. All she needs to say is that the generalization 4) abbreviates the infinite conjunction 5) in the same way that the generalization i) abbreviates the infinite conjunction ii) in the following:

- i) (x) (x is human \rightarrow x is mortal)
- ii) If Madonna is human \rightarrow Madonna is mortal; if Lady Gaga is human \rightarrow Lady Gaga is mortal; and so on.

In order to say that i) abbreviates ii), we don't have to say that i) means the same as ii); all we need to say is that i) is materially equivalent to ii). Nobody would say that i) means the same as ii), but everybody accepts that i) is a generalization of ii). But then, why does Gupta say so? Gupta seems to be confused between the two slightly different but related aspects of deflationism. The first is the deflationist's claim that the truth predicate is eliminable when it is ascribed to a single sentence (or proposition) that is explicitly given. For example, the truth predicate in iii) in the following is eliminable:

- iii) 'Snow is white' is true.

But then, iii), as Field says, iii) is cognitively equivalent to the truth-less statement iv) as follows:

- iv) Snow is white.

The second aspect of deflationism is that the truth predicate isn't eliminable when it appears in a generalization such as 2). But if the truth predicate is *not* eliminable in a generalization such as 2), then 2) cannot be considered cognitively equivalent to the infinite conjunction 3). So, it seems that Gupta wrongly assimilates the second aspect of deflationism into its first aspect. These two aspects are related because both of them are grounded on the equivalence schema (E). But they should be distinguished from each other for a proper understanding of the deflationary theory of truth.

→ one becomes healthier> and one exercises → one becomes healthier, then one is more likely to succeed in achieving one's goal by exercising; and so on.

As Quine says, second, we cannot generalize 3) using ordinary objectual quantification because of some technical difficulties.¹⁶⁵ We may appeal to non-objectual quantification—i.e., substitutional quantification—in the way below:

- 4) (Πp) (If one believes that p and p , then one is more likely to succeed in achieving one's goal).

Even if we accept 4) as a formal generalization of 3), we still need an informal generalization that is couched in ordinary language. The truth predicate serves this purpose; indeed, the truth predicate may have been created exactly for such a purpose. Viewed this way, it is clear why we need to assume nothing more or less than what is implied by an instance of the equivalence schema (E). Here, I will quote one of the most often-quoted passages in the literature of deflationism:

It is not surprising that we should have use for a predicate P with the property that " '____' is P " and "____" are always interdeducible. For we frequently find ourselves in a position to assert each sentence in a certain infinite set z (e.g., when all the members of z share a common form); lacking the means to formulate infinite conjunctions, we find it convenient to have a single sentence which is warranted precisely when each member of z is warranted. A predicate P with the property described allows us to construct such a sentence: $(x)(x \in z \rightarrow P(x))$. Truth is thus a notion that we might reasonably want to have on hand, for expressing semantic ascent and descent, infinite conjunction and disjunction.¹⁶⁶

¹⁶⁵ There are two well-known technical difficulties; first, the second occurrence of 'one nods → one gets a beer' or 'one exercises → one becomes healthier' is a sentence, but a sentence is not a name. So, it cannot be replaced by an objectual variable. Second, the same sentence-type occurs both inside and outside a 'that'-clause in 3). See W. O. Quine, *Philosophy of Logic*, Eaglewood Cliff; Prentice-Hall, 1970.

¹⁶⁶ S. Leeds, "Theories of Reference and Truth," *Erkenntnis* 13 (1978): 111-29, p. 121.

What Leeds says above offers a clear and intuitive account of why we consider the instances of the equivalence schema (E) epistemologically fundamental. Given the genetic explanation offered by Leeds, we cannot but consider the instances of the schema (E) conceptually fundamental.

As was stated in the footnote above, the second aspect of the deflationary conception of truth is that the notion of truth is ineliminable when it is invoked in a generalization such as 2). According to the deflationist, the truth predicate exists solely for this purpose. So, although the truth predicate appears in a generalized success-explanation such as 2), its role is purely logical—it appears merely as a device of generalization. This is the reason why the deflationist doesn't accept the inference from iii) to iv). The deflationist explains why the truth predicate is appealed to in a causal-explanatory generalization while she construes truth as a purely logical property. So, she doesn't accept iv).

3.2.2 Damnjanovic's Criticism of the Standard Deflationist Response

A. The Distinction Between Causal Efficacy and Causal Relevancy

The main inspiration of Damnjanovic's argument comes from Jackson and Pettit's distinction between causally *efficacious* properties and causally *relevant* properties.¹⁶⁷ I have no intention to go into the details of their account except for noting that the distinction between causally efficacious properties and causally relevant properties was drawn as a means of

¹⁶⁷ Frank Jackson and Philip Pettit, "Causation in the Philosophy of Mind," *Philosophy and Phenomenological Research*, Fall, 1990

vindicating the causal relevance of “functionally defined higher-order properties” (of which the particular emphasis being laid upon commonsense psychological properties).¹⁶⁸ According to Jackson and Pettit, higher-order functional properties are not themselves causally efficacious properties but they are causally relevant properties in the sense that they supervene on some lower-level basic properties which are themselves causally efficacious. The examples presented by Damjanovic are properties such as being a sleeping-pill and having a pain. Let’s suppose, for example, Ann ingested a sleeping-pill and is now face down in her bowl of soup. By pleonastic transformation, we can also say that she is face down in her soup-bowl because she ingested something having the property of being a sleeping-pill. However, there arises a question whether we can say that the property of being a sleeping-pill is a causal-explanatory property, in other words, whether it played a role in causing Ann to be face down in her soup-bowl now.

Damjanovic’s answer, as may be expected, is that even though the property of being a sleeping-pill is not itself a causally *efficacious* property, it is a causally *relevant* property. Damjanovic says that we cannot suppose that the property of being a sleeping-pill—in addition to the lower-level property the property of being a sleeping pill supervenes on—is causally efficacious because to do so would result in the problem of causal over-determination. The problem of causal over-determination occurs when we suppose that there are two distinct properties both of which are causally efficacious in bringing about an event or a phenomenon simultaneously, where each property alone is sufficient in bringing about the event or phenomenon at the time. Assuming that this is a legitimate problem, we cannot say that both the property of being a sleeping pill and its supervenient basis are causally efficacious, because a property is causally efficacious if it doesn’t need another property in order to cause an event or

¹⁶⁸ J. Kim offers a good criticism of Jackson and Pettit’s program explanation. See J. Kim, *Mind in a Physical World*, London: Massachusetts Institute of Technology, 1998.

phenomenon.¹⁶⁹ A causally efficacious property is one that can alone cause an event or a phenomenon. Hence, to say that both the property of being a sleeping pill and its supervenient basis are causally efficacious is to say that there are two distinct properties both of which are sufficient in causing Ann's facing down in her soup-bowl. Therefore, to avoid the problem of causal over-determination, we should say that only one of them is a causally efficacious property. In this case, we should say that the supervenient basis of the higher-order property—the property of being a sleeping pill—is the causally efficacious property.

Nonetheless, the property of being a sleeping-pill, according to Jackson and Pettit's account, can still be said to be causally explanatory because the higher-order property of being a sleeping pill "ensures" that there is a lower-level, chemical, property that instantiates the higher-order property. In Jackson and Pettit's account, the fact that a higher-order property isn't reduced to a lower-level property confers an explanatory power on the higher-order property. Their argument is that if a higher-order property can be instantiated by distinct lower-level properties, we cannot explain the explanatory power of the higher-order property in terms of the explanatory powers of lower-level properties that instantiate the higher-order property. In particular, they argue that an explanation that invokes a higher-order property provides the kind of modal information that an explanation invoking only a lower-level property cannot supply.¹⁷⁰ It is possible that the same higher-order property be instantiated by a lower-level property that is different from what underlies it now. Hence, the information provided by an explanation such as 1) below is different from the information provided by an explanation such as 2):

¹⁶⁹ See F. Jackson and P. Pettit, "Program Explanation: A General Perspective," *Analysis*, Vol. 50. No. 2, 1990: 107-117.

¹⁷⁰ A higher-order property can be instantiated by different lower-level properties at different possible worlds. So, an explanation that invokes a higher-order property conveys information that the same phenomenon can be caused by different lower-level properties at different possible worlds. See F. Jackson and P. Pettit, "Causation in the Philosophy of Mind," *Philosophy and Phenomenological Research*, Fall, 1990.

- 1) Ann is face down in her soup-bowl because she ingested something that has the property of being a sleeping-pill.
- 2) Ann is face down in her soup-bowl because she ingested something that has the chemical property *C*.

The information provided by 1) cannot be reduced to what is provided by 2), because the higher-property of being a sleeping pill may be instantiated by a lower-level property that is distinct from the chemical property *C*. The presence of the higher-order property of being a sleeping pill, however, ensures the presence of a lower-level property (whatever it might be). In this sense, the higher-order property of being a sleeping pill is causally “relevant” to the effect although it is not causally efficacious.¹⁷¹

Now, the upshot of Damnjanovic’s argument is that the deflationary concept of truth can be understood in terms of Jackson and Pettit’s program explanation. More precisely, his idea is that, first, we can understand the deflationary concept of truth in terms of a supervenience relation, and second, we can justify the causal relevance of deflationary truth by applying Jackson and Pettit’s program explanation to the property of truth. So, Damnjanovic’s goal is to prove that deflationary truth is a causally relevant property by applying Jackson and Pettit’s program explanation to the deflationary concept of truth. But then, the plausibility of Damnjanovic’s idea depends on, first, whether the deflationary concept of truth can be defined in terms of a supervenience relation, and second, whether Jackson and Pettit’s program explanation

¹⁷¹ The ultimate goal of Jackson and Pettit’s program explanation is to justify the causal-explanatory role of a higher-order, mental, property in an account of behavior by utilizing the non-reductive/supervenience theory of a mental property.

is correct. The logical order of these two questions is clear—only after we answer the first question, we need to be worried about the second.

The crux of my criticism of Damnjanovic’s argument will be centered on the first question—whether the deflationary concept of truth can be understood in terms of the notion of supervenience. The notion of supervenience is a jack of all trades—as McLaughlin and Bennett say, it has been invoked in almost every corner of analytic philosophy.¹⁷² The supervenience relation that Jackson and Pettit’s theory draws upon, however, is a supervenience relation of a special kind—it is a supervenience relation that obtains between a higher-order property and lower-level properties, where the lower-level properties are physical properties.¹⁷³ So, in order to apply Jackson and Pettit’s program explanation to the deflationary concept of truth, the deflationary concept of truth should be defined in terms of the supervenience relation of this specific kind. The most serious problem with Damnjanovic’s argument, however, is that he totally misses this point—he wrongly thinks that a supervenience relation of any kind can do the trick.¹⁷⁴

B. Deflationary Truth is a Supervenient Property

According to Damnjanovic, the deflationary concept of truth can be understood in terms of the notion of supervenience because what is implied by the instances of the equivalence schema can be re-constructed in terms of the notion of supervenience. According to Horwich’s

¹⁷² B. McLaughlin and K. Bennett, “Supervenience,” *Stanford Encyclopedia of Philosophy*, 2010.

¹⁷³ Jackson and Pettit’s program explanation is a kind of physicalistic theory of a mental property (this is made clear by Jackson and Pettit themselves). So, Jackson and Pettit claim that the lower-level properties that higher-order, mental, properties supervene on are physical properties. See Jackson and Pettit, *ibid*.

¹⁷⁴ A supervenience relation in its generic sense is defined as follows: a property *A* supervenes on a property *B* if and only if there cannot be a difference in the property *A* without there being a difference in the property *B*. As McLaughlin and Bennett say, this simple slogan can be cashed out in many different ways, which results in supervenience relations of different kinds.

deflationary theory of truth, the concept of truth is implicitly defined by the instances of the equivalence schema (E):

(E) $\langle P \rangle$ is true iff P .

So, on this theory, to say that

$\langle P \rangle$ is a true proposition. .

is to say that

$\langle P \rangle$ is true = $\langle \text{Snow is white} \rangle$ & snow is white; or $\langle P \rangle$ = $\langle \text{Grass is green} \rangle$ & grass is green;

or . . .

In other words, to accept the deflationary theory of truth is to accept 1) in the following:

- 1) $\langle P \rangle$ is a true proposition iff $\langle P \rangle$ = $\langle \text{Snow is white} \rangle$ & snow is white; or $\langle P \rangle$ = $\langle \text{Grass is green} \rangle$ & grass is green; or . . .

Unlike other critics of deflationism, Damnjanovic agrees with all the core ideas of the theory; for example, he agrees with the idea that the concept of truth is implicitly but exhaustively defined by 1) above. So, he agrees with deflationists that we need to assume no theory of truth other than the deflationary one. In particular, Damnjanovic accepts the deflationist idea that whatever role played by the notion of truth can exhaustively be explained by the instances of the equivalence schema (E). So, he doesn't find Horwich's response to the

success argument to be inadequate *per se*; he thinks that a Horwich-style success-explanation is all we need for an account of the success of behavior. Damnjanovic doesn't take this, however, as showing that truth isn't a causal-explanatory property. To the contrary, he thinks that a Horwich-style success-explanation shows that truth *is* a causal-explanatory property. How on earth does a Horwich-style success-explanation show that truth is a causal-explanatory property? Damnjanovic's super-economic, low-cost strategy is to insert the concept of supervenience into Horwich's deflationary theory of truth.

Damnjanovic's idea is simple—maybe too simple. First, he claims that the deflationary definition of truth 1) above entails 2) in the following:

- 2) The truth of $\langle P \rangle$ supervenes on $\langle P \rangle = \langle \text{Snow is white} \rangle \ \& \ \text{snow is white}$; or $\langle P \rangle = \langle \text{Grass is green} \rangle \ \& \ \text{grass is green}$; or . . .

Second, from 2), Damnjanovic derives 3) in the following:

- 3) The subvenient properties of the property of truth are disjuncts such as ' $\langle P \rangle = \langle \text{Snow is white} \rangle \ \& \ \text{snow is white}$ ' or ' $\langle P \rangle = \langle \text{Grass is green} \rangle \ \& \ \text{grass is green}$ '; or . . .¹⁷⁵

¹⁷⁵ Damnjanovic says:

But in just the same way, the property of truth requires a combination of truth-bearer and state of the world. If truth is instantiated in an agent's set of beliefs, for example, then either he believes that p , and p ; or he believes that q , and q ; or In fact, the standard deflationist method for explaining the generalization linking truth and practical success relies heavily on this point. *What is more, truth's subvenient disjuncts (like believing that p when p) are, at least relatively, efficacious.* (N. Damnjanovic, "Deflationism and the Success Argument," *The Philosophical Quarterly*, Vol. 55, pp. 53-67. Emphasis is done by me.)

Third, Damnjanovic applies Jackson and Pettit's theory to 3), the result of which, according to Damnjanovic, is as follows:

In particular, the property of being in pain requires a certain combination of agent and neural event. If pain is instantiated, then either John has his C-fibres firing or Dave has his D-fibres firing, or ... But in the same way, the property of truth requires a combination of truth bearer and state of the world. If truth is instantiated in an agent's set of beliefs, for example, then either he believes that p , and p ; or he believes that q , and q ; or ... In fact, the standard deflationist method for explaining the generalization linking truth and practical success relies heavily on this point. What is more, truth's subvenient disjuncts (like believing that p when p) are, at least relatively, efficacious.¹⁷⁶

According to Jackson and Pettit's program explanation, the causal relevance of a higher-order property such as being a sleeping pill or being in pain is explained in terms of the causal efficacy of its lower-level properties on which the higher-order property supervenes. Damnjanovic's grand proposal is that the causal relevance of the property of truth can be explained in the same way—in terms of the causal efficacy of its subvenient properties. Although the property of truth doesn't directly enter into the explanation of the success of behavior, it is indirectly relevant to bringing about the success of behavior. What makes this possible, says Damnjanovic, is that deflationary truth is a sort of supervenient property.

3.2.3 The Problems with Damnjanovic's Argument

I have always wondered whether anyone attempted a supervenience theory of truth; if a reductive theory of truth is unlikely to be forthcoming, what about a supervenience theory of

¹⁷⁶ N. Damnjanovic, p. 65. Underlining by myself.

truth? A correspondence theorist may benefit himself by developing a supervenience theory of truth that explicates the nature of the correspondence relation in terms of its subvenient properties. I couldn't find anyone, however, who has made a serious attempt of this kind.¹⁷⁷ The reason may be that it is extremely difficult to develop a supervenience theory of truth. Let's suppose, as suggested by Damnjanovic, that we model a supervenience theory of truth on a supervenience theory of a mental property. A difference between the reductive theory and the supervenience theory of a mental property is that the latter allows that one and the same mental property can be realized by different lower-level physical properties in different domains. For example, the supervenience theory of a mental property may characterize the property of being in pain in the manner of 1) as follows:

- 1) x is in pain iff x is a human & x 's C-fibres are firing; or x is a Martian & x 's D-fibres are firing; or . . .

Hence, if there is a supervenience theory of truth and it is to be modeled upon the supervenience theory of a mental property, it should proceed in the manner of 2) as follows:

- 2) x is a true proposition iff x is a proposition belonging to the domain D^1 & x has the physical property F^1 ; or x is a proposition belonging to the domain D^2 & x has the physical property F^2 ; or . . .

¹⁷⁷ David Armstrong seems to defend a supervenience theory of truth of a very simple kind in connection with his theory of truth and truthmakers. Indeed, Damnjanovic's way of using the notion of supervenience is very similar to Armstrong's—both of them simply replace 'iff' with 'supervenies on'. See D. Armstrong, *Truth and Truthmakers*, Cambridge: Cambridge University press, 2004.

It is not clear what those domains D^1, \dots, D^n should be, and even less clear what those physical properties F_1, \dots, F^n should be. Most of all, in order to have a supervenience theory of truth in the form of 2), it is essential that we should be able to give a functionalist (of functional) definition of the concept of truth. Notice that the supervenience theory of mental properties such as 1) wouldn't even get off the ground unless we can give a functionalist definition of the concept of being in pain.¹⁷⁸ Only after we show that the concept of being in pain can be defined in terms of a functionalistic property ϕ , we can give a supervenience theory of the property of being in pain by showing that the functionalistic property ϕ can be multiply realized by distinct lower-level physical properties. Indeed, the very motivation of Jackson and Pettit's program explanation is the worry that the functionalist theory of mental concepts may not explain the causal relevance of mental properties if an agent's behavior can exhaustively be explained by the agent's internal physiological properties.¹⁷⁹ The program explanation is the solution offered by Jackson and Pettit—the causal relevance of mental/functionalistic properties can be justified via the causal efficacy of the lower-level physical (neural) properties. Now, Damnjanovic's goal is to defend the causal relevance of the property of truth in the way Jackson and Pettit defends the causal relevance of mental properties. But then, he should show, first, how to define the concept

¹⁷⁸ The functionalist definition of pain is to define the concept of pain in terms of its causal-functional role; for example, an organism is in pain iff it exhibits a set of behavior b^* when it receives a sensory stimulation s^* , and then move to another mental state m^* . See N. Block, "Functionalism," in N. Block, ed., *Readings in the Philosophy of Psychology*, Cambridge: Harvard University Press, 1980; N. Block, "What is Functionalism?" *The Encyclopedia of Philosophy Supplement*, 1996.

¹⁷⁹ Jackson and Pettit says:

Our problem is that if the popular functionalist approach to mental properties is correct, the very same style of argument appears to be available to cast doubt on the causal relevance of mental properties. The shadow of physiology seems to raise a problem for functionalists as well as for dualists despite the fact that functionalism is compatible with a purely materialistic view of the mind. (F. Jackson and P. Pettit, "Causation in the Philosophy of Mind," *Philosophy and Phenomenological Research*, Vol. 50, 1990, pp. 195-214, p. 1998)

of truth in terms of a certain functionalist property Ψ ; second, how the functionalistic property Ψ can be multiply realized by distinct lower-level physical properties in distinct domains.

The supervenience theory of truth advocated by Damnjanovic is basically as follows:

- 3) x is a true belief (or proposition) iff $x = \langle \text{Snow is white} \rangle \ \& \ \text{snow is white}$; or $x = \langle \text{Grass is green} \rangle \ \& \ \text{grass is green}$; or . . .

Now, let's compare 3) with 2)—what similarities do they bear? They only bear a superficial similarity—both of them are defined in terms of multiple disjunctions. Other than that, they are very different, where their difference may be summed up in terms of the following two statements:

- i) First, unlike 2), 3) doesn't define the concept of truth in terms of lower-level physical concepts; *therefore*
- ii) Second, unlike 2), 3) is not a metaphysical or ontological analysis of the concept of truth.

As a result, Damnjanovic's supervenience theory of truth cannot be combined with Jackson and Pettit's theory—it is not the kind of supervenience theory that is needed by Jackson and Pettit's theory.

As McLaughlin and Bennett say, the notion of supervenience has been used almost every corner of analytic philosophy.¹⁸⁰ So, it may not be incorrect to construe the deflationary theory of truth in terms of a concept of supervenience. A supervenience relation, however, comes in many

¹⁸⁰ B. McLaughlin and K. Bennett, "Supervenience," in *Stanford Encyclopedia of Philosophy*, 2005, 2011.

different kinds, and so it is important to invoke a supervenience relation that fits the purpose at hand. Damnjanovic's argument fails in this respect—he thinks that to invoke supervenience relation of any kind will give him what he needs. What Damnjanovic needs, however, is a supervenience relation of a specific kind—the one that is connected to *ontological dependence*. Consider 1); i.e., the supervenience theory of the property of being in pain. 1) states that the property of being in pain supervenes on lower-level physical properties (e.g. the property of C-fibres' firing) in the sense that the higher-order mental property depends for its existence on the lower-level physical properties.¹⁸¹

The question, however, is whether 3) fits this type of supervenience theory of truth. We can construe 3), as Damnjanovic insists, by invoking the notion of supervenience—that the truth of a belief supervenes on the belief is <Snow is white> & snow is white; or the belief is <Grass is green> & grass is green; or . . . This doesn't give, however, Damnjanovic what needs—the kind of supervenience invoked here is not concerned with ontological dependence. What he needs is the supervenience theory that connects the property of truth—being a higher-order property—with lower-level physical properties that the property of truth depends on for its existence. 3) can never be turned into a supervenience theory of truth of this kind, because it is the (Horwich-style) *deflationary* definition of the concept of truth. According to Horwich, the notion of truth is implicitly defined by the instances of the equivalence schema (E), and 3) embodies this very idea. Recall, however, how Horwich characterizes the notion of truth—he characterizes it as a non-natural/purely logical property which cannot be assimilated into properties such as being turquoise, being a tree, or being made of tin. So, 3) is characteristically

¹⁸¹ The nature of the ontological dependence of mental properties on physical properties should further be explained. J. Kim says that it can be explained by invoking modal operators or possible worlds. See J. Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, MIT Press, 2000, Ch. 1.

different from 1)—whereas 1) is aimed to reveal the ontological bases of the concept of being in pain, 3) isn't. What is then the aim of 3)? The statement 3) is aimed at revealing the functional property of the concept of truth—it is a *functional characterization* of the property of truth. It tells us when we can use the truth predicate 'true'—we can use the truth predicate when a belief (or proposition) is <Snow is white> & snow is white; or when a belief (or proposition) is <Grass is green> & grass is green; or . . . So, according to 3), we can derive a statement of the form '<P> is true' from a statement of the form 'P', and vice versa. According to Horwich's deflationary theory, however, this exhausts whatever role played by the notion of truth. But then, there is no point of trying to uncover the ontological bases of this concept—its role is in its nature purely logical in the sense that the role of a logical constant such as ' \wedge ' is purely logical.

So, Damnjanovic's comparison between the property of being in pain and the property of truth is incorrect—we shouldn't understand 3) on the par with the supervenience theory of the property of being in pain; i.e., the statement 1). If Damnjanovic is to vindicate the causal relevance of the property of truth in the way Jackson and Pettit vindicates the causal relevance of the property of being in pain, he should have given a supervenience theory of truth of the form 2) above. Namely, he should have attempted to connect the property of truth with lower-level physical properties with respect to different domains. I have no objection to construing the deflationary definition of truth—the statement 3)—in terms of the concept of supervenience. To do so, however, doesn't give what Damnjanovic needs. Only when he offers a supervenience theory of truth of the type 2) can he appeal to Jackson and Pettit's program explanation in order to vindicate the causal relevance of the property of truth. The real problem with Damnjanovic's overall proposal is that this can never be done if he accepts the core ideas of deflationism. Damnjanovic's overall proposal is to show that a deflationist can recognize the causal-

explanatory role of truth without stepping outside the deflationary theory of truth. This is the reason why Damnjanovic turns the deflationary definition of truth into a supervenience theory of truth leaving the core ideas of deflationism intact. Supervenience is an elastic notion, so the deflationary definition of truth may be turned into a supervenience theory of truth of *some* kind. But the kind of a supervenience theory of truth that Damnjanovic needs is the supervenience theory of truth of the type 2). Only in such a case can he plug Jackson and Pettit's theory into the property of truth. A supervenience theory of truth of the type 2), however, is inconsistent with the Horwich-stype deflationary theory of truth. Someone can pursue a supervenience theory of truth of the type 2) only if he assumes that truth is a natural or physicalistic notion. Horwich denies this, which is implied by his contention that truth is a purely logical property.

In conclusion, Damnjanovic makes a typical mistake in dealing with the deflationary conception of truth—he fails to understand what it means to say that truth is a *purely* logical property. He is right in saying that a logical property can also have a causal-explanatory role. The deflationist's view of truth, however, isn't simply that truth is a logical property but that truth is nothing but a logical property, implying that it isn't a causally explanatory property. In order to see the point let's compare 3) with 4) in the following:

- 4) x is in pain iff x produces the output O^1 under the input I^1 & goes to another mental state M^1 ; or x produces the output O^2 under the input I^2 & goes to another mental state M^2 ; or . . .

Like 3), 4) is a functional characterization of the property of being in pain. Indeed, if Damnjanovic is to make an analogy between the concept of truth and the concept of being in

pain, he should have begun by comparing 3) with 4), not with 1). Although 3) and 4) are both functional characterizations, their epistemological statuses are different—whereas 4) is an empirical truth, 3) is a purely conceptual truth. This is the reason why the role of truth characterized by 3) is purely logical, whereas the role of being pain characterized by 4) isn't.¹⁸² Moreover, 4) leads to 1)—namely, we need 1) as an explanation of why the mental property of being in pain has the functional role that is characterized by 4). On the other hand, we don't need such an explanation for 3)—it is a purely conceptual truth that the property of truth has the functional role that is characterized by 3). Consequently, as long as we accept the deflationary definition of truth 3), we don't need a supervenience theory of truth of the type 2). However, this is exactly what Damnjanovic needs in order to justify the causal explanatory role of truth by appealing to Jackson and Pettit's program explanation. Hence, Damnjanovic's proposal collapses due to an internal conflict—either he has to deny the deflationary theory or he should give up the causal-explanatory role of truth. He cannot have both.

3.3 Kitcher's Realist Attack against Deflationism: A Case For the Causal-Explanatory Role Of Correspondence Truth In the Success-To-Truth Rule

3.3.1 Introduction¹⁸³

¹⁸² In other words, we don't need any explanation why we can derive a statement of the form '*P*' from a statement of the form '<*P*

¹⁸³ Kitcher's articles referred to in this section include P. Kitcher, "On the Explanatory Role of Correspondence Truth," *Philosophy and Phenomenological Research*, Vol. 64(2), 2002, pp:346-364; P. Kitcher, "Real Realism: The Galilean Strategy," *The Philosophical Review*, Vol. 110, No. 2. 2001, pp:151-197; and P. Kitcher, "Explanatory

Is there any rational relationship between the problem of truth and the problem of realism? According to Horwich, there is no interesting connection between a theory of truth and the debate between realism and anti-realism. A commitment to realism or anti-realism, says Horwich, doesn't make one embrace a particular theory of truth. Or to speak conversely, to embrace the deflationary theory of truth doesn't constitute a basis to take up or avoid any particular position in the realism debate. In other words, a theory of truth is neutral in the debate between realism and anti-realism.¹⁸⁴

In his "On the Explanatory Role of Correspondence Truth", Kitcher aims to show otherwise; he aims to show that realism—more precisely, scientific realism—requires the correspondence theory of truth.¹⁸⁵ Scientific realism comes in many varieties (and so does scientific anti-realism). The purpose of my discussion of Kitcher isn't to determine the plausibility of scientific realism *per se* but to determine whether scientific realism, as Kitcher claims, requires a non-deflationary theory of truth; i.e., the correspondence theory of truth. That said, it is inevitable to say a few words about scientific realism. I will here briefly define scientific realism in terms of two doctrines respectively called 'SR' and 'SSR' by Michael Devitt:

(SR) Most of the essential unobservables of well-established current scientific theories exist mind-independently.

Unification," *Philosophy of Science*, Vol. 48 (4), 1981, pp:507-531.

¹⁸⁴ See P. Horwich, "Realism and Truth" in his *From a Deflationary Point of View*, Oxford University Press, 2004.

¹⁸⁵ P. Kitcher (2002), "On the Explanatory Role of Correspondence Truth," *Philosophy and Phenomenological Research*, Vol. 64(2), 346-364. The deflationary notion of truth that Kitcher challenges in this work is mainly the one advocated by Horwich. This is the reason why Kitcher's main argument is concerned with criticizing Horwich's response to the success argument. Kitcher's criticism of Horwich's response to the success argument will be discussed in the next section.

(SSR) Most of the essential unobservables of well-established current scientific theories exist mind-independently, and mostly have the properties attributed to them by science.¹⁸⁶

Kitcher is committed to both (SR) and (SSR) and so he states that a scientific statement that has passed appropriate scrutiny of its epistemic credentials—including the one that involves an expression for an *unobservable* entity—should be considered *true* or *approximately true*.¹⁸⁷

Kitcher also characterizes scientific realism he defends as a “homely doctrine, vested in commonsense appraisals rather than weighty metaphysics.”¹⁸⁸ To explain what Kitcher means by this, it has to be pointed out that Kitcher builds the correspondence notion of truth into his characterization of scientific realism. According to Kitcher, a scientific statement that has passed scrutiny of its epistemic credentials isn’t merely true but true in the correspondence sense. So, on Kitcher’s view, scientific realism presupposes the correspondence theory of truth. What Kitcher means by ‘the correspondence theory of truth’, however, is a theory that accepts only i) and ii), but not iii) in the following:

- i) Reference relations are causal relations between mind-independent entities and linguistic tokens.

¹⁸⁶ M. Devitt, “Scientific Realism,” *The Oxford Handbook of Philosophy*, eds. By F. Jackson and M. Smith, Oxford University Press, 2005, p. 769.

¹⁸⁷ Richard Boyd characterizes scientific realism using the expression ‘approximate truth.’ An opposing view is Van Fraassen’s constructive empiricism. According to his constructive empiricism, a scientific statement involving an expression for an unobservable entity is not true. On this view, a scientific statement involving an expression for an unobservable entity should be evaluated in terms of empirical adequacy instead of truth or falsity. See R. N. Boyd, “Realism, Underdetermination, and a Causal Theory of Reference,” *Noûs*, Vol. 7, No. 1, 1973, pp:1-12; Bas C. van Fraassen, *The Scientific Image*, Oxford: Clarendon Press, 1980.

¹⁸⁸ P. Kitcher, “On the Explanatory Role of Correspondence Truth,” p. 347.

- ii) It's possible to give a naturalistic account of these relations; that is, semantics can be conducted as a natural science.
- iii) There's a physicalist reduction of semantics.¹⁸⁹

So, the reason why Kitcher calls his realism a 'homely' or 'modest' doctrine is that the correspondence theory of truth that he considers an essential aspect of scientific realism doesn't require the notions of reference and truth to be reduced to physicalistic terms. Kitcher says

Here I should note explicitly that the type of correspondence theory I envisage is one like that originally sketched by Hartry Field in "Tarski's Theory of Truth," *Journal of Philosophy* 69 (1972): 347-75. One important difference, however, is that I don't see a naturalistic account of reference as necessarily involving a reduction to a physicalist vocabulary. Naturalists who are antireductionists with respect to some sciences (for example, parts of biology) should allow that the basis for a reduction of the notion of reference may out-strip the resources of physics.¹⁹⁰

But why does Kitcher consider the correspondence theory of truth an essential aspect of scientific realism. ? Namely, what is Kitcher's reason for holding that scientific realism requires a non-deflationary theory of truth? Kitcher's reason in a nutshell is that the correspondence notion of truth plays an essential role in the so-called 'success-to-truth' rule that can be formulated as follows:

(Success-to-Truth Rule)

S plays a crucial role in a systematic practice of fine-grained prediction and intervention.

¹⁸⁹ P. Kitcher, "On the Explanatory Role of Correspondence Truth," *Philosophy and Phenomenological Research*, Vol. 62 (2) 2002, pp:346-364, P. 347. .

¹⁹⁰ P. Kitcher, "Real Realism: The Galilean Strategy," *Philosophical Review* Vol. 110 (2), 2001: 151-197.

S is approximately true.¹⁹¹

Two things should be noted about the success-to-truth rule: first, this rule embodies a success-explanation of a certain type—a success-explanation that accounts for the success or effectiveness of a scientific theory. A success-explanation of a scientific theory, however, can be construed as a success-explanation of behavior of a specific type; i.e., behavior caused by a scientific belief. That is to say, a success-explanation of a scientific theory is supposed to explain how scientists who follow a certain scientific theory are so successful in achieving their goals; for example, how bio-engineers who accept the main claims of molecular genetics are so successful in manufacturing new organisms.¹⁹² So, the success-to-truth rule doesn't have to be construed as a success-explanation of an entirely new kind. Kitcher also treats it as continuous with a success-explanation of behavior. Second, the success-to-truth rule embodies an important argument for scientific realism; namely, the argument called 'the no-miracle argument' or 'the abductive argument' for scientific realism.¹⁹³ According to the no-miracle argument, the empirical success of a scientific theory would not be explained unless it is true or approximately true. The abductive argument defends the same idea; from the empirical success of a theory, we can derive the conclusion that the theory is true as a hypothesis that best explains the success of the theory. Thus, Kitcher's argument proceeds as follows: first, scientific realism relies on the success-to-truth rule, and second, the success-to-truth rule relies on the correspondence notion of truth. So, according to Kitcher, the correspondence notion of truth plays an essential role in the

¹⁹¹ Here, *S* is a (scientific) statement. P. Kitcher, "On the Explanatory Role of Correspondence Truth," (from the same source as was indicated above), p. 347.

¹⁹² This is Kitcher's example. See P. Kitcher, "On the Explanatory Role of Correspondence Truth," p. 346.

¹⁹³ This argument is often traced back to Putnam's slogan: "Realism is the only philosophy that does not make the success of science miracle." (H. Putnam, *Mathematics, Matter and Method: Philosophical Papers Vol. 1*, London: Cambridge University Press, 1975, p. 73.) This idea was made famous by Putnam, J. J. C. Smart, and Boyd. According to Kitcher (in his "Real Realism: The Galilean Strategy"), this argument was already employed by Descartes.

success-to-truth rule, and in that sense scientific realism requires the correspondence theory of truth.

In the remaining of this chapter, I will try to discern whether Kitcher is right about this—whether the success-to-truth rule requires the correspondence notion of truth. By maintaining that the success-to-truth rule and thereby scientific realism requires the correspondence theory of truth, Kitcher aims to criticize the deflationary theory of truth *from a realist point of view*. On Kitcher's view, deflationism about truth is not compatible with scientific realism—as a real realist, Kitcher says he cannot embrace the deflationary theory of truth. For a while I was somewhat sympathetic to Kitcher's view, mainly because many scientific realists presuppose the correspondence notion of truth when they invoke the notion of truth to characterize scientific realism. According to Horwich, however, it is a commonly made mistake to think that scientific realism has anything to do with a theory of truth of a certain kind. I now think that Horwich is right about the neutrality of a theory of truth in the realism debate. At least, the correspondence notion of truth doesn't seem to play a role in the success-to-truth rule. Some other arguments for scientific realism might need the correspondence theory of truth, but the success-to-truth rule doesn't. This is what I will argue for in the remaining of this chapter.

3.3. 2. Horwich's View on Truth and the Realism Debate

Among all the deflationists, Horwich seems to be Kitcher's major opponent. Horwich's views are in direct opposition to Kitcher's views in two respects. First, whereas Kitcher holds that deflationism is not compatible with realism, Horwich holds that deflationism is absolutely neutral between realism and anti-realism. Second, whereas Kitcher thinks that the explanatory

role played by truth in the success-to-truth rule cannot be accommodated by the deflationary theory of truth, Horwich thinks it can. As we have seen 3.2.1, this is the gist of Horwich's response to the success argument. Here, in order not to mislead readers with similar terms, I shall make the following distinction:

- Success-Explanation: an explanation that accounts for the success of behavior or the effectiveness of a scientific theory.
- Success-Argument: an argument criticizing the deflationary theory of truth on the grounds that the explanatory role of truth in a success-explanation cannot be accommodated by the deflationary theory of truth.

Kitcher's argument begins with his criticism of Horwich's response to the success-argument.

Before examining Kitcher's argument, however, I shall first explain briefly why Horwich claims that deflationism is neutral between realism and anti-realism.

Let's first consider what Horwich says:

The important thing, rather, is to see that it is not settled one way or the other by deflationism about truth . . . Not only is deflationism neutral with respect to the disputes between realism and antirealism, but so are the other well-known theories of truth . . . [N]ot only is the *right* account of truth (namely, deflationism) absolutely neutral regarding realism, but even if one adopted what I have suggested are *wrong* points of view on truth (either the correspondence or the coherence accounts) it would be a further mistake to think that any of the epistemological or metaphysical issues that constitute the realism debate would be settled.¹⁹⁴

What Horwich says here is that not only can deflationists be either realists or anti-realists but also that endorsing a non-deflationary theory doesn't help solve problems facing the realism

¹⁹⁴ P. Horwich, "Realism and Truth," in P. Horwich, *From a Deflationary Point of View*, Oxford University Press, 2004, p. 41.

issue. Consider one of the most challenging problems facing realism: that we cannot have epistemological access to phenomena involving unobservable entities. One may try to solve this problem by endorsing the verificationist theory of truth—that is, by “promoting the accessibility of facts at the expense of their autonomy.”¹⁹⁵ The problem with the verifiability theory of truth, according to Horwich, is that it assumes the equivalence schema: $\langle P \rangle$ is true iff P . Horwich says,

[O]ne might reason that, since it is relatively easy to tell regarding, for example, the proposition that there are infinitely many stars, whether or not it is verifiable, then—given the identification of truth with verifiability—it would be equally easy to tell if that proposition is true, hence easy to tell if there are infinitely many stars. And, moreover, since the verifiability of the proposition is a fact about our methodology, then so would be the truth of the proposition, i.e. the fact that there are infinitely many stars.

To elaborate Horwich’s point (which wasn’t easy at all), the verificationist would first identify the truth of \langle There are infinitely many stars \rangle with its verifiability (e.g. its verification condition), and then, second, infer that there are infinitely many stars (or that there aren’t infinitely many stars) from the fact about the verifiability of \langle There are infinitely many stars \rangle .

The problem with this approach, however, is that it stealthily appeals to the equivalence schema (E): $\langle P \rangle$ is true iff P .¹⁹⁶ The verificationist can infer that there are infinitely many stars (or there aren’t infinitely many stars) from the fact about the verifiability of \langle There are infinitely many stars \rangle because he assumes the truth of an instance of the equivalence schema; i.e., \langle There are infinitely many stars \rangle is true iff there are infinitely many stars. Can the verificationist assume the equivalence schema? The answer is no, because that’s exactly what needs to be justified by the verificationist theory of truth. So, the attempt to defend realism by identifying truth with verifiability doesn’t work—it faces the problem of circularity. A similar problem confronts a

¹⁹⁵ *ibid.*

¹⁹⁶ *ibid.*

correspondence theorist who attempts to defend realism by defining truth with a correspondence relation. Horwich says,

[T]he correspondence theory of truth provides no quick path to sceptical anti-realism or to metaphysical realism. For again, those conclusions can be obtained only given the equivalence schema which, as in the previous case, will be no easier to establish than the realist or anti-realist theses it is being used to support.¹⁹⁷

Consider again the proposition that there are infinitely many stars. The correspondence theorist claims that <There are infinitely many stars> is true iff it corresponds to the fact that there are infinitely many stars. From the correspondence theorist point of view, however, the truth of this claim should be justified by the correspondence theory. If the correspondence theorist is to simply assume that <*P*> is true iff <*P*> corresponds to the fact that *P*, there is no point of invoking the correspondence theory of truth in arguing for realism or anti-realism. The whole point of invoking a non-deflationary theory of truth is that to invoke a non-deflationary theory of truth—for example, the correspondence theory of truth—is essential to argue for realism or anti-realism. But if so, it is unlikely that anyone would be able to argue for realism or anti-realism—to establish a non-deflationary theory of truth is as difficult as establishing realism or anti-realism.

But then, what explains the prevalence of the idea that the realism debate is about the notion of truth? Horwich's explanation is that those who think that the realism debate is about the notion of truth are misled by the role played by the notion of truth appearing in the theses of realism and anti-realism. The notion of truth is often employed in articulating the theses of realism and anti-realism. For example:

¹⁹⁷ P. Horwich, *ibid.*

All truths are verifiable.

Theoretical hypotheses are truth-value-less.

Science aims at truth.

No contingent statement about the future can be true.

Horwich's point is that the notion of truth appears as a purely logical device in all of these claims; it appears as a device of generalization. To see why, let's consider the third claim in the list above; the claim that science aims at truth. This statement is a generalization of particular statements such as 1) in the following:

- 1) Scientists want it to be that they believe that there are infinitely many stars only if there are infinitely many stars; or scientists want it to be that they believe that electrons ARE almost perfectly round only if electrons are almost perfectly round; or . . .

By appealing to the equivalence schema (E), we can derive from 1) the following: 1*).

- 1*) Scientists want it to be that they believe that there are infinitely many stars only if it is true that there are infinitely many stars.

Now, we can generalize 1*) into 1**) in the following:

- 1**) Scientists want it to be that they believe x only if it is true that x .

The claim above—that science aims at truth—is a colloquial way of expressing 1^{**}).¹⁹⁸ So, the role of truth in this claim is purely logical—we appeal to the notion of truth in order to generalize an infinite disjunction such as 1) above. Similar accounts can easily be given for other claims listed above.

On the basis of these considerations, Horwich concludes that truth is irrelevant to the realism debate. Horwich isn't alone thinking this way; for example, Devitt also contends that realism is compatible with a totally deflationary theory of reference and truth. A deflationist, says Devitt, can be a realist or an anti-realist.¹⁹⁹ Like Horwich, Devitt says that the notion of truth appears as a device of generalization in the realism debate. Of course, Kitcher disagrees, and his arguments will be presented in the next section.

3.3.3 Kitcher's Arguments

A. Kitcher's Criticism of Horwich's Explanation

As was mentioned earlier, Kitcher's main argument is that the correspondence notion of truth plays an essential role in the success-to-truth rule. In order to establish this point, Kitcher examines Horwich's response to the success-argument. The gist of Horwich's response to the success-argument is that the explanatory role played by truth in an account of the success of behavior or a scientific theory can exhaustively be accommodated by the deflationary theory of

¹⁹⁸ This exposition is based on Horwich's account. See P. Horwich, "Realism and Truth," in P. Horwich's *From a Deflationary Point of View*, Oxford University Press, 2004.

¹⁹⁹ M. Devitt, "Scientific Realism," in *The Oxford Handbook of Philosophy*, eds., F. Jackson and M. Smith, Oxford University Press, 2005. Hartry Field seems to be in accordance with this view as well, although Field's position on this is much less clear. Field once said that deflationism is somewhat anti-realistic, although it aims to be neutral in the debate between realism and anti-realism. See "The Deflationary Conception of Truth," in G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 55-117.

truth. An example of Horwich's explanation was already introduced in the previous section (3.2.1), but since Kitcher's criticism is also concerned with the same example, it should be repeated here:

- 1) Bill wants <Bill has a beer>. [Premise]
- 2) Bill believes <Bill nods \rightarrow Bill has a beer>. [Premise]
- 3) [Bill wants <Bill has a beer> & Bill believes <Bill nods \rightarrow Bill has a beer>] \rightarrow Bill nods. [Premise]
- 4) Bill nods. [from 1), 2), and 3)]
- 5) Bill's belief is true. [Premise]
- 6) <Bill nods \rightarrow Bill has a beer> is true. [From 2) and 5)].
- 7) <Bill nods \rightarrow Bill has a beer> is true iff Bill nods \rightarrow Bill has a beer. [An instance of the equivalence schema]
- 8) Bill nods \rightarrow Bill has a beer. [From 6) and 7)]
- 9) Bill has a beer. [From 4) and 8)]
- 10) <Bill nods \rightarrow Bill has a beer> is true iff Bill nods \rightarrow Bill has a beer. [An instance of the equivalence schema]
- 11) Bill nods \rightarrow Bill has a beer. [From 6) and 7)]
- 12) Bill has a beer. [From 4) and 8)]
- 13) <Bill has a beer> is true iff Bill has a beer. [From the equivalence schema]
- 14) <Bill has a beer> is true. [From 9) and 10)]
- 15) Bill gets what he wants. [From 1) and 11)]²⁰⁰

As was stated in the previous section (3.2.1), the derivation from 1) to 15) is Horwich's deflationary explanation of the success of behavior—by explaining the success of behavior this way, Horwich responds to the success argument against deflationism. According to Kitcher, however, Horwich's deflationary explanation "stops rather at a shallow level," implying that

²⁰⁰ P. Kitcher, *ibid.* (I added explanations inside brackets as they appear in Horwich's original exposition. See P. Horwich, *Truth*, 2nd ed. Clarendon Press: Oxford University Press, 1999, p. 23.

there is more to be explained about the success of behavior.²⁰¹ If we stop with Horwich's explanation, we wouldn't need the correspondence notion of truth. But Horwich's explanation, says Kitcher, doesn't tell us the whole story of what underlies the success of behavior. The following is what Kitcher says about Horwich's explanation of the success of behavior:

This appears to be a very powerful debunking of the realist's insistence that correspondence truth was required to explain success, for we seem to be able to give an explanation without looking at any notion of correspondence. All that's needed is the minimal theory of truth. But we should ask exactly what has been explained. The answer, surely, is that if an agent has a true belief about means-ends relations, then that agent is likely to be successful. That isn't quite the *explanandum* that realists have taken to be crucial in their defenses of correspondence truth. Correspondence truth has been supposed to be necessary because of the way in which true beliefs about means-ends relations, or behavior that is as if the agent has true beliefs about means-ends relations, result from true beliefs about the objects that figure in the desired goal-state. The problem with Horwich's explanation is that it stops at a rather shallow level. Indeed, it's akin to the classic paradigms of "explanation" that subsumed facts about bird plumage under putative ornithological laws. The major instances of scientific explanation go much deeper, and I think we can emulate them in this case. Specifically, we can deepen our understanding of what is going on in successful actions by considering, among other things, the causal relations that connect items in the world with the tokens that figure in the agent's psychological states.²⁰²

Kitcher's criticism may be summarized as follows:

- i) Horwich's deflationary explanation gets the *explanandum* wrong;
- ii) Horwich's deflationary explanation is shallow because it doesn't invoke causal relations that connect the constituents of psychological states with the entities in the mind-independent world.

I shall elaborate these points as follows. First, according to Kitcher, Horwich gets the *explanandum* wrong because what needs to be explained isn't merely how successful behavior results from an agent's true beliefs about means-ends relations. From the realist's perspective,

²⁰¹ P. Kitcher, "On the Explanatory Role of Correspondence Truth," p. 354.

²⁰² P. Kitcher, *ibid.*,

what needs to be explained is how the agent obtains his true beliefs about means-ends relations—it is the *source* of an agent’s true beliefs about means-ends relations that should be explained. An example of an agent’s true beliefs about means-ends relations is Bill’s belief that if he nods, then he will get a beer (i.e., Bill’s believing <Bill nods → Bill gets a beer>). Horwich doesn’t explain where Bill got this belief, and this, according to Kitcher, makes Horwich’s explanation remain at a “shallow” level.

Kitcher’s quest for the source of a belief about means-ends relations, however, is connected to his quest for the principles that can explain “systematic successes”. The notion of systematic successes looms large in Kitcher’s criticism of deflationism. To understand what he means by that notion, let’s consider the following:

Bill’s performance is extremely local. We hear one success on an occasion in which Bill had a commonplace belief about means-ends relations. But I’ve insisted, from the beginning, that the realist’s argument concerned *systematic successes*. What exactly does that mean? I answer that the cases of concern are those in which our agent has a device, or set of devices, from which beliefs about means-ends relations (or behavior that is as if the agent had beliefs about means-ends relations) flow.²⁰³

Kitcher is one of those who regard “systematic unification” as a criterion of a good scientific explanation.²⁰⁴ A good explanation is the one that has the capacity to unify a wide variety of phenomena under one or a small number of ultimate principles. Kitcher’s emphasis on systematic successes should be understood from this perspective as well—the concept of systematic successes assumes that a wide variety of successful behavior can be explained in

²⁰³ P. Kitcher, “On the Explanatory Role of Correspondence Truth,” p. 355. Italicization is done by me.

²⁰⁴ See P. Kitcher, “Explanatory Unification,” *Philosophy of Science* 48, (1981) pp. 507-531. It seems that Kitcher’s view of a “good explanation” affected his view of scientific realism. Kitcher’s view that a good explanation should have the power of unifying a wide variety of phenomena under one or a small number of basic principles led him to advocate not only scientific realism but also scientific realism with the correspondence theory of truth.

terms of one or a small number of underlying principles. To take a related example, we don't use the term 'systematic failure' when one or two schools in the nation are doing badly; we use the term when a large number of schools in the nation are doing badly for a long period of time. Some may object that there isn't necessarily a systematic failure just because a large number of schools are failing. Kitcher disagrees; if a large number of schools are failing, there should be a common cause in terms of which those massive failures can be explained. Similarly, a success-explanation should offer an account of systematic successes. A Horwich-style deflationary explanation, however, cannot account for systematic successes; it can explain only individual successes.

But why can't a Horwich-style deflationary explanation offer an account of systematic successes? The reason, says Kitcher, is that a Horwich-style deflationary explanation fails to relate all the true beliefs about means-ends relation to their ultimate sources. In order to explain systematic successes, we should be able to explain a wide variety of successful behavior in terms one or a small number of ultimate principles. To do so in turn requires tracing all the true beliefs about means-ends relations back to their ultimate source, and the explanatory role of correspondence truth emerges in the search of the ultimate source of true belief. In order to argue for these points, Kitcher brings in the example of a map.

B. The Map-Example

Suppose Ophelia uses a map to go to various places she wants to; she obtains various true beliefs about means-end relations from the map, and she adjusts her behavior in accordance with these beliefs. By doing so, she succeeds in arriving various places that she wants to go. One day,

Ophelia wants to go to the brook in her vicinity. So, she refers to the map, and she obtains the belief that if she follows the path past the willows, she will get to the brook. She follows the path past the willows, and she succeeds in arriving at the brook.

The standard deflationary explanation of the success of Ophelia's behavior proceeds in the same way that its explanation of the success of Bill's behavior does. It begins with Ophelia's having the desire <Ophelia wants to get to the brook>, and Ophelia's believing <Ophelia follows the path past the willows → Ophelia gets to the brook>. The rest of the explanation will be omitted here, since it proceeds in the exactly same way that the explanation of Bill's success does. As was stated above, Kitcher finds this explanation stops at a shallow level, because it doesn't trace Ophelia's belief about means-ends relation back to its source. The source of her belief is the map. So, Kitcher expands the Horwich-style explanation of Ophelia's success by adding the following explanation:

- 1) Ophelia believes <the map is accurate>.
- 2) Ophelia is competent to read the map.
- 3) Ophelia desires q [possible instances: <Ophelia gets to the brook>, <Ophelia gets to the cliff>, <Ophelia gets to the graveyard>].
- 4) [Ophelia believes <the map is accurate> & Ophelia is competent to read the map & Ophelia desires q] → Ophelia does $A(q)$.
- 5) The map is accurate.
- 6) [The map is accurate & Ophelia does $A(q)$] → q .
- 7) q .

Here, ' q ' is to be replaced by Ophelia's desires; ' $A(q)$ ' is to be replaced by the action that Ophelia performs to achieve her desire q . What Kitcher is concerned with in this derivation is the contrast between step 4) and step 6). According to Kitcher, step 4) tells us that the action $A(q)$ is

psychologically pertinent to achieve Ophelia's desire q . An action $A(q)$ is *psychologically pertinent* to an agent's goal as long as the agent *thinks* that the action will lead to the satisfaction of her goal. The derivation from 1) to 4) explains why Ophelia *thinks* that doing the action $A(q)$ is pertinent to her goal. The psychological pertinence of an action, however, should be distinguished from the objective pertinence of an action. An action $A(q)$ is objectively pertinent to the agent's goal q only if the agent succeeds in achieving her goal by doing $A(q)$. Step 6) tells us that Ophelia's action $A(q)$ is objectively pertinent to her goal q .

Now, the question that Kitcher is most interested in is what explains an action $A(q)$ that is both psychologically and objectively pertinent to an agent's goal. Kitcher says, "The explanatory mystery is precisely why there should be any action that meets these two constraints."²⁰⁵ And here comes Kitcher's main criticism of the deflationary explanation of the success of behavior. He says, "To read 5) in the deflationist's sense leaves at least one of 4), 6) *unexplained*."²⁰⁶ That is to say, the deflationist can explain either the psychological pertinence or the objective pertinence of $A(q)$ to q , *but not both*. On Kitcher's view, the deflationist can easily explain the psychological pertinence of an action to a goal. The deflationist, however, cannot explain how a psychologically pertinent action can also be objectively pertinent. Kitcher's reason, in short, is that the deflationist cannot explain systematic success; that is to say, on Kitcher's view, if the deflationist cannot explain systematic success, she cannot explain the objective pertinence of an action $A(q)$ to a goal q , either. I will argue in the next section, however, that this view is wrong.

3.3.4. The Criticisms of Kitcher's Argument

²⁰⁵ P. Kitcher, "On the Explanatory Role of Correspondence Truth," p. 359.

²⁰⁶ P. Kitcher, *ibid*. I adjusted the numbers of the statements.

What I will argue for in this section is as follows; first, although the deflationist cannot explain systematic successes, he can explain the objective pertinence of an action to a goal; and second, for Kitcher's ultimate purpose—the defense of realism—the objective pertinence of an action is all that needs to be explained. In particular, I will argue that the success-to-truth rule can be sustained as long as the objective pertinence of an action is explained. Kitcher's requirement for an account of systematic successes seems orthogonal to the defense of the success-to-truth rule.

First, let's examine whether the deflationist can explain how there can be an action $A(q)$ that is both psychologically and objectively pertinent to an agent's goal q . To do so, let's return to Kitcher's expansion of a Horwich-style explanation of the success of Ophelia's action. I will reproduce it here for the sake of convenience:

- 1) Ophelia believes <the map is accurate>.
- 2) Ophelia is competent to read the map.
- 3) Ophelia desires q [possible instances: <Ophelia gets to the brook>, <Ophelia gets to the cliff>, <Ophelia gets to the graveyard>].
- 4) [Ophelia believes <the map is accurate> & Ophelia is competent to read the map & Ophelia desires q] \rightarrow Ophelia does $A(q)$.
- 5) The map is accurate.
- 6) [The map is accurate & Ophelia does $A(q)$] $\rightarrow q$.
- 7) q .

According to Kitcher, when we understand 5) in the deflationary sense, we cannot explain how there can be an action $A(q)$ that is both psychologically and objectively pertinent to Ophelia's goal q . But the deflationist has no problem explaining it. First, the psychological pertinence of $A(q)$ is fully explained by the step from 1) to 4) above. To say that an action $A(q)$ is

psychologically pertinent to an agent's goal q is to say that the agent *believes* that doing the action $A(q)$ will satisfy his goal. The step from 1) to 4) shows that Ophelia believes that doing the action $A(q)$ will satisfy her goal q . Given that each step from 1) to 4) is satisfied, Ophelia will do the action $A(q)$. But then, in order to explain the psychological pertinence of Ophelia's action, it isn't even necessary to invoke 5). Namely, the psychological pertinence of an action can be explained independently of the truth of a belief or the accuracy of the map. So, at this stage, it doesn't matter how we understand the notion of accuracy or truth; we don't even have to invoke 5).

At the second stage, however, we need to invoke the accuracy of the map; an account of the objective pertinence of an action requires the step 5). To say that an action $A(q)$ is objectively pertinent to a goal q is to say that if an agent does $A(q)$, he will succeed in satisfying the goal q ; namely, it is to say that OS) in the following is true:

OS) Ophelia does $A(q) \rightarrow q$.

But then, to say that OS) is true amounts to saying that the action is successful; it means that if an agent does $A(q)$, he will succeed in achieving his goal q . Therefore, to explain the objective pertinence of an action is to explain the success of behavior; they amount to the same thing. But then, it is weird for Kitcher to say all of sudden that the deflationist cannot explain the objective pertinence of an action. We all agreed from the beginning that the success of an action should be explained, among other things, by the truth of a belief. But then, the objective pertinence of an action should be explained by the truth of a belief. In the map-example, the truth of Ophelia's

beliefs about means-ends relations originates from the accuracy of the map. Kitcher questions whether the deflationist can make sense of the role played by the accuracy of map in an account of the objective pertinence of Ophelia's action. Contrary to Kitcher's skepticism, the deflationist can easily explain it by offering the derivation in the following:

5) The map is accurate.

5¹) The map says $\langle \text{One does } A(q) \rightarrow q \rangle$.

5²) $\langle \text{One does } A(q) \rightarrow q \rangle$ is true. (from 5) and 5¹))

5³) One does $A(q) \rightarrow q$. (from 5²) and the equivalence schema (E))

OS) Ophelia does $A(q) \rightarrow q$. (an instantiation of 5³))

The derivation from 5) to OS) shows how the objective pertinence of Ophelia's action $A(q)$ is derived from the accuracy of the map. This explanation assumes no principle of truth other than the equivalence schema. So, unlike what Kitcher says, the deflationist has no problem explaining both the psychological and objective pertinence of an action.

But then, what is the nature of Kitcher's complaint of the deflationary account of the success of behavior? The nature of Kitcher's complaint is that the deflationary account cannot explain *systematic successes*. As was noted above, to explain systematic successes is to explain a wide variety of successful behaviors in terms of a small number of ultimate principle(s). In order to do so, we have to find a common feature shared by all successful behaviors. In the map-example, the common feature is that they all flow from 5)—the accuracy of the map. More

precisely, any successful behavior in the map-example is caused by a true belief about means-ends relations, and a true belief about means-ends relations originates from the accuracy of the map. So, it appears, at a first glance, that the deflationist should have no problem explaining systematic successes—the underlying principle of all successful behaviors is the accuracy of the map. At a second glance, however, we find something wrong with this explanation—when we understand the notion of truth in the deflationary sense, it is trivial to say that the underlying principle of successful behaviors is the accuracy (truth) of the map. When we understand the accuracy of the map in the deflationary sense, 5) is equivalent to 5^D) in the following:

5^D) The map says <One walks past the willows → one gets to the brook> & (One walks past the willows → one gets to the brook; or the map says <One crosses the river → one gets to the castle> & (One crosses the river → one gets to the castle); or . . .

To say that 5^D) is the underlying principle of all successful behaviors trivializes the very idea of systematic successes—the idea that a wide variety of successful behavior can be explained in terms of a small number of principles. 5^D) consists of infinite disjunctions, and this already betrays the idea that 5^D) can be an underlying principle of a wide variety of successful behaviors. Given that 5) is equivalent to 5^D), to say that an action $A(q)$ is successful because the map is accurate amounts to saying 5^{D*}):

5^{D*}) <Ophelia follows the path past the willows → Ophelia gets to the brook> because the map says <One follows the path past the willows → one gets to the brook> & (One follows the path past the willows → one gets to the brook); or <Ophelia crosses the

river → Ophelia gets to the castle> because the map says <One crosses the river → one gets to the castle> & (One crosses the river → one gets to the castle); or . . .²⁰⁷

To be sure, 5^{D*}) isn't what Kitcher means by an account of systematic successes; there is nothing systematic about 5^{D*}). But this is all we can afford if we understand 5) in the deflationary sense.

Since the deflationist denies that phenomena about truth can be reduced to a small number of more basic facts, to invoke the accuracy of the map in an account of the success of behavior means nothing more or nothing less than 5^{D*}). The utility of a truth-expression—for example, 'true' or 'accurate'—is that it enables us to generalize 5^{D*}) into G) in the following:

G) An action $A(q)$ is successful because its ultimate source—the map—is accurate.

The explanatory role of truth or accuracy is exhausted by such a logical, generalizational, function. But to say that the nature of the explanatory role of truth is purely logical is to say that the notion of truth doesn't have any other explanatory function. To make the point vivid, let's compare the explanatory role played by notion of truth with the explanatory role played by the notion of oxygen in Lavoisier's theory of combustion. Roughly speaking, Lavoisier's theory of combustion is that oxygen is the underlying principle in terms of which various phenomena of combustion can be explained. In this theory, the notion of oxygen doesn't play a merely logical, generalizational, function; the notion of oxygen determines well-defined physical properties, and these properties cause combustion by interacting with other physical properties. By contrast, the

²⁰⁷ Relevant agents go beyond Ophelia; they can be anybody.

deflationary notion of truth doesn't determine well-defined properties, let alone physical properties. Therefore, it doesn't make sense to say that the *property of truth* explains a wide variety of successful behaviors; truth isn't even the kind of property that can have such an explanatory power.

Michael Levin—whose view Kitcher mentions in his discussion—claims that the notion of truth has nothing to do with the success of behavior (including successful predictions and interventions made by scientists).²⁰⁸ Levin says,

Why do airplanes stay up? Surely the reason airplanes stay up is not

(1) “The pressure on the underside of a moving airfoil is greater than the pressure on its overside” is true,

but rather

(2) The pressure on the underside of a moving airfoil *is* greater than the pressure on its overside.²⁰⁹

Levin isn't here advocating the redundancy theory of truth; he only tries to identify what really does the causal-explanatory work; the causal-explanatory work is done by the properties invoked in 2). The notion of truth invoked (1) is a purely disquotational device doing no causal-explanatory work. Similarly, the notion of truth invoked in G) above doesn't do a causal-explanatory work—it is a device of generalization. The causal-explanatory work is done by Ophelia's believing <One follows the path past the willows → one gets to the brook> & (One follows the path past the willows → one gets to the brook); Ophelia's believing <One crosses the river → one gets to the castle> & (One crosses the river → one gets to the castle); and so forth.

²⁰⁸ M. Levin, “What Kind of Explanation is Truth?”, in J. Leplin (ed.), *Scientific Realism*, Berkeley: University of California Press, 1984, pp:124-39.

²⁰⁹ M. Levin, *ibid.*, p. 126.

The notion of truth is appealed to only as a device of generalization such an infinite conjunction (or a disjunction).

So, Kitcher is right when he claims that if we construe the notion of truth in the deflationary sense, we cannot expect truth to explain systematic successes. Let's recall, however, what Kitcher aims to prove eventually—he aims to prove that scientific realism requires a non-deflationary notion of truth; in particular, correspondence truth. To prove this point, Kitcher brings in the success-to-truth rule:

(Success-to-Truth Rule)

S plays a crucial role in a systematic practice of fine-grained prediction and intervention.

S is approximately true.

This rule embodies one of the most important arguments for scientific realism—it embodies ideas underlying the “no-miracle” argument and the abductive argument for scientific realism. So, according to this rule, we would not be able to explain how a theory yields successful predictions and interventions unless we take the theory to be true; put differently, from the premise that a theory yields successful predictions and interventions, we are allowed to infer that the theory is true. Kitcher maintains that the deflationist cannot make sense of this argument for scientific realism on the grounds that the deflationist cannot accommodate the role of truth in this argument; i.e., the success-to-truth rule.

The deflationist has no problem, however, accommodating the role of truth in the success-to-truth rule—he can explain it in the same way he explains the role of truth in an account of the success of Ophelia's behavior. The nature of the role played by truth in the

success-to-truth rule, says the deflationist, is purely logical, because the success-to-truth rule generalizes an infinite disjunction 8) in the following:

- 8) If Newton's theory is composed of propositions $\langle P_1 \rangle, \langle P_2 \rangle, \dots, \langle P_n \rangle$, and Newton's theory plays a crucial role in a systematic practice of fine-grained prediction and intervention, then P_1, P_2, \dots, P_n ; or if Einstein's theory is composed of propositions $\langle R_1 \rangle, \langle R_2 \rangle, \dots, \langle R_n \rangle$, and Einstein's theory plays a crucial role in a systematic practice of fine-grained prediction and intervention, then R_1, R_2, \dots, R_n ; or . . .

Given that the success-to-truth rule is purported as an argument for scientific realism, 8) exhausts what is stated by the success-to-truth rule. The reason the success-to-truth rule supports scientific realism is that if this rule is accepted, the two main theses of scientific realism will be accepted as well: (SR) and (SSR) (see 3.3.1 and the footnote below).²¹⁰ That is to say, if we accept the success-to-truth rule, we will accept the existence of the entities—including unobservable entities—that feature in well-established scientific theories; we will also believe that those entities have the properties attributed to them by science. The same effect will be obtained if we accept 8). For example, suppose we infer the conclusion that masses warp space and time from the premise that Einstein's theory states $\langle \text{Masses warp space and time} \rangle$ and Einstein's theory plays a crucial role in a systematic practice of prediction and intervention. This is enough to show that we accept (SR) and (SSR); it shows that we believe that masses, space, and time exist mind-independently, and these entities have the properties attributed by Einstein's theory.

²¹⁰ (SR) Most of the essential unobservables of well-established current scientific theories exist mind-independently.

(SSR) Most of the essential unobservables of well-established current scientific theories exist mind-independently, and mostly have the properties attributed to them by science.

Therefore, as long as the success-to-truth rule is construed as a generalization of the infinite disjunction 8), the success-to-truth rule can support scientific realism. In order to explain how the success-to-truth rule generalizes the infinite disjunction 8), we need to assume no theory of truth other than the instances of the equivalence schema (E): $\langle P \rangle$ is true iff P . Therefore, the deflationist has no problem accommodating the role of truth in the success-to-truth rule.

Kitcher may respond to this response by pointing out that the deflationist cannot offer an account of systematic successes. Kitcher may argue that if the success-to-truth rule is nothing but a generalization of 8), then we could not explain systematic successes. What he would mean by ‘systematic successes’ in this context is that many, diverse, scientific theories—for example, Newton’s theory, Einstein’s theory, Boyle and Charles’ theory, and so forth—result in successful predictions and interventions, and the success of all these theories should be explained in terms of a small number of ultimate principles. If the meaning of the success-to-truth rule is exhausted by the infinite disjunction 8), however, we cannot expect that we will find some ultimate principles in terms of which we can explain how all those theories result in successful predictions and interventions. Therefore, says Kitcher, we cannot give an account of systematic successes if we construe the notion of truth invoked in the success-to-truth rule in the deflationary sense.

Two responses can be offered by the deflationist. First, for the purpose of defending scientific realism, it isn’t essential that we give an account of systematic successes. As was discussed earlier, Kitcher places two very different requirements on a success-explanation—the requirement of explaining the objective pertinence of an action and the requirement of explaining systematic successes. Of these two requirements, it is only the requirement of explaining the objective pertinence of an action that is relevant for the defense of scientific realism. The core

theses of scientific realism are (SR) and (SSR), but to defend (SR) and (SSR) doesn't require us to give an account of systematic successes. Whether or not we can derive the success of all those theories from a small number of ultimate principles has no bearing upon endorsing (SR) or (SSR). To be sure, if a success-explanation—an explanation of the success of behavior—can derive all the successful action from a small number of ultimate principles, its explanatory power would increase. The success-to-truth rule, however, isn't any kind of a success-explanation—the purpose of the success-explanation is to defend realism. If the success-to-truth rule were any kind of success-explanation, then one may insist that the notion of truth should play a role in accounting for systematic successes. The purpose of the success-to-truth rule, however, isn't simply to explain the success of an action—it is to show that only the doctrine of scientific realism can explain the success of predictions and interventions. The notion of truth is invoked to make this point—to state that the extra-mental and extra-linguistic conditions of reality plays an essential role in an account of the success of scientific predictions and interventions. Explaining systematic successes is totally orthogonal to the role of truth in the success-to-truth rule. Therefore, it is wrong for Kitcher to require the deflationist to explain systematic successes.

Another response that can be offered by the deflationist is that to invoke the correspondence notion of truth isn't helpful in giving an account of systematic successes. Kitcher maintains that if we construe the notion of truth in terms of the correspondence notion, we can have both—not only can we make sense of the success-to-truth rule but we can also have an account of systematic successes. This expectation can be met, however, only if the correspondence notion of truth is properly defined. Unless the correspondence notion of truth is defined in a clear and substantive way, we would have no idea how the empirical success of many, diverse, scientific theories are derived from the property of correspondence truth. Besides,

unless the correspondence notion of truth is properly defined, there would be no difference between invoking truth and invoking correspondence truth in an account of the success of scientific theories. Kitcher's characterization of correspondence truth, however, is far from being clear and substantive. He rarely explains what the property of correspondence truth is. In particular, Kitcher uses the notion of correspondence relations in two different ways without differentiating one from the other. He sometimes means by it a congruence-relation between a mental/linguistic representation and the states in the world. At other times, he means by it a causal relation between the elements of a mental/linguistic representation and the entities in the world. It hardly makes sense, however, to construe the nature of a correspondence relation between a linguistic/mental representation and the world in terms of a congruence relation. To construe a correspondence relation in terms of a congruence relation makes sense only when we deal with a pictorial representation such as a map. A mental/linguistic representation, however, isn't a pictorial representation.

So, the notion of a correspondence relation, if there is such a relation, should be construed in terms of a causal relation, not a congruence relation. As was discussed in Ch. 1, however, there are many difficulties in defining a correspondence relation in terms of a causal relation. Kitcher's problem, however, is even more acute—Kitcher has to show that there is a causal relation that essentially enters into an explanation of the empirical success of a scientific theory. Kitcher's idea is that we could not explain how scientists succeed in making predictions and interventions without invoking the causal relations between the elements of their beliefs and the entities in the world. Is there however, any causal relation that should exist in order for scientists to make successful predictions and interventions? It seems that as long as scientists have true beliefs—in other words, the scientific theories they accept are true—their predictions

and interventions would be successful. How the entities in the world are causally related to the scientists' beliefs doesn't seem to play an essential role in their making successful predictions and interventions. It is possible that two scientists have the same belief about science while the causal relations between the elements of their beliefs and the entities in the world are not the same. For example, a Martian scientist and a scientist on the earth have the same scientific belief, although the way the elements of a Martian's belief are causally connected with the entities in the world is different from the way the elements of the earthling's belief are causally connected with the entities in the world. To say this doesn't deny the role played by a scientific method in an account of successful predictions and interventions. The causal relation between the elements of a belief and the entities in the world, however, has no bearing upon a scientific method—it bears upon the *semantics* of scientific beliefs. Unfortunately, Kitcher often confuses a methodological question with a semantic question. The deflationist doesn't deny the importance of a methodological question in bringing about successful predictions and interventions. He denies, however, that a semantic question has anything to do with the empirical success of a scientific theory. In order to argue for the correspondence theory of truth, Kitcher has to prove that a semantic question has any bearing upon the empirical success of a scientific theory. It is doubtful, however, that he can do so; a Martian, a blind person, a deaf person, or even an android may have the same success in their predictions and interventions. What matters, thankfully, is the truth of their beliefs, not the semantic mechanism involved in their beliefs.

Chapter 4. Deflationist Views of Meaning and Content

4.1. Introduction

What should a theory of meaning and content be like when the notion of truth is construed in the deflationary sense? This is a big question, and involves vagueness as well. First, the term ‘theory of meaning’ can mean a theory that explicates the nature of the meaning of an expression, a theory that explains the underlying facts by virtue of which an expression has the particular meaning it has, or something else.²¹¹ Second, there are a variety of deflationary theories of truth. So, before discussing what theory of meaning is best suited for a deflationary theory of truth, it is essential to be clear about what the kind of deflationary theory of truth is being considered.²¹² In this chapter, the term ‘theory of meaning’ will be used both as an account of the nature of meaning and as an account of the underlying facts by virtue of which an expression has the particular meaning it has.²¹³ The kinds of deflationary theory of truth being

²¹¹ David Lewis says:

I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and, second, the description of the psychological and sociological facts whereby a particular one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics. (D. Lewis, “General Semantics,” *Synthese*, Vol. 22, 1970, pp. 18-67)

Drawing upon Lewis’s distinction, Jeff Speaks (“Theories of Meaning,” *Stanford Encyclopedia of Philosophy*, 2010) distinguishes a semantic theory from a foundational theory of meaning, where a semantic theory is to assign semantic contents to expressions, and a foundational theory of meaning is to explicate the facts by virtue of which expressions come to have their semantic contents. There seems to be, however, the third kind of a theory of meaning which, like Davidson’s, aims to generate theorems of knowledge of which is sufficient to understand what a speaker of a language says in uttering a sentence in the language. See D. Davidson, *Inquires into Truth and Interpretation*, Oxford: Clarendon Press, 2nd ed., 2001.

²¹² Following Horwich, we may call ‘semantic deflationism’ a combination of a deflationary theory of truth and a deflationary theory of meaning. See P. Horwich, *Meaning*, Oxford: Clarendon Press, 1998, Ch. 1.

²¹³ Horwich uses the term ‘theory of meaning’ as meaning a theory explicating underlying facts (or properties) by virtue of which an expression possesses the particular meaning it has. (P. Horwich, *Meaning*, Ch. 3, p. 43) Field, on the other hand, uses the term ‘theory of meaning’ as meaning both a theory explicating the nature of the meaning of an expression and a theory explicating the facts by virtue of which an expression possesses the particular meaning it

considered are those developed and advocated by Horwich and Field. In this thesis, I have concentrated more on their commonalities than their differences. In discussing their matching, deflationist theories of meaning, I will adopt the same approach—I aim to show that there are more commonalities than there appears to be.

In discussing the commonalities between Horwich’s and Field’s theories of meaning, I will focus on a specific issue—how their theories of meaning deflate the role of truth-theoretic notions in an account of meaning. According to Devitt, one of the most important characteristics of a deflationary theory of meaning and truth is that it denies a systematic connection between an account of meaning and an account of truth.²¹⁴ Considering what motivates the deflationary concept of truth in the first place, what Devitt says makes perfect sense. The most important argument for the deflationary concept of truth is that the role played by the notion of truth is purely expressive as opposed to being explanatory. The truth term such as ‘is true’ is a device that enables us to express things that we could not do easily without it—it enables us to abbreviate a long series of conjunctions (or disjunctions); it enables us to express an agreement (or a disagreement) with a belief that we cannot articulate; and so forth. In this very sense, deflationists construe truth as a purely logical notion not playing an important explanatory role.

So, a deflationist theory of meaning should be deployed without having the notion of truth playing a central role. In this regard, Horwich says, “[W]hen it comes to semantics, truth really has got nothing to with it,” implying that his deflationist view of meaning is opposed to orthodox, mainstream semantics characterizing facts of meaning in terms of truth-theoretic

has. Ironically, Field’s deflationist theory of meaning denies there is either a meaning expressed a word or facts by virtue of which an expression possesses a particular meaning. (See H. Field, *Truth and the Absence of Fact*, Oxford University Press, 2001)

²¹⁴ This is one of the many valuable comments given by Prof. Devitt on this thesis.

terms.²¹⁵ Field agrees—he says, “[A]ccepting deflationism requires dethroning truth conditions from the central place in the theory of meaning and the theory of intentionality . . .”²¹⁶ Thus, Horwich and Field offer theories in which non-truth-theoretic notions play a central role in an account of meaning and content. Horwich’s use theory of meaning and Field’s linguistic view of meaning attributions are those theories respectively.

Both Horwich’s use theory of meaning and Field’s linguistic view of meaning attributions deal with a wide range of issues concerning a theory of meaning and content. Obviously, it is beyond the scope of this thesis to discuss all of them. So, the aim of this chapter is modest—in the first part of this chapter, I will explain the core ideas of Horwich’s use theory of meaning. In discussing some of the core ideas, I will bring in Horwich’s response to Kripke’s skeptical argument regarding Wittgenstein’s use theory of meaning. A brief excusatory remark is needed here—this discussion wasn’t part of my original plan. Working on this topic upon Prof. Levin’s request, however, I got interested in the later Wittgenstein’s view of meaning especially with regard to its influence on contemporary deflationist views of meaning. This is the reason why I will give somewhat lengthy discussion on Kripke’s treatment of Wittgenstein’s view of meaning. In the second part of this chapter, I will compare Horwich’s view of meaning with Field’s view of meaning. As was mentioned earlier, I will deal with this topic by concentrating more on their commonalities than differences. In the last part of this chapter, I will discuss Field’s treatment of a version of success arguments against the deflationist theory of truth and content. The purpose of this discussion is to see how a deflationist can explain the success of

²¹⁵ P. Horwich, “Semantics: What’s Truth Got to Do with It,” in his *Truth—Meaning—Reality*, Oxford: Oxford University Press, 2010, p. 165.

²¹⁶ H. Field, “Disquotational Truth and Factually Defective Discourse,” *Philosophical Review*, 1994, 103, pp.405-52. As Field emphasizes in many of his writings, Field regards one’s stance on the role of truth as being decided by one’s stance on the role of truth conditions. A more discussion on this point will be given shortly.

behavior while she explains the content of a belief-state without assigning a central role to a truth-theoretic notion.

4. 2. Horwich's Use Theory of Meaning and its Response to Kripke's Skeptical Paradox

Horwich says that his use theory of meaning is intended to answer the question: in virtue of which of its underlying properties does a word come to possess a particular meaning.²¹⁷ To understand how Horwich's use theory answers this question, we should first be clear about what he means by the 'meaning' of a word. The meaning of a word, in his theory, is a concept expressed by the word, where a concept is an abstract entity that constitutes a proposition and state of mind (e.g. a belief, desire, etc.).²¹⁸ Horwich uses capitalization convention to name a concept expressed by an expression. For example, 'DOG' refers to the concept, DOG, and it is the meaning of 'dog'; 'SNOW IS WHITE' refers to the concept, SNOW IS WHITE, and it is the meaning of 'snow is white'; and so on. Now, in Horwich's use theory, ordinary meaning attributions of the forms 1) and 2) are construed in terms of 3) and 4) in the following:

- 1) x means <dog>.²¹⁹
- 2) x means that snow is white.
- 3) x means DOG.

²¹⁷ P. Horwich, *Meaning*, Oxford: Clarendon Press, p. 43.

²¹⁸ *Ibid.*, p. 44.

²¹⁹ Here, I use Field's formulation of an ordinary meaning-attribution for a sub-sentential expression, which is intended to be neutral to the question of what sort of entity if the bracket-term refers to. In Horwich's use theory of meaning, the bracket-term refers to the concept expressed by the word inside the bracket-term. Since I will use this formulation in discussing Field's deflationist theory of meaning, I use this formulation for the sake of consistency. See H. Field, "Attributions of Meaning and Content," in his *Truth and the Absence of Fact*, Oxford University Press, 2001, p. 157.

4) x means SNOW IS WHITE.

As will be argued later, however, Horwich's theory of meaning doesn't seem to involve the claim that ordinary meaning-attributions of the forms 1) and 2) are *reducible* to meaning-attributions of the forms 3) and 4). Instead, it involves the claim that meaning-attributions of the forms 3) and 4) are reducible to the use-properties of relevant expressions; i.e., 'dog' and 'snow is white' as they are understood in English. In this sense, Horwich says that his use theory of meaning involves a series of reductive analyses such as 5):

5) x means DOG = $u(x)$
 x means ELECTRON = $v(x)$
 . . . and so on,

where the underlying use-properties $u(x)$, $v(x)$, . . . , and so on stem from the basic acceptance properties possessed by relevant expressions.²²⁰ The basic acceptance properties of an expression are *explanatorily fundamental* properties that designate circumstances in which certain specified sentences containing the term are accepted.²²¹ For example, the basic acceptance property underlying the use-property of 'red' is the disposition to apply 'red' to an observed surface when and only when it is clearly red. Similarly, the explanatorily fundamental acceptance property underlying the use-property of 'dog' is our disposition to utter or accept a sentence such as 'This is a dog' when and only when there is an object having the property of doggyness.²²² The

²²⁰ P. Horwich, *Meaning*, Ch. 3 ("Meaning as Use").

²²¹ P. Horwich, *Meaning*, p. 45.

²²² Here, as Horwich says, the term 'property' is to be understood in a sufficiently fine-grained sense; that is to say, the property of 'doggyness' is distinct from the property of 'caninity' (just as the property of 'water' is distinct from the property of 'H₂O'). According to Horwich, the identity criterion of properties is as follows:

primary explanatory function of these basic acceptance properties is to account for the acceptance of other sentences containing relevant words; for example, the basic acceptance property of ‘dog’ accounts for our acceptance of sentences such as ‘Dogs are friendly animal’ and ‘Dogs are humans’ favorite pet’. So, according to Horwich’s use theory of meaning, what explains a word x ’s having a particular meaning property such as ‘ x means DOG’ is that x has a certain use-property $u(x)$. For example, if the word ‘chien’ in French means DOG, then it should have the use-property $u(x)$ to which the meaning-property ‘ x means DOG’ is reduced. Given that it is a trivial truth that ‘dog’ means DOG, if ‘chien’ means DOG, then the two words ‘chien’ and ‘dog’ should have the same basic acceptance property.²²³ In this sense, Horwich’s use theory of meaning involves the claim that two words express the same concept in virtue of having the same basic acceptance property.²²⁴

What makes Horwich’s use theory of meaning distinctively deflationary is the way in which Horwich appeals to this theory in explaining why many constraints that philosophers have traditionally placed in a theory of meaning are “pseudo”-constraints.²²⁵ There are several pseudo-constraints examined throughout his book *Meaning*—the understanding constraint, the relationality constraint, the representation constraint, the aprioricity constraint, the compositionality constraint, and the normativity constraint.²²⁶ By calling them ‘pseudo-constraint’, Horwich doesn’t intend to weaken the importance of these constraints; he rather intends to remove obstacles in the way of developing a theory of meaning satisfying these

the property f -ness = the property g -ness iff the predicate ‘ f ’ means the same as the predicate ‘ g ’.

See P. Horwich, *Meaning*, p. 21.

²²³ To say that ‘dog’ means DOG is trivially true because ‘DOG’, by definition, names the meaning of the word ‘dog’. See P. Horwich, *Meaning*, p. 15.

²²⁴ P. Horwich, *Meaning*, Ch. 3 (“Meaning as Use”), p. 46.

²²⁵ See P. Horwich, *Meaning*, Ch. 1 (“Introduction”).

²²⁶ P. Horwich, *Meaning*, Oxford: Clarendon Press, 1998, Ch. 2 (“Pseudo-Constraint on an Adequate Account of Meaning”).

constraints.²²⁷ The main obstacle in satisfying these constraints, says Horwich, is a cluster of misguided assumptions. For example, the difficulty with meeting the understanding constraint emerges from the misguided assumption that our understanding the meaning of a word should be explained by an explicit knowledge.²²⁸ Horwich explains why this assumption is misguided by showing how his use theory of meaning satisfies the understanding constraint while it doesn't presuppose the assumption.²²⁹ Taking another example, the normativity constraint is the requirement that an adequate theory of meaning should explain the normative import of a meaning-property. For example, the meaning-property 'x means DOG' has the normative import that *x ought* to be applied only to dog. Hence, the property that reduces this meaning-property must be capable of having such normative import as well. This constraint has been considered extremely difficult to satisfy, because no naturalistic (factual) property is intrinsically normative.²³⁰ According to Horwich, however, this impression emerges from the misconception that a meaning-property is *intrinsically* normative—that the property reducing a meaning-property should be intrinsically normative. Horwich explains why this is a misconception, and then proposes his use theory of meaning as an alternative approach in which the normative import of a meaning-property is explained in terms of a factual property; i.e., a use-property.²³¹

In this way, Horwich argues for the dual aspects of his deflationist view of meaning—eliminating various misconceptions plaguing a theory of meaning, and providing an alternative

²²⁷ *ibid.*

²²⁸ P. Horwich, *Meaning*, pp. 16-18.

²²⁹ Since the main focus of the discussion will be the relationality constraint and the representation constraint, the detailed account of other constraints will be omitted. To be brief, however, Horwich's point about the understanding constraint is that the knowledge constituting our understanding the meaning of a word is *implicit*. Our knowledge of the use-property of a word is implicit in that although we cannot articulate it to ourselves consciously or unconsciously, we know how to deploy the word in various circumstances.

²³⁰ As Horwich says, Kripke's skeptical argument in his *Wittgenstein on Rules and private Language* is a good example of this kind. See P. Horwich, *Meaning*, Ch. 8 ("Norms of Language") and S. Kripke, *Wittgenstein on Rules and Private Language*,

²³¹ See P. Horwich, *Meaning*, Ch. 2 & 8.

theory of meaning that meet all the constraints listed above. It would be nice if I could go over each of his solutions to these constraints. Instead, I will in the following focus on Horwich's solutions to two of those constraints—the relationality constraint and the representation constraint. Horwich's solutions to these constraints are particularly relevant to his response to Kripke's skeptical argument against Wittgenstein's use theory of meaning. To be sure, Kripke's argument is concerned with the normativity constraint—that there is no fact in the world that can explain the normative import of a meaning-property. Kripke's skeptical argument, however, is also concerned with the relationality and representation constraint. Especially, Horwich responds to Kripke's skeptical argument by relating it to the relationality and representation constraints. Moreover, Horwich's treatment of these constraints shows how the inflationary conception of truth engenders some of the misguided assumptions presupposed by traditional theories of meaning.

The relationality constraint stems from the surface form of a meaning-attribution: ' x means y ', where x is to be replaced by a phonological and orthographic entity and y is to be replaced by its meaning. Given that a meaning-attribution has such a relational form, an adequate theory of meaning must explain, first, the nature of meaning-entities, and second, the nature of the meaning-relation in which expressions stand to the meaning entities. As was mentioned, in Horwich's use theory of meaning, meaning-entities are concepts expressed by words or sentences; for example, DOG or SNOW IS WHITE. But then, to meet the relationality constraint, Horwich's use theory of meaning should explain the nature of the meaning-relation between expressions and concepts.

In explaining the nature of the meaning-relation, Horwich's use theory of meaning makes an important departure from other theories of meaning. Most theories of meaning that have been

seriously entertained assume that the meaning-relation should be reduced to a relational property.²³² If we accept this assumption, we would require that the relational property ‘*x* means DOG’ be analyzable into a relational property of the form 6) in the following:

6) $T(x, \text{dog})$.²³³

Horwich argues, however, that nothing compels us to analyze a meaning-property into 6); to think otherwise involves several misconceptions. First, it involves committing the *Constitution Fallacy*—to suppose that what constitutes a relational property has to be itself a relational property.²³⁴ The surface form of a meaning-property is relational, but it doesn’t follow from this that the property by virtue of which an expression has a meaning-property should be relational as well.²³⁵ One may object here that if what constitutes a meaning-property isn’t relational, a meaning-property isn’t genuinely relational, either. The problem with this objection is the obscurity with the notion of being “genuinely relational”; the distinction between a relational and non-relational property is obscure in the first place, and so it is by no means clear what a

²³² According to Horwich, the Davidsonian truth-conditional theory, the teleological theories defend by Papineau, Dretske, Millikan, and Jacob, and Kripke’s skeptical argument against the naturalistic reduction of meaning-facts implicitly assume this idea. See P. Horwich, *Meaning*, Ch. 2.

²³³ I omitted an explanation for the sake of simplicity. The omitted explanation is as follows. Assuming that a meaning-property should be reduced to a relational property, ‘*x* means DOG’ would first be analyzed into $T^*(x, \text{DOG})$. Since ‘DOG’ names the meaning of ‘dog’, T^* couldn’t be construed as a non-semantic relation. In order to reduce the semantic property ‘*x* means DOG’ into a non-semantic, relational property, ‘*x* means DOG’ has to be further analyzed into $T(x, \text{an exemplification of DOG (i.e., doggyness)})$. This leads us to 6): $T(x, \text{dog})$. See P. Horwich, *Meaning*, pp. 21-27.

²³⁴ More precisely, the Constitution Fallacy (Horwich uses capital letters) is an error to suppose that reductive analysis must preserve logical structure. In other words, it is a fallacy to assume that whenever a fact has a certain component, then whatever constitutes this fact must contain the same component or something that constitutes it (see P. Horwich, *Meaning*, p 21).

²³⁵ I suppose Horwich is right about this; for example, John may have the relational property ‘*x* is taller than Tom Cruise’ by virtue of having the non-relational property ‘*x* is 6 feet tall’. Field expresses a disagreement with Horwich’s constitution fallacy (in his *Truth and the Absence of Fact*, p. 71). I won’t discuss it here.

“genuinely relational” property consists in.²³⁶ That doesn’t mean that there is no distinction at all between relational and non-relational properties; for example, there is a difference between properties such as ‘*x* kicks *y*’ or ‘*x* loves *y*’ and properties such as ‘*x* is sick’ or ‘*x* is alive’.

Horwich takes note of such difference by distinguishing a *strongly* relation property and a *weakly* relational property. A meaning-property would be strongly relational if it were constituted by a relational property at a deeper level; and if it were, we would be justified in analyzing ‘*x* means DOG’ into 6) above: $T(x, \text{dog})$. But the surface form of a meaning-property doesn’t by itself show that a meaning-property is constituted by a relational property at its deepest level. In this sense, Horwich argues that most meaning theories have assumed a misconception that a meaning-property is a *strongly* relational. Horwich calls this misconception ‘the strongly relational constraint’, and this is another misconception associated with the relationality constraint.

Freed from these misconceptions, Horwich argues that we can explain the nature of the meaning relation without invoking a relational property such as ‘ $T(x, \text{dog})$ ’. Thus, Horwich claims that what constitutes a meaning-property such as ‘*x* means DOG’ is simply the following:

- 7) The concept, DOG, is present within some mental state of a speaker who utters *x*.²³⁷

We infer 7) from the use-property $u(x)$ which the meaning-property ‘*x* means DOG’ is reduced to; namely, by observing how a speaker uses a certain word, we infer 7). So, Horwich’s use theory

²³⁶ For example, is a property such as ‘*x* is red’ a relational property or not? We tend to think it to be a non-relational property. However, in order for an object to have this property, it should stand in relations to certain conditions; indeed, according to Locke, an object must stand in relations to a perceiver. But then, it is hard to say that the property ‘*x* is red’ is “genuinely” relational or non-relational.

²³⁷ See P. Horwich, *Meaning*, pp. 20-21.

satisfies the relationality constraint easily while it doesn't presuppose the strong relationality constraint.

Horwich's solution to the relationality constraint is closely connected to his solution to the representation constraint. Interestingly, the representation constraint is in turn closely related to Kripke's skeptical argument against Wittgenstein's use theory of meaning. More precisely, Kripke's complaint about Wittgenstein's use theory of meaning is that it doesn't meet the representation constraint. The representation constraint is that an adequate theory of meaning must explain how an expression comes to represent a certain part of the world by virtue of possessing a certain meaning-property. A rationale for this constraint is that we accept a meaning-to-truth conditional such as 8) in the following:

8) x means DOG $\rightarrow x$ is true of all and only dogs.

For example, we accept that if the word '개' means DOG, then '개' should be true of all and only dogs. The question is then how to explain the connection between the antecedent and the consequent of a meaning-to-truth conditional such as 8).

The core idea of Kripke's argument against Wittgenstein's dispositional (use) theory of meaning is that meaning-properties such as 'x means DOG' cannot be reduced to a dispositional, use property $u(x)$ because a use-property $u(x)$ doesn't determine a representational (truth-theoretic) property such as 'x is true of all and only dogs'.²³⁸ Here, Horwich summarizes Kripke's argument as follows:

²³⁸ The main references for Kripke's skeptical argument and Horwich's response are: S. Kripke, *Wittgenstein on Rules and Private Language*, Cambridge: Harvard University Press, 1982; L. Wittgenstein, *Philosophical Investigations*, tran., by G. E. M. Anscombe, Basil Blackwell, 1958; P. Horwich, *Meaning*, Oxford: Clarendon

- i) Whatever constitutes the meaning of a predicate must determine its extension.
- ii) The use of a predicate doesn't determine its extension.

Therefore,

- iii) The meaning of a predicate isn't constituted by its use.

To substantiate the premise ii) Kripke brings in the famous “quus”-example; when I answer ‘125’ to the problem ‘68+57’, is my answer justified by my disposition to use the plus sign (‘+’ or ‘plus’) in a certain way? Kripke’s answer is that no fact about my disposition to use the plus sign justifies that ‘+’ means PLUS as opposed to QUUS. Kripke (the self-proclaimed skeptic) says:

The sceptic argues that when I answered '125' to the problem '68+57', my answer was an unjustified leap in the dark; my past mental history is equally compatible with the hypothesis that I meant quus, and therefore should have said '5'.²³⁹

Here, an assumption is that I have never performed a computation that involves a number bigger than 57.²⁴⁰ Another assumption is that there is the quus-function (symbolized ‘ ω ’ as) that is defined as follows: $x \omega y = x + y$ if $x, y < 57$, and $x \omega y = 5$ otherwise. But then, we can justify ‘68 + 57 = 125’ only if we can justify the metalinguistic assertion that my employment of the plus sign (‘+’) means PLUS, not QUUS. Kripke’s skeptical conclusion is that nothing justifies this metalinguistic assertion; in particular, my disposition to use the plus sign can never justifies that it means PLUS as opposed to QUUS. Kripke emphasizes that the point of his skeptical

Press, 1998; P. Horwich, “Meaning, Use, and Truth,” *Mind* Vol. 104, 1995, pp. 355-368; P. Boghossian, “The Rule-Following Considerations,” *Mind* Vol. 98, 1989, pp. 507-549; A. Miller, “Horwich, Meaning, and Kripke’s Wittgenstein,” *The Philosophical Quarterly*, Vol. 50, No. 199, 2000, pp. 161-174.

²³⁹ S. Kripke, *Wittgenstein on Rules and Private Language*, Cambridge: Harvard University Press, 1982, p. 15.

²⁴⁰ S. Kripke, *ibid.*, pp. 8-9.

argument isn't epistemological—it isn't that we don't have an access to the fact by virtue of which the plus sign means PLUS as opposed to QUUS.²⁴¹ The point is rather semantic—that there is simply no fact of the matter determining that I meant PLUS by the plus sign.²⁴² Once he establishes such skeptical conclusion, Kripke deploys his skeptical solution. Before discussing Kripke's skeptical solution, however, I shall first examine Horwich's response to Kripke's skeptical argument.

Horwich offers a two-pronged response to Kripke's skeptical argument. Its first prong is that Kripke's argument is threatened to be invalidated by equivocation—the concept of determination can be construed in two different ways, but Kripke's argument rests on only one of them.²⁴³ Its second prong is that the concept of determination Kripke's argument rests on involves the inflationary conception of truth. I shall both points in the following. First, concerning the problem of equivocation, Horwich claims that we should make the distinction between the strong and the weak determination of the representational property of an expression by its use-property. That is to say, the use property of a predicate either strongly determines (DETERMINES) or weakly determines (determines) its extension. The use property of a predicate DETERMINES its extension when we can *read off* the extension of the predicate from its use-property.²⁴⁴ According to Horwich, it is this notion of determination that Kripke's skeptical argument rests on; namely, the skeptical conclusion iii) above follows from the premises i) and ii) only when 'determines' is understood as meaning DETERMINES. Construed

²⁴¹ S. Kripke, *ibid.*, p. 21.

²⁴² Kripke says that even God looked at my mind, he could not find the fact by virtue of which I meant addition (PLUS) by 'plus'. Kripke's skeptical argument, however, isn't just intended to show that there is no internal fact by virtue of which an expression means anything; it is also intended to show that there is no external fact determining the meaning of an expression. Kripke makes this point several times throughout his *Wittgenstein on Rules and Private Language*. See especially the part in which he makes an analogy to Hume's skeptical argument against private causation (S. Kripke, *ibid.*, p. 69).

²⁴³ P. Horwich, "Meaning, Use, and Truth," *Mind* Vol. 104, 1995, pp. 355-368.

²⁴⁴ P. Horwich, *ibid.*, and *Meaning*, Ch. 10.

this way, the crux of Kripke's objection to Wittgenstein's use theory of meaning is that we cannot read off the extension (any other truth-theoretic properties) of an expression from its use property.

What's wrong, however, with the concept of DETERMINATION? Horwich's answer is that it assumes the inflationary concept of truth. To see why let's examine what we need in order to be able to read off the extensional (truth-theoretic) property such as 'x is true of all and only dogs' from the use-property, $u(x)$. What we need, says Horwich, is an inflationary principle of the form 9) in the following:

$$9) \quad u(x) \rightarrow T(x, \text{dog}),$$

where ' T ' names a non-semantic, naturalistic relation. It goes without saying that we would be able to read off the extension of a predicate from its use property if we had a principle of the form 9). But consider what it amounts to accept 9)—it amounts to accepting an inflationary definition of *being true of* such as 10) in the following:

$$10) \quad x \text{ is true of all and only dogs} \leftrightarrow T(x, \text{dog}).$$

A definition of the form 10) is inflationary, since it offers a substantive analysis of the truth-theoretic concept *being true of*—it analyzes this truth-theoretic property in terms of a naturalistic property. From the deflationist perspective, however, it is neither needed nor possible to analyze the property of being true of in terms of a naturalistic property. The deflationist construes truth as a purely logical or expressive device the nature of which can exhaustively explained by the

equivalence schema (E): $\langle P \rangle$ is true iff P . Similarly, in the deflationary truth, the concept of being true of can exhaustively explained by what is implied by the instance of schemas such as

The concept F is true of $y \leftrightarrow F(y)$.²⁴⁵

So, when we construe truth-theoretic concepts such as being true of in the deflationary sense, we don't need inflationary principles such as 9) and 10). But we need these principles if we are to be able to *read off* the extension of a predicate from its use property. Therefore, when Kripke criticizes Wittgenstein's use theory on the grounds that we cannot read off the extension of a predicate from its use property, Kripke assumes the inflationary conceptions of truth-theoretic notions. Put more precisely, Kripke's assumes that meaning-constituting properties are *strongly relational* in the sense that meaning-properties strongly determine (DETERMINE) the truth-theoretic properties of expressions. Horwich's point is that these are "pseudo-constraints", which are ultimately rooted in the inflationary conceptions of truth-theoretic notions. When we construe truth-theoretic notions in the deflationary sense, it is only expected that we cannot read off truth-theoretic properties from use properties. But then, Kripke's skeptical argument doesn't provide a legitimate reason to object to Wittgenstein's use theory of meaning.

I agree with Horwich's response to Kripke's skeptical argument—most of all, the concept of meaning presupposed by Kripke's skeptical argument is not Wittgensteinian.²⁴⁶ Since

²⁴⁵ See P. Horwich, *Meaning*, Ch. 2.

²⁴⁶ Horwich makes the same point. He says:

Just because most of the properties we encounter have one, we shouldn't assume that all do. Such an assumption about truth-the inflationary view-would seem to be a paradigmatic Wittgensteinian example of a philosophical misconception and pseudo-problem generated by an overdrawn linguistic analogy.

the Wittgensteinian concept of meaning seems to be closely related to the deflationist view of meaning, I will briefly discuss in the following some problems with Kripke's criticism of Wittgenstein's use theory of meaning.

Kripke's skeptical argument involves two distinct arguments neither of which shows that dispositional use properties don't constitute meaning properties. First, it involves the argument questioning whether I will have the same disposition to use the plus sign the disposition that I have had. Kripke says that it is possible that I proceed to compute—in conflict with my previous linguistic intentions—that $68 + 57 = 5$.²⁴⁷ On the basis of this consideration, Kripke concludes that nothing in my mind determines whether I mean PLUS or QUUS by the plus sign.²⁴⁸ This argument doesn't seem valid; even if we agree that my past and present dispositions don't determine my future dispositions, from this we cannot infer that my dispositions to use the plus sign don't determine its meaning-constituting properties. Suppose I proceed to compute '68+57=5.' If this happens, we would have to say that I have been meaning QUUS by the plus sign all along. By Kripke's own definition, meaning QUUS involves an abrupt change of dispositions—to compute 'x \oplus y = x + y if x, y < 57' but to compute 'x \oplus y = 5 otherwise'. This only proves that meaning-properties are constituted by dispositional properties. Kripke may respond back that at the present moment we cannot determine whether my plus sign means PLUS or QUUS (because my present disposition is compatible with meaning PLUS and meaning QUUS). The problem with this response is that the issue, as Kripke says himself, isn't

See P. Horwich, "Meaning, Use, and Truth," *Mind* Vol. 104, 1995, pp. 355-368, p. 359.

²⁴⁷ See S. Kripke, *ibid.*, p. 9. (He says that although it sounds insane to say so, it is still logically possible.)

²⁴⁸ S. Kripke, *ibid.*, p. 82. Kripke says,

The point is the sceptical problem, outlined above, that anything in my head leaves it undetermined *what* function 'plus' (as I use it) denotes (plus or quus), what 'green' denotes (green or grue), and so on.

epistemological—whether *we* can determine the meaning of the plus sign is orthogonal to the issue. The issue is whether meaning-properties *are* constituted by dispositional properties. So, even if we cannot determine, by observing how I am disposed to use the plus sign, whether my plus sign means PLUS or QUUS, it shouldn't undermine the dispositional theory of meaning. The question, therefore is whether Kripke's counterexample (the quus-example) shows that meaning-properties are not constituted by dispositional use-properties. The answer is that it doesn't; to the contrary, it only adds weight to the dispositional theory. Notice how Kripke defines the meaning-property 'x means QUUS'—he defines it as 'x \circ y = x + y if x, y < 57, and x \circ y = 5 otherwise'. This definition can easily be accommodated by the dispositional theory of meaning—what constitutes meaning QUUS is the disposition to compute numbers in the same way you compute an addition (x + y) if numbers are smaller than 57, but compute them 5 otherwise. Similarly, what constitutes 'x means PLUS' is the disposition to compute numbers in the way we compute addition for all numbers. Surely, my past and present disposition to use an expression are compatible with many different semantic hypotheses; that it means PLUS, it means QUUS, it means QUUS*, and so forth. So, we cannot *determine*, just on the basis of dispositional use-properties—whether an expression means anything. But by Kripke's own definition, meaning-properties are constituted by dispositional use-properties—what makes a sign mean QUUS is how a user of the sign is disposed to compute numbers. Despite that Kripke warns us not to confuse an epistemological question with a semantic question, he himself mixes up these two questions.

Kripke's another argument is that whereas the notion of meaning has an infinitary character, the notion of dispositions doesn't; we are finite beings, so our dispositions are also finite. Therefore, meaning-properties cannot be constituted by dispositional, use-properties. It is

doubtful, however, whether meaning has an infinitary character; is it really an infinite matter that the plus sign ('+') means PLUS? Although Wittgenstein says at times as if meaning is an infinite matter (a rule is like a rail to infinity), he also says that a rule can be altered.²⁴⁹ Wittgenstein says that following a rule is like a practice or custom.²⁵⁰ We usually follow a practice or custom blindly, without asking for a justification, but a practice or custom can change. Therefore, it seems wrong to take what Wittgenstein means by 'infinity' literally. Instead, we should take it as saying that although the first step of following a rule contains its future steps, it does so only finitely, not infinitely.²⁵¹ But then, the notion of a disposition best captures Wittgenstein's notion of meaning; a dispositional property has the kind of the regularity that a meaning property has, but a dispositional property, like a meaning property, can change over time.

So, Kripke's argument begs the question—it involves the dubious assumption that a meaning property is fundamentally different from a dispositional use property. Kripke assumes that whereas it is an infinite matter that the plus sign ('+') means PLUS, it is a finite matter that a speaker (or a linguistic community as a whole) has a relevant disposition. It is by no means clear, however, that we have to understand the notion of meaning in such an exaggerated way. Neither does Wittgenstein construe the notion of meaning in this way.

²⁴⁹ See §218 for Wittgenstein's analogy between a rule and a rail. Also consider the following:

The whole time they are playing a ball-game and following definite rules at every throw. And is there not also the case where we play and—make up the rules as we go along? And there is even one where we alter them—as we go along. (L. Wittgenstein, *ibid.*, §83, p. 39).

²⁵⁰ See L. Wittgenstein,

²⁵¹ Wittgenstein says,

It may now be said: "The way the formula is meant determines which steps are to be taken". What is the criterion for the way the formula is meant? It is, for example, the kind of way we always use it, the way we are taught to use it. (*Philosophical Investigations*, §190.)

The most serious problem with Kripke's criticism of the use theory, however, is that his own solution to the skeptical problem—the skeptical solution—isn't really different from the use theory of meaning. The core idea of Kripke's skeptical solution is that we should understand a meaning-attribution such as 'John means PLUS by the plus sign ('+')' in terms of its assertability condition, not truth condition.²⁵² Kripke says,

All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives. No supposition that 'facts correspond' to those assertions is needed.²⁵³

So, on this view, we are entitled to assert 'John means PLUS by the plus sign' if John's responses to sum-questions seem acceptable from the perspective of the members of his linguistic community. This is the reason why Kripke considers Wittgenstein's argument against private language a key to the solution to the skeptical paradox. Just as Hume solves the problem of causation by recognizing the impossibility of private causation, so does Wittgenstein solve the skeptical paradox by recognizing the impossibility of private language.²⁵⁴ When a person is *considered* isolated, the meaningfulness of a meaning attribution such as 'Jones means PLUS by the sign '+' cannot be made sense of.²⁵⁵ Kripke says:

Wittgenstein finds a useful role in our lives for a 'language game' that licenses, under certain conditions, assertions that someone 'means such-and-such' and that his present application of a word 'accords' with what he 'meant' in the past. It turns out that this role, and these

²⁵² S. Kripke, *Ibid.*, p. 74 & P. 87.

²⁵³ S. Kripke, *ibid.*, p. 78.

²⁵⁴ S. Kripke, *ibid.*, pp. 62-113.

²⁵⁵ S. Kripke, *ibid.*, p. 110. Kripke says here, interestingly, that his point is not that a physically isolated person but a person who is considered to be isolated cannot be said to follow rules. But then, can Robinson Crusoe be said to follow rules if he compares his present use with his past use of a word? It seems that Kripke would say yes, because according to his account it is not another person but another point of view that is essential for a meaning-attribution.

conditions, involve reference to a community. They are inapplicable to a single person considered in isolation. Thus, as we have said, Wittgenstein rejects 'private language'. . .²⁵⁶

We expect another person's usage of '+' to agree with ours because our lives are dependent upon it. Therefore, the meaning-attribution 'John means PLUS by '+' doesn't involve John alone; it involves the whole linguistic community that he belongs to. Only when a meaning attribution is conceived from this non-individualistic point of view can we make sense of the meaningfulness of a meaning attribution.²⁵⁷ This is the gist of the skeptical solution that Kripke attributes to Wittgenstein.

But if this is Kripke's own solution to the skeptical problem, it is mysterious why he objects to the use theory of meaning in the first place.²⁵⁸ How can Wittgenstein's argument against private language be different from the dispositional use theory of meaning? There seems no way to understand the 'assertability condition' associated with an expression independently of the way the members of a linguistic community are disposed to use the expression. So, the problem with Kripke's skeptical solution to the skeptical problem isn't that his solution is wrong; it is that his skeptical solution is inconsistent with his criticism of the use theory of meaning. Kripke's skeptical solution *supports*—rather than criticizes—the use theory of meaning.

4.3 The Commonalities Between Horwich's and Field's Deflationist Views of Meaning

²⁵⁶ S. Kripke, *ibid.*, p. 79.

²⁵⁷ S. Kripke, *ibid.*, p. 97 & p. 109.

²⁵⁸ As Horwich says, Kripke sympathetically presents Wittgenstein's view of meaning, not attacking it. (P. Horwich, "Meaning, Use, and Truth," *Mind*, Vol. 104 (New Series), 1995, pp. 355-368, p. 355) So, Kripke's skeptical argument can be understood as an indirect means to advocate Wittgenstein's view of meaning. There still remains the question the legitimacy of this method. A hospitable way of construing Kripke's intention is that he employs the skeptical paradox as a means of highlighting the contrast between the early Wittgenstein's view of meaning with the later Wittgenstein's.

In this section, I will compare Horwich's view of meaning with another deflationist's—Field's—view of meaning. In comparing their deflationist views of meaning, I found some interesting and complicated dynamics among Wittgenstein, Kripke, Horwich, and Field. First, Wittgenstein's *Philosophical Investigations* seems to contain all the important seeds for contemporary deflationist views of meaning. Its influence on Horwich's view is obvious—Horwich says that the purpose of his book *Meaning* is to clarify and support Wittgenstein's view that the meaning of a word is its use in the language.²⁵⁹ Its influence on Field's deflationist view is less obvious—Field indeed doesn't mention the later Wittgenstein's view of meaning at all.²⁶⁰ Field's deflationist view of meaning, however, bears interesting similarities with the later Wittgenstein's view of meaning.²⁶¹ Consider what Dummett says about the later Wittgenstein's view of meaning:

[T]he *Investigations* contains implicitly a rejection of the classical (realist) *Frege—Tractatus* view that the general form of explanation of meaning is a statement of the truth conditions.²⁶²

Compare this passage with what Field says in the following:

I see the philosophy of language, and the part of philosophy of mind concerned with intentional states like believing and desiring and intending and the like, as pretty much bifurcated into two traditions. The traditions differ over the role that the notion of truth conditions plays in the theory of meaning and in the theory of the content of intentional

²⁵⁹ P. Horwich, *Meaning*, Oxford: Clarendon Press, 1999, Ch. 10, "A Straight Solution to Kripke's Skeptical Paradox," p. 224.

²⁶⁰ As far as I know, Field only mentions the early Wittgenstein as an advocate of the "inflationary" theory of meaning and content.

²⁶¹ One of the important similarities is that both the later Wittgenstein and (the later) Field are skeptical about the explanatory value of invoking meaning as entities. For example, Wittgenstein seems to suggest that we understand the notion of a proposition in terms of the use properties of words (propositions are not some queer entities doing some queer things). See *Philosophical Investigations*, §93-96 & §120. H. Field, "Deflationist Views of Meaning and Content" and "Attributions of Meaning and Content" both in his *Truth and the Absence of Fact*.

²⁶² M. Dummett, "Wittgenstein's Philosophy of Mathematics," *The Philosophical Review*, Vol. 68, 1959, pp. 324-348, p. 348.

states. One of the traditions, whose early advocates include Frege, Russell, the early Wittgenstein, and Ramsey, has it that truth conditions play an extremely central role in semantics and the theory of mind; a theory of meaning or content is, at least in large part, a theory of truth conditions. . . . However, it isn't easy to say precisely what this central role is that truth conditions allegedly play; that I think is a main motivation for the alternative tradition.²⁶³

Field considers the Frege-Russell-(the early)Wittgenstein-Ramsey tradition the “inflationary” tradition, where its distinctive feature is assigning a central role to the notion of a truth condition in an account of meaning and content. Field finds this inflationary tradition of meaning and content troublesome; he doubts the role that a truth condition allegedly has.

Besides, Field denies that there is a fact of the matter about meaning-attributions, which concurs with Kripke’s skeptical interpretation of the later Wittgenstein’s view of meaning. Relatedly, Field denies that translations and meaning-attributions are matters of being *correct* or *incorrect*; they are matters of being *better* or *worse*.²⁶⁴ Again, according to Kripke, this is the later Wittgenstein’s view of meaning. According to Kripke, all we can afford for a meaning-attribution is the assertability condition, and not the truth condition. So, although Field also defends a deflationist view of meaning, his view of meaning has more skeptical elements than Horwich’s.²⁶⁵ On Horwich’s view, skepticism about a theory of meaning is largely due to mistaken constraints placed on an adequate theory of meaning.²⁶⁶ Related to this point, Horwich disagrees with Kripke’s claims that meaning-attributions have only assertability conditions, and not truth condition. Horwich says “[T]he ‘use-theoretic’ truth conditions for attributions of

²⁶³ H. Field, “Deflationist View of Meaning and Content,” in his *Truth and the Absence of Fact*, Oxford University Press, 2001, p. 104.

²⁶⁴ H. Field, “Deflationist Views of Meaning and Content,” in *Truth and the Absence of Fact*, p. 129.

²⁶⁵ In a sense, Field’s view of meaning is located in the middle way between Kripke’s and Horwich’s views of meaning. Like Kripke, Field’s view is leaning toward skepticism about meaning. Like Horwich, however, Field is a thoroughgoing deflationist about the concept of truth. This dilutes the seriousness of Field’s skepticism about meaning; although Field may be sympathetic to Kripke’s skeptical paradox (that there is no fact of the matter as to meaning), he is less concerned about it.

²⁶⁶ See P. Horwich, *Meaning*, Ch. 1.

meaning are entailed—and are certainly not precluded—by the assertibility conditions that Kripke endorses”.²⁶⁷

In situations like this, a natural way to discuss Horwich’s and Field’s deflationist views of meaning may be to take sides with one of them. That’s not, however, how I will deploy the discussion in the following—I will rather focus on the commonalities between Horwich’s and Field’s deflationist views of meaning. There are two reasons why I take this approach; first, the approach is helpful in having an overarching view of what a deflationist view of meaning and truth is about. After all, they are two leading deflationists whose views of truth are more or less in agreement. So, by focusing on the commonalities of their views of meaning, one can develop a holistic understanding of what the deflationary approach to truth and meaning is about. Second, Horwich’s and Field’s accounts of meaning have more in common than they appear to. This shouldn’t be taken as saying that there is no important difference between their accounts of meaning. Many of their differences, however, are not the differences in principle—they are differences in how to accomplish the ultimate goal of the deflationary concept of truth; i.e., the idea that the role of truth is expressive rather than explanatory. In order to discuss the commonalities between Horwich’s and Field’s views of meaning in an orderly manner, I will divide the discussion into two large topics: meaning-entities and a fact of the matter (about meaning-attributions).

4.3.3 Meaning-Entities

In Field’s deflationist view of meaning and content, the question of the object of a meaning-relation looms large. More precisely, one of the main questions discussed in Field’s deflationist view of meaning is whether we are to construe a meaning-attribution as a relation to

²⁶⁷ P. Horwich, *Meaning*, p. 223 (“A Straight Solution to Kripke’s Sceptical Paradox”).

a *meaning-entity*, where a meaning-entity is construed as an extra-mental/extra-linguistic, abstract entity. According to the orthodox (traditional) view of a meaning-attribution, we should construe a meaning-attribution as a relation to a meaning entity.²⁶⁸ After all, the term ‘mean’ involved in a meaning-attribution is a two-place transitive verb just like ‘love’ in ‘Mary loves John’. This surface structure of a meaning-attribution—that it has the form ‘*x* means *y*’—supports the traditional view that there should be something referred to by an expression that replaces ‘*y*’. Moreover, what is referred to by an expression replacing ‘*y*’ should be an abstract entity—it cannot be located in space and time; it should also be a language-independent entity, since what is meant by an expression (for example, what is meant by the Korean word ‘가’’) isn’t dependent on the existence of a particular language; and it should also be a mind-independent entity, since the meaning of an expression exists independently of what anyone says or thinks.

According to Field, however, we’d better not construe a meaning-attribution as a relation to a meaning-entity. Instead, we’d better construe it as a relation to a linguistic entity; namely, an expression in a meaning attributor’s language. Field calls this view ‘the linguistic view of a meaning attribution’.²⁶⁹ So, if we accept this view, we are to understand ordinary meaning attributions of the form 1) and 2) in terms of 1^L) and 2^L) in the following

- 1) *x* means <dog>.²⁷⁰
- 2) *x* means that snow is white.

²⁶⁸ I extended Schiffer’s face-value theory of a belief-attribution to a meaning-attribution. See S. Schiffer, *The Things We Mean*, Oxford University Press, 2003, Ch. 1 (“The Face-Value Theory”).

²⁶⁹ H. Field, “Attributions of Meaning and Content,” in his *Truth and the Absence of Fact*, Oxford University Press, 2001.

²⁷⁰ Field’s use of brackets can be understood in the same way Horwich uses capital letters such as DOG to formally represent a meaning attribution for a subsentential expression. Field says that a formulation such as 1) is intended to be neutral to the question of what sort of entity if any bracket-term refers to. See H. Field, “Attributions of Meaning and Content,” in his *Truth and the Absence of Meaning*, Oxford University Press, 2001, p. 157.

1^L) Means-the-Same-as (x , ‘dog’).

2^L) Means-the-Same-as (x , ‘snow is white’),

where ‘dog’ and ‘snow is white’ are a meaning-attributor’s words as they are computationally typed.²⁷¹ On this view, to say that an expression x means ‘dog’ is to say that it means the same as the word ‘dog’ does in my own language; also, to say that a sentence x means that snow is white is to say that it means the same as the sentence ‘Snow is white’ does in my own language. So, this view not only avoids invoking a meaning-entity in an analysis of meaning-attributions, it also revises the logical form of meaning-attributions; whereas ‘means’ is a transitive verb, ‘means the same as’ is a comparative verb on the par with ‘is as tall as’. Both involve relations but relations of different kinds; whereas ‘means’ expresses a non-symmetric relation, ‘means the same as’ expresses a symmetric relation. Field isn’t the first advocating the linguistic analysis of meaning- and content-attributions; it seems that his linguistic view of meaning- and content-attributions was influenced by Carnap’s linguistic analysis of propositional-attitude attributions.²⁷² In my thesis, however, I won’t discuss the historical background of Field’s linguistic view of a meaning- and content-attribution. Instead, I will devote the discussion for the examination of its implication in the debate between the inflationary and the deflationary views

²⁷¹ In Field’s theory, to say that a word is computationally typed is to say that a word is typed in terms of its computational role, which is in turn characterized in terms of its inferential/conceptual roles and indication relations. See H. Field, “Deflationist Views of Meaning and Content.”

²⁷² Field mentions Church’s criticism of Carnap’s sententialist (non-propositionalist) analysis of a propositional attitude ascription. (A. Church, “On Carnap’s Analysis of Statements of Assertion and Belief,” *Analysis* Vol. 10. 1950, pp. 97-99) Quine’s sententialist approach may be another influence on Field’s linguistic view of meaning attributions. (W. V. Quine, “Quantifier and Propositional Attitudes,” *The Journal of Philosophy* Vol. 53, 1956, pp. 177-187.) (See H. Field, “Deflationist Views of Meaning and Content” and “Attributions of Meaning and Content,” both in his *Truth and the Absence of Fact*.)

See H. Field, “Deflationist Views of Meaning and Content” and “Attributions of Meaning and Content,” both in his *Truth and the Absence of Fact*.

of meaning. For that purpose I shall first contrast this view with Horwich's analysis of a meaning attribution.

Horwich says that he defends a non-revisionist (traditional or orthodox) approach to a meaning attribution;

Attributions of meaning in the style of “‘dog’ means DOG”, “‘Il pleut’ means IT'S RAINING”, and so on, will be taken at face value as claiming, in each case, that the relation ‘x means y’ links a certain noise or mark (such as the English word ‘dog’ or the French sentence ‘Il pleut’) with a further, more mysterious entity—a meaning (such as DOG or IT'S RAINING).²⁷³

So, on Horwich's deflationist view of meaning, meaning attributions of the form 1) and 2) are analyzed in terms of 1^M) and 2^M) in the following:

1^M) Means (x, DOG).

2^M) Means (x, SNOW IS WHITE),

where DOG and SNOW IS WHITE are meaning-entities; they are non-linguistic, abstract entities which are the meanings of ‘dog’ and ‘Snow is white’ respectively.²⁷⁴

Viewed this way, Field's and Horwich's analyses of meaning-attributions are vastly different from each other; whereas Horwich analyzes the logical form of a meaning-attribution

²⁷³ P. Horwich, *Meaning*, Oxford: Clarendon Press, 1999, p. 4.

²⁷⁴ Horwich identifies meanings with concepts; for example, DOG is the concept expressed by ‘dog’, and as such it is the meaning of ‘dog’. SNOW IS WHITE is the concept expressed by the indicative sentence ‘Snow is white’, and as such it is the meaning of ‘Snow is white.’ (P. Horwich, *Meaning*, Ch. 1 & Ch. 3) Horwich also invokes the notion of a proposition; the primary bearer of his deflationary theory of truth is a proposition. Given that a proposition is the meaning of an indicative sentence, SNOW IS WHITE should also be considered the proposition referred to by ‘that snow is white’ embedded in an ordinary meaning-attribution of the form ‘x means that snow is white.’ Indeed, Horwich says that the concept, DOG, is an ingredient of the proposition that dogs bark. (P. Horwich, *Meaning*, p. 4) So, I will consider the concept expressed by an indicative sentence the proposition referred to by a ‘that’-clause involving the sentence. Horwich says, however, that a proposition is the meaning of a sentence-token, and the use theory of meaning is primarily applied to the meaning of a sentence-type. See P. Horwich, *Meaning*, p. 82.

by invoking abstract meaning-entities, Field analyzes it by invoking a non-abstract, linguistic entity. Despite this appearance, however, their analyses of meaning-attributions are not fundamentally different from each other. To the least, there are very important commonalities underneath the superficial difference. In the debate over deflationism, whether one invokes a meaning-entity is less important than how one characterizes the nature of a meaning-entity. As I will argue in the following, Horwich characterizes the nature of a meaning-entity in such a way that his non-revisionist analysis is compatible with Field's revisionist analysis of meaning-attributions. To argue for this point, I shall first explain why Field objects to invoking a meaning-entity in an analysis of a meaning-attribution.

There are three reasons for which Field is opposed to invoking a meaning entity in the analysis of a meaning-attribution. First, there is an ontological (or metaphysical) reason: being a physicalist, Field is declined to invoke abstract entities such as propositions or concepts. This is, however, the least important reason among three.²⁷⁵ The other two, much more important, reasons are that, first, invoking a meaning-entity tends to create a conflict with the deflationary conception of truth; and second, it isn't a meaning-entity but a linguistic entity that plays an explanatory role in an account of meaning. In the following, I shall first elaborate these two reasons, and then argue that if these are the reasons why Field objects to invoking meaning-entities, there is no irreconcilable gap between Field's and Horwich's deflationist views of meaning.

First, why does invoking a meaning-entity tend to create a conflict with the deflationary theory of truth? The deflationary theory of truth defends the idea that truth is a purely logical notion that serves as an expressive—not as an explanatory—device. Field thinks that this view of

²⁷⁵ Field says himself that the ontological reason is the least important reason. See H. Field, "Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*, p. 122.

truth will get into trouble if meaning-entities are invoked as the objects of the meaning-attributions. Consider a meaning-attribution of the form 2) illustrated above:

2) x means that snow is white.

To say that we construe a meaning-attribution of this form as a two-place relation between x and a meaning-entity is to say that we construe the ‘that’-clause embedded in 2)—‘that snow is white’—as a singular term referring to a meaning-entity.²⁷⁶ In philosophy, we call a meaning-entity referred to by a ‘that’-clause embedded in a meaning- or content-attribution a ‘proposition’.²⁷⁷ A ‘proposition’ is a term of art, and therefore, it isn’t defined uniformly—different theories of propositions define the nature of propositions in different ways.²⁷⁸ A proposition, however, is a meaning-entity. Therefore, it has the basic features that are attributed to a meaning-entity; that it is an abstract, mind- and language-independent entity. These are, however, not the features that make Field avoid a proposition. As was mentioned above, Field’s main reason for avoiding a proposition isn’t ontological. What makes Field avoid a proposition is that a proposition is typically defined as a *truth-theoretic* entity; as Schiffer says, a proposition is typically defined as an entity that has a truth condition essentially and absolutely.²⁷⁹ Now, it is

²⁷⁶ There seems no alternative here—there seems no other way but considering ‘that snow is white’ as referring to a meaning-entity as long as we construe a meaning-attribution of the form 2) as a relation to a meaning-entity.

²⁷⁷ In the literature of a meaning- and propositional-attitude attribution, the relational view is often taken to be equivalent to the propositional view. Namely, it is often taken for granted that if we consider ‘means’ and ‘believes’ as two-place relational verbs, then we should consider ‘that’-clauses following these verbs referring to propositions. (See M. McGrath, “Propositions,” *Stanford Encyclopedia*, F. Moltmann, “Propositional Attitudes Without Propositions,” *Synthese* 135, 2003, pp. 77-118.)

²⁷⁸ According to Schiffer, propositions can be construed either as structured entities or unstructured entities. Structured propositions are further divided into Fregean propositions and Russellian propositions, where both Fregean and Russellian propositions are further divided into neo-Fregean and neo-Russellian propositions. An example of an unstructured proposition is a pleonastic proposition; a product of the something-from-nothing transformation. For more details, see S. Schiffer, *The Things We Mean*, Oxford University Press, 2003.

²⁷⁹ A proposition has a truth condition absolutely because the proposition that snow is white has its truth condition without being relativized to anything (language, conceptual scheme, etc.). These basic features—including that a

clear why Field avoids a proposition in his analysis of a meaning-attribution—as long as a proposition is construed as an *intrinsically truth-theoretic entity*, if we analyze a meaning-attribution by invoking a proposition, we should assign a central role to a truth-theoretic notions (e.g. the notions of truth, reference, and a truth-condition). The central thesis of the deflationary theory of truth is that truth-theoretic notions don't play a heavy-duty explanatory role. Therefore, to invoke a proposition as an object of a meaning-relation may contradict the core idea of the deflationary theory of truth. To make the point vivid, suppose we analyze 2) (*x* means that snow is white) into 2^P) in the following:

2^P) Means (*x*, the proposition that snow is white).

Suppose further that a proposition is construed as an intrinsically truth-theoretic entity—that it is primarily defined in terms of the notion of a truth-condition. But then, by analyzing the meaning-attribution 2) in terms of 2^P), we would assign a central role to the notion of a truth-condition in an account of meaning. In this account, to say that a sentence *x* means that snow is white is to say that it stands in a meaning-relation to an entity that has the truth-condition that snow is white essentially and absolutely. The consequence is quite unpalatable to the deflationary theory of truth; the notion of truth is at the center of an account of the meaning of a sentence. Thus, Field's qualm about invoking a meaning-entity as the object of a meaning-relation is that once we invoke a meaning-entity as the object of a meaning-relation, then via the traditional theory of a proposition, we would end up assigning a serious explanatory role to truth-theoretic notions.

proposition is an abstract and mind- and language-independent entity—make a proposition essentially different from a sentence, a linguistic entity. See S. Schiffer, *The Things We Mean*, Oxford University Press, 2003, p. 14.

Field's argument is compelling, but it is open to one possible objection—a deflationist may invoke a proposition in his analysis of meaning-attributions if he defines a proposition in a non-traditional way. Horwich attempts this approach by characterizing a proposition as something other than an intrinsically truth-theoretic entity. Thus, Horwich suggest that we understand the notion of a proposition along the lines of (UP) in the following:

(UP) u expresses the proposition *that* $p \leftrightarrow \text{Int}(u) \in *p*$.²⁸⁰

Here, 'Int(u)' abbreviates 'the interpretation of u ', where to interpret u means nothing more or less than translating u into a sentence ' p ' in my language.²⁸¹ ' $*p*$ ' picks out the word type with a certain shape and a certain meaning, which is motivated in order to handle ambiguous expressions. For example, ' $*bank*$ ' picks out two different sets of that combination; one is composed of 'bank' and the use-property of a 'river bank', and the other composed of 'bank' and the use-property of a 'money-bank'. So, on this view, to say that an utterance u in another language expresses the proposition that p is to say that u is translated into ' p ' in my own language, where the meaning of ' p ' is in turn explicated in terms of $*p*$. Nowhere is a truth-theoretic notion such as the notion of reference or truth condition invoked. In this account, 2) is analyzed into 2^P), but 2^P) is analyzed into 2^M): x means SNOW IS WHITE. But SNOW IS WHITE—which amounts to the proposition that snow is white in Horwich's sense—is a *use-theoretic*, not truth-theoretic, entity. In the same breath, the concept, SNOW or the concept, WHITE is a use-theoretic entity which is primarily characterized in terms of a use-property. To be sure, from ' x means SNOW IS WHITE' we can infer ' x is true iff snow is white'; and from ' x

²⁸⁰ P. Horwich, *Meaning*, Oxford: Clarendon Press, 1999, p. 101.

²⁸¹ *Ibid.*

means SNOW' we can infer 'x stands for (or is true of) the property of being snow'. But this is the *consequence*, not constitutive, of 'x means SNOW IS WHITE' and 'x means SNOW'. So, Horwich's theory of propositions accommodates the ordinary conception of proposition within the framework of the deflationary theory of meaning. That is to say, it accommodates the commonsense idea that if a sentence expresses the proposition that *p*, then it has the truth condition that *p*. In doing so, however, Horwich doesn't assign an explanatory role to a truth-theoretic notion, because a proposition is primarily characterized in terms of a use-theoretic property, and only derivatively in terms of a truth-theoretic property. But then, Horwich's invocation of a meaning-entity—such as a proposition and concept—doesn't conflict with Field's linguistic view of a meaning attribution. Therefore, although Horwich and Field superficially disagree on the logical form and the object of a meaning-relation, they agree at the fundamental level—they both object to invoking an intrinsically truth-theoretic entity in an analysis of meaning-attributions. That is to say, they both deny that truth-theoretic notions play a central role in an account of meaning-attributions.

There is, however, another reason why Field objects to invoking a meaning-entity—a meaning-entity plays no explanatory role. To invoke a meaning-entity such as a proposition, says Field, "obscure(s) the fact that the explanatory role of propositions derives from that of the sentences that they 'shadow'."²⁸² Consider a meaning-attribution such as 3) in the following:

3) 'La neige est blanche' means that snow is white.

What explains our acceptance of 3) is not that we somehow know that 'La neige est blanche' stands in a meaning-relation to the proposition that snow is white. It is rather that we take the

²⁸² H. Field, "Attributions of Meaning and Content," in his *Truth and the Absence of Fact*, p. 167.

meaning-characteristics of ‘La neige est blanche’ are the same or sufficiently similar to those of ‘Snow is white’ in our language. Field defines ‘meaning-characteristics’ as the combination of the conceptual/inferential roles and indication relations of an expression. But then, what plays an explanatory role in an account of meaning-attributions is not a proposition; it is a linguistic entity. Therefore, from the perspective of an explanatory role, a proposition is nothing but a shadow of a sentence; whatever explanatory roles we may attribute to a proposition, they are derived from the explanatory roles played by the meaning-characteristics of a linguistic entity; i.e., a sentence.²⁸³

This argument is appealing; one piece of evidence is that if we don’t understand a sentence ‘*p*’ embedded in ‘the proposition that *p*’, invoking the proposition that *p* is of no use. For example, if someone tells me that a sentence *x* means the proposition that eukaryote is pelagic, I learn nothing about the meaning of *x*. On the other hand, if someone tells me that a sentence *x* means that a cell having a nucleus pertains to oceans, I can understand the meaning of *x*. Let’s suppose here that ‘Eukaryote is pelagic’ and ‘A cell having a nucleus pertains to oceans’ are synonymous; namely, they express one and the same proposition. If a proposition is more than the shadow of a sentence, there shouldn’t any difference between my understanding the proposition that eukaryote is pelagic and my understanding the proposition that a cell having a nucleus pertain to oceans. But there is a difference; I understand the one, but not the other. So, whatever properties a proposition has, they are derived from the properties of a sentence.

Field says that he isn’t strongly opposed to an intentional entity as long as an intentional entity is construed in a sufficiently bland way.²⁸⁴ What he means by this is that if a proposition is construed as nothing but a shadow of a sentence, and a concept as nothing but a shadow of a

²⁸³ H. Field, “Attributions of Meaning and Content,” in his *Truth and the Absence of Fact*.

²⁸⁴ *ibid.*, p. 166.

substantial expression, he isn't opposed to invoking a proposition or concept.²⁸⁵ But then, Field shouldn't be opposed to Horwich's invoking a proposition and concept as objects of meaning-relations. To see why let's first consider what the later Wittgenstein says of a proposition. He says,

'A proposition is a queer thing!' Here we have in germ the subliming of our whole account of logic. The tendency to assume a pure intermediary between the propositional *signs* and the facts. Or even to try to purify, to sublime, the signs themselves.—For our forms of expression prevent us in all sorts of ways from seeing that nothing out of the ordinary is involved, by sending us in pursuit of chimeras.²⁸⁶

The later Wittgenstein, the father of the use theory of meaning, invokes a proposition in his account of meaning. He objects, however, to the idea of construing a proposition as a mysterious entity that is distinct from a propositional *sign*; he says that our forms of expressions engender the misguided idea that a proposition is a sublime, extraordinary, chimerical entity. It isn't clear what he mean by 'our forms of expressions' here, but a good example of those expressions may be a 'that'-clause embedded in a meaning- and content-attributions. The surface structure of a meaning-attribution such as '*x* means that snow is white' makes us construe 'that snow is white' as a singular term referring an abstract entity such as a proposition. Given that we cannot change the surface structure of a meaning-attribution, it may be inevitable that we appeal to a proposition in an analysis of a meaning-attribution. That said, it isn't inevitable that we construe the nature of a proposition as entirely distinct from the nature of a propositional sign; i.e., a

²⁸⁵ *ibid.*

²⁸⁶ L. Wittgenstein, *Philosophical Investigations*, §94.

sentence. We may construe a proposition, as Schiffer says, as a pleonastic entity; an entity that emerges from our use of a linguistic entity.²⁸⁷

This pleonastic character of a proposition is also found in Horwich's notion of a proposition. Speaking generally, an advocate of the use theory of meaning should admit that the primary explanatory role in an account of meaning is played by a linguistic entity, and not by an abstract entity such as a concept or proposition. The core idea of the use theory of meaning is that a meaning-property is constituted by a use-property. A use-property, however, is a property possessed by a linguistic entity, not by an abstract entity. This shouldn't be taken as saying that an advocate of the use theory of meaning isn't allowed to invoke a proposition or a concept in an analysis of a meaning-attribution. Nevertheless, it is a linguistic entity, not an abstract entity, that plays a primary explanatory role in the use theory of meaning. Consider what Horwich says of a concept:

[G]eneralizing, our concept of *concept* will be explained by the schematic principle

The concept *F* = that which is meant (i.e. indicated) by our word "*f*" and by any word with the same meaning-constituting property.

Thus we cannot arrive at the underlying analysis of, for example, '*x* means DOG' in the way suggested by its superficial form—that is, by adding an analysis of the component, '*x* means *y*', to an analysis of the concept, DOG. Rather, we must first uncover the nature of meaning properties and then arrive (by 'subtraction' as it were) at the character of meaning entities (i.e. concepts).²⁸⁸

It isn't that a meaning-entity exists first, and then we derive a meaning-constituting property from the meaning-entity. The order of explanation is the opposite—we first uncover a meaning-

²⁸⁷ Schiffer calls this 'the something-from-nothing transformation'—the existence of a pleonastic entity depends on the way we speak a language. For example, a fictional entity such as Buck Mulligan is a pleonastic entity in that its existence is secured by the something-from-nothing transformation. This entity is created by the hypostatizing use of a fictional name. See S. Schiffer, *The Things We Mean*, Ch. 2 ("Pleonastic Propositions").

²⁸⁸ P. Horwich, *Meaning*, p. 26.

constituting property—a use-property—and then derive the nature of a meaning-entity from the nature of a use-property. A similar account should apply to the nature of a proposition; the order of explanation is that we first uncover the use-property constituting the meaning-property ‘*x* means that snow is white’, and then arrive (by subtraction) at the nature of the proposition that snow is white. But then, if we didn’t have the concept of a use-property of a linguistic entity, we couldn’t have the concept of a meaning-entity such as DOG or SNOW IS WHITE. So, conceptually speaking to the least, a meaning-entity such as a concept or a proposition is a shadow of a linguistic entity such as a word or a sentence. But then, Field has no reason to be strongly opposed to Horwich’s invoking a meaning-entity. Both in Field’s and Horwich’s accounts of meaning-attributions, the primary explanatory role is played by a linguistic entity, and not by an abstract, meaning-entity.

4.3.4 A Fact of The Matter

Horwich’s and Field’s deflationist theories of meaning seem to disagree more sharply in another aspect of their accounts—whether there is a fact of the matter in virtue of which an ordinary meaning attribution of the form 1) or 2) is true:

- 1) *x* means <dog>.
- 2) *x* means that snow is white.

As was mentioned, Field denies that there is a fact of the matter in virtue of which a meaning attribution of the form 1) or 2) is true; a meaning-attribution isn’t a matter of being correct or incorrect but a matter of being better or worse. Horwich, by contrast, claims that there *is* a fact of

the matter by virtue of which a meaning attribution of the form 1) or 2) is true; a meaning attribution has a truth condition, and therefore, it is a matter of being correct or incorrect. Viewed this way, the difference between Field's and Horwich's views of meaning in this respect is too big to be reconciled. When we give a closer look at their views of meaning, however, we can find room for reconciliation.

First, what is Field's argument for denying that there is a fact of the matter in virtue of which a meaning attribution is true? In Field's linguistic view of a meaning attribution, an ordinary meaning attribution of the form 2) above—

2) x means that snow is white.

is analyzed into 2^L) in the following:

2^L) x means the same as 'Snow is white'.

So, on Field's theory, to say that there is no fact of the matter for a meaning attribution of the form 2) is to say that there is no fact of the matter in virtue of which 2^L) is true. A statement of the form 2^L), however, states that the sentence replacing x and the sentence 'Snow is white' in a meaning-attributor's language mean the same; i.e., synonymous. Here, the sentence replacing x may be a sentence uttered by a speaker in my own language or in a foreign language. But then, 2^L) may be understood in terms of 2^{LP}) and 2^{LL}) in the following:

2^{LP}) x uttered by another English speaker means the same as 'Snow is white'.

2^{LL}) x uttered by a speaker in a foreign language L means the same as 'Snow is white'.

To say that there is a fact of the matter for a statement is to say that there is an objective standard for the truth of the statement. But then, to say that there is *no* fact of the matter by virtue of which a meaning statement of the form 2) is true is to say that there is no objective standard for interpersonal synonymy (2^{LP}) or interlinguistic synonymy (2^{LL}). In short, by denying that there is a fact of the matter for a meaning attribution, Field expresses his skepticism about interpersonal and interlinguistic synonymy.

But why is Field skeptical about interpersonal or interlinguistic synonymy? Field offers several related arguments all of which point toward one final conclusion. To see what it is, I shall first examine what he says about skepticism about synonymy:

The skeptic about interpersonal synonymy holds that though we of course have standards of translation, they are a matter of being better or worse, not of right and wrong, and also are highly context-dependent (since the purposes for which they are better or worse vary from one context to the next); and that these features together keep them from generating any useful notion of synonymy.²⁸⁹

Let's consider a claim about interlinguistic synonymy; namely, a claim involving a translation of an expression in a foreign language into an expression in our own language. In order to do so, we have to compare their meaning-characteristics; i.e., inferential/conceptual roles and indication relations. In this sense, Field says that we *have* standards of translation; for example, translating 'Comment ça va?' into 'You have just passed your French exam' grossly violates any reasonable standard of good translation, because the meaning-characteristics of these sentence-types are vastly different.²⁹⁰ But then, why does Field deny that there is an objective standard of translation? He offers two related reasons in the passage quoted above; the first is that one and

²⁸⁹ H. Field, "Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*, p. 129.

²⁹⁰ See H. Field, "Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*, p. 129.

the same expression in a foreign language can be translated into two different expressions in my own language, where none of them can reasonably be said to be “wrong”.²⁹¹ In such cases, there is no way of determining which translation is “the correct” one—both of them meet minimal objective standards, but beyond that it is a matter of personal taste or preference. Some may say that this is better, and others may say the other is better; but there is no objective standard that can determine who is right. In this sense, Field claims that the standards of translation are a matter of being better or worse, not of right or wrong. Another, related, reason is that translation or interpretation is context- and goal-relative—when we compare the meaning-characteristics of a foreign expression with those of our own expression, we often do so *relative to* a certain context and a goal (interest or purpose). But then, on Field’s view, we are to understand a synonymy-claim such as 2^{LL}) above along the line of 3) in the following:

- 3) x in a foreign language L is translated into ‘ p ’ in my own language relative to a context and a goal.

To say that translation (or interpretation) is often relativized to a context and goal, however, doesn’t necessarily show that there is no fact of the matter or objective standard for interlinguistic synonymy. We can still talk about “the best” translation within the framework of a

²⁹¹ This my elaboration, and the following is my example, too. Suppose English texts are translated into Asian languages; e.g. Koran. Translating English into Korean is very different from translating English into French or German in two respects. First, these are two entirely different kinds of languages (one belongs to the Indo-European family, and the other belongs to the Altaic). So, their grammars are vastly different; for example, the orders of their words are just opposite. Second, cultural backgrounds are vastly different as well; it is often impossible to find an expression that has exactly the same meaning. Translations in such cases should aim for finding expressions with sufficiently “similar” meaning-characteristics. For this reason, there are sometimes more than one plausible translations of the same expression. In such cases, it may be futile trying to determine the correct translation; we may have to accept that personal taste or preference affects the final decision. If this is Field’s point, I agree with him.

given context and goal. Being aware of this, Field gives a much stronger reason than that translation is relative to a context and goal. He says:

I'm inclined to think that the problem isn't just that interpersonal synonymy ('is a good translation of') is a highly vague notion, but that it isn't a fully factual notion at all; rather, it is an evaluative notion. This is compatible with many claims of interpersonal synonymy being factual relative to the goals of the translator; but not all interpersonal synonymy claims need be factual even relative to the goals.²⁹²

According to Field, meaning-related notions—such as 'synonymy' and 'is a good translation'—are not *straightforwardly factual*—they are *evaluative* notions. So, the main reason why Field denies that there is a fact of the matter as to a meaning-attribution is that meaning and synonymy are not straightforwardly factual notions—they are evaluative notions.

By maintaining that synonymy and meaning are evaluative notions, however, Field doesn't intend to defend "radical relativism"—the view that different people have radically different standards for a synonymy-claim and meaning-attribution. Field instead aims to defend "moderate relativism"—the view that different people may have *slightly* different standards for a synonymy-claim and meaning-attribution.²⁹³ Now, if this is Field's idea behind his 'no-fact-of-the-matter' thesis, there doesn't seem to be an unbridgeable gap between Field's and Horwich's views. To see why, I shall consider Quine's example for the indeterminacy of translation.²⁹⁴

Suppose a native speaker in a remote language utters 'Gavagai'. Suppose there are two different translation manuals both of which are acceptable according to all available

²⁹² H. Field, "Deflationist Views of Meaning and Content," Fn. 27.

²⁹³ Field doesn't explain clearly what he means by the claim that meaning and synonymy are evaluative notions. But he offers a detailed explanation of what he means by his claim that apriority is an evaluative notion (see H. Field, "Apriority as an Evaluative Notion," in his *Truth and the Absence of Fact*).

²⁹⁴ The discussion on Quine in the following may not be literally faithful to Quine's own presentation of his thesis of indeterminacy of translation. But my concern isn't so much Quine's thesis *per se* as its connection to Field's view of meaning.

observational evidence.²⁹⁵ Suppose further that we don't accept mentalism: the idea that the meaning of an expression is determined in the mind just as a label is attached to an exhibit in a museum.²⁹⁶ According to Quine, from these premises we can derive the conclusion that there is no fact of the matter by virtue of which a statement such as 4^T) in the following is correct.

4^T) 'Lo, gavagai' is translated into 'There is a rabbit'.

Quine's well-known reason involves a counterfactual consideration; imagine a counterfactual world in which people use a different translational manual, *T** according to which 'Lo, gavagai' is translated into 'There is an undetached rabbit-part':

4^{T*}) 'Lo, gavagai' is translated into 'There is an undetached rabbit-part'.²⁹⁷

According to Quine, the difference between 4^T) and 4^{T*}) wouldn't be manifested by observational evidence. The reason is that if we make a compensatory adjustment in the translation of more basic terms in native language, both 4^T) and 4^{T*}) would turn out to be equally

²⁹⁵ There is a controversy whether Quine's indeterminacy of translation and meaning is essentially grounded on Skinnerian behaviorism. Quine's own answer is that the 'observable evidence' isn't limited to assent or dissent (overt behavior). He says that it includes "any observable behaviour that one can exploit in trying to get a clue to how to translate the language." That's why I use here the phrase 'according to all available observational evidence'. W. V. Quine, "Comment on Hintikka," in R. B. Barrett & R. Gibson (eds.), *Perspectives on Quine*, Blackwell, Oxford, 1990. P. Raatikainen, "On How to Avoid Indeterminacy of Translation," *The Southern Journal of Philosophy*, 43, 2005.

²⁹⁶ W. O. Quine, "Ontological Relativity," in his *Ontological Relativity and Other Essays*, Columbia University Press, 1969.

acceptable according to all available observational evidence.²⁹⁸ To be more precise, if we construe ‘is’ in 4^{T*}) in terms of the notion of genidentity, it would compensate for the difference between ‘rabbit’ and ‘undetached rabbit-part’ thereby making 4^{T*}) as equally acceptable as 4^T).²⁹⁹ Therefore, the translation of ‘Lo, gavagai’ is indeterminate between 4^T) and 4^{T*}); there is no fact of the matter by virtue of which either 4^T) or 4^{T*}) is correct.³⁰⁰

Now, let’s consider Field’s and Horwich’s responses to Quine’s example in turn. First, Field basically agrees with Quine—we cannot say only one of 4^T) or 4^{T*}) is correct. Neither can we say that both of them are correct.³⁰¹ Rather, we should construe 4^T) as being inseparably correlated with a semantic standard involving the notion of identity, and 4^{T*}) with a semantic standard involving the notion of genidentity.³⁰² So, neither 4^T) nor 4^{T*}) is *straightforwardly* true; 4^T) is acceptable relative to the notion of identity, and 4^{T*}) is acceptable relative to the notion of

²⁹⁸ W. V. Quine, “Ontological Relativity,” in his *Ontological Relativity and Other Essays*, Columbia University Press, 1969.

²⁹⁹ Quine doesn’t specifically use the term ‘genidentity’, but it is obvious that what he means by a ‘compensatory adjustment’ is an adjustment of this kind; that whatever difference between translation the native’s sentence into ‘There is a rabbit’ and ‘There is an undetached rabbit-part’ is compensated for by a difference at a more deeper level (by correlating ‘undetached rabbit part’ with genidentity). Field calls this ‘correlative indeterminacy’ (H. Field, “A Reply to Anil Gupta and Jose Martinez-Fernandez,” *Philosophical Studies*, 124, 2005, pp. 105-128). The notion of ‘genidentity’ is introduced by Kurt Lewin, and it means here the identity-relation between two states of the same object. *Blackwell Reference Online*.

³⁰⁰ Quine says that the indeterminacy of translation of this kind cuts across both extension and intension—both meaning and reference are rendered indeterminate or inscrutable. W. V. Quine, *ibid*.

³⁰¹ Field proposed the partial-reference solution to Quine’s indeterminacy (inscrutability) of *reference*. It isn’t clear, however, this solution can be extended to solve the problem of indeterminacy of meaning. Even if it can, it doesn’t support the conclusion that both 4^T) and 4^{T*}) are correct. Instead, it may support that ‘Lo, Gavagai’ partially means that there is a rabbit, and it partially means that there is an undetached rabbit-part. See H. Field, “Quine and the Correspondence Theory,” in his *Truth and the Absence of Fact*.

³⁰² Field says,

[T]he basic indeterminacy is in the identity predicate, and that (putting aside ordinary vagueness) predicates like ‘rabbit’ get a fixed extension relative to the interpretation of identity. (For example, its extension is the set of rabbits relative to identity and the set of time-slices of rabbits relative to genidentity.) (H. Field, “A Reply to Anil Gupta and Jose Martinez-Fernandez,” *Philosophical Studies*, 124, 2005, pp. 105-128.)

genidentity.³⁰³ This view is consistent with Field's moderate relativism—different people can have slightly different standards, and there is sometimes no way of telling who is right in an entirely objective sense. We prefer 4^T) to 4^{T*}), but we do so because 4^T) meets *our* semantic standard, which is different from saying that our semantic standard is “correct” in an entirely objective sense. This shouldn't be taken as saying that our choice of 4^T) is entirely arbitrary—that isn't supported by any objective evidence. As was stated earlier, Field's view isn't that any translation will do; an acceptable translation should preserve the meaning-characteristics. For example, translating 'Lo, gavagai' into 'There is an apple' isn't acceptable, because their meaning-characteristics are too dissimilar. Field doesn't think, however, that a translation and meaning-attribution can *fully* be explained by evidence pertaining to meaning-characteristics. Just as Quine says, we cannot choose one of meaning-attributions between 5) and 5*)

5) 'Lo, gavagai' means that there is a rabbit.

5*) 'Lo, gavagai' means that there is an undetached rabbit-part.

as *the* correct one because if we made an adjustment at a more fundamental level, both 5) and 5*) would be acceptable. 5) is acceptable relative to the background assumption that 'is' is construed as identity, whereas 5*) is acceptable relative to the background assumption that 'is' is construed as genidentity. Surely, we choose 5), but that doesn't show that 5) is correct. It only shows that meaning isn't a *straightforwardly factual* notion; it shows that meaning is an *evaluative* notion.

³⁰³ This is what Field says about nonfactualism about a moral sentence. Field says, “The factualist/nonfactualist debate about evaluatives, then, is the debate about whether evaluatives are straightforwardly true.” See H. Field, “Disquotationalism and Factually Defective Discourse,” in his *Truth and the Absence of Fact*, p. 436.

Unlike Field, Horwich responds negatively to Quine's thesis of indeterminacy of translation. In the section called 'Quelling Quine's Qualms' in his book *Meaning*, Horwich offers a detailed criticism of Quine's arguments for the thesis.³⁰⁴ I will here discuss only parts of Horwich's argument against Quine's indeterminacy of translation.

First, Horwich argues that even if there are nonequivalent multiple translations all of which are acceptable on observational grounds, from this meaning skepticism that Quine intends to establish doesn't follow.³⁰⁵ What Horwich means by 'meaning skepticism' is the idea that no word means anything. According to Horwich, in order for this radical view to follow from the indeterminacy of translation, the thesis (M) in the following has to be assumed:

(M) If a word has a meaning then an adequate translation of that word must have the same meaning.³⁰⁶

But the thesis (M) isn't true; it subscribes to the implausible verificationist view.³⁰⁷ Therefore, meaning-skepticism doesn't follow from Quine's thesis of indeterminacy of translation.

Horwich's view of (M) seems right—as Horwich says, (M) can easily be refuted by the plausible anti-verificationist view that being true (or correct) should be distinguished from being justified.³⁰⁸ That is to say, although multiple translations are all justified on empirical grounds, it

³⁰⁴ P. Horwich, *Meaning*, Oxford: Clarendon Press, 1999, Ch. 9. The following discussion on Horwich's criticism of Quine's thesis of indeterminacy of translation is all based on Ch. 9 of Horwich's *Meaning*.

³⁰⁵ As will be discussed shortly, Horwich raises an objection to the narrowness of Quine's notion of observational evidence. Here, for the argument's sake, that objection is put aside.

³⁰⁶ P. Horwich, *Meaning*, p. 199.

³⁰⁷ See P. Horwich, *Meaning*, Oxford: Clarendon Press, 1999, p. 199.

³⁰⁸ There is a debate, however, whether Quine's thesis of indeterminacy of translation rests upon verificationism. See P. Raatikainen's "Is Quine a Verificationist?" *The Southern Journal of Philosophy*, 41, 2003, pp. 399-409. As Raatikainen states, the dispute seems to be concerned with how the notion of verification is to be understood. If the notion of verification is construed broadly, the verification theory of meaning amounts to nothing but the empiricist theory of meaning. That is to say, by defending the verification theory of meaning, Quine might have defended the empiricist approach—as opposed to the mentalist approach—to an explication of meaning. If so, Quine's approach

is still possible that only of them is correct.³⁰⁹ Therefore, Quine's meaning skepticism doesn't follow from his thesis of indeterminacy of translation. Now, given that Horwich's criticism of Quine's meaning skepticism is correct, how can this be balanced with Field's moderate relativism? According to Field's moderate relativism, there is no fact of the matter as to the truth of a meaning attribution; a meaning attribution isn't a matter of being correct or incorrect, but a matter of being better or worse. But then, isn't Field's view of meaning committed to meaning skepticism?

The answer seem 'no'.³¹⁰ Field distinguishes two kinds of skepticism; skepticism being defined as the view that we should abstain from making all meaning-related claims, and skepticism being defined as the view that there is no fact of the matter by virtue of which a meaning attribution is true.³¹¹ Field says that skepticism of the first kind is idiotic—it is foolish to say that no word means anything or we are barred from making a meaning-related claim.³¹² Indeed, a main virtue of evaluativism about meaning is that it enables us to avoid skepticism of this radical type; it is intended to render our meaning talk legitimate. On the other hand,

to meaning isn't substantially different from Horwich's. Horwich seems to agree with this point. See P. Horwich, *Meaning*, Ch. 9.

³⁰⁹ Horwich says,

[O]ne might distinguish between determinate truth, which requires decidability on the basis of underlying facts, and truth simpliciter, which does not. . . . And in this way one would be able to hang on to meanings as those entities that are preserved in correct (not merely adequate, and not necessarily determinate). P. Horwich, *Meaning*, p. 199.

³¹⁰ Field doesn't discuss skepticism about meaning directly. But he discusses skepticism about epistemology; in particular, epistemological properties such as default reasonableness. Field's view of epistemological properties is similar to his view of semantic properties (such as meaning and synonymy)—he argues that epistemological properties don't reduce to either natural or non-natural properties. So, on Field's view, reasonableness doesn't consist in anything; it is not a factual property but an evaluative property. I extrapolated Field's view of epistemological properties to semantic properties. See H. Field, "Apriority as an Evaluative Notion," in his *Truth and the Absence of Fact*.

³¹¹ H. Field, "Apriority as an Evaluative Notion," in his *Truth and the Absence of Fact*.

³¹² H. Field, *ibid.*, p. 371.

evaluativism about meaning denies that there is a fact of the matter as to a meaning-attribution. For this reason one might think that Field's view of meaning embraces skepticism of the second type to the least. But Field also objects to this idea—he says that it is a totally “perverse” definition of skepticism.³¹³ It is perverse because unless one assumes the factualist view of meaning, one couldn't derive skepticism from the absence of a fact of the matter about a meaning attribution. From the perspective of an evaluativist, it is only the nature of meaning that it doesn't have a fact of the matter corresponding to a meaning-related statement. So, on Field's view, semantic evaluativism shouldn't be considered advocating skepticism; to the contrary, it should be considered a *solution* to get out of meaning skepticism.³¹⁴

I brought up Quine's thesis indeterminacy of translation in order to find a commonality between Field's and Horwich's views of a fact of the matter by examining their responses to Quine's thesis. One commonality found is that neither of them takes Quine's thesis to support meaning skepticism. Although this is an important commonality, it isn't enough to show that their views don't conflict with each other. The problem is that although they both reject deriving skeptical conclusion from Quine's thesis, they seem to do so on different grounds. Field's reason, as was just explained, is that meaning isn't a straightforwardly factual notion; it is an evaluative notion. On the other hand, Horwich's reason is that Quine's thesis doesn't show that there is *no*

³¹³ As I stated above, this isn't what Field states *explicitly*; I derived this view from Field's discussion of epistemological properties. But I am quite confident that this is what Field would have said about his evaluativism about meaning. See H. Field, *ibid.*

³¹⁴ Field says,

Indeed, a main virtue of evaluativism is that it removes the force of most skeptical arguments. Most skeptical arguments depend on assuming that reasonableness is a factual property of beliefs or of rules, and on the understandable resistance to stripping away the normative nature of reasonableness by identifying it with a natural property like reliability (for rules; or being arrived at by reliable rules, for beliefs)

fact of the matter as to a meaning attribution. Even if Quine's right in claiming that two non-equivalent translation manuals are justified on empirical grounds, from this we cannot infer that none of them is "correct".³¹⁵ So, one big difference between Horwich and Field seems that whereas Field takes meaning to be an evaluative notion, Horwich doesn't—he seems to take it to be a straightforwardly factual notion.

I doubt, however, this is correct; I will argue that the notion of meaning in Horwich's account isn't straightforwardly factual, either. A caveat is needed here; Horwich has never expressed his view on this issue explicitly, so what I will argue in the following is my interpretation of Horwich's view of meaning.

Horwich's view of a meaning attribution is subtler and more complicated than it appears to be. Its subtlety and complexity are well captured by what Gupta says of Horwich's use theory of meaning. Gupta says,

Horwich's theory is unlike any other use theory of meaning in the philosophical literature. Whereas the other use theories aim to explain meaning in terms of some aspect of use (e.g., inferential or conceptual role), Horwich aims to explain not meaning but *meaning properties*, properties such as "x means RED in English," "x means AND in Paul's language," etc. Horwich is not concerned to reduce meaning to use.³¹⁶

Although what Gupta says here is important in understanding Horwich's deflationist view of meaning, he doesn't elaborate what he means by the distinction—the distinction between saying that *meaning properties* are reduced to use-properties and saying that *meaning itself* is reduced to a use-property.³¹⁷ I will here attempt to provide an explanation of the distinction as follows.

³¹⁵ P. Horwich, *Meaning*, Ch. 9, "Quelling Quine's Qualms," pp. 199-200.

³¹⁶ A. Gupta, "Deflationism, the Problem of Representation, and Horwich's Use Theory of Meaning," *Philosophy and Phenomenological Research*, Vol. 67, 2003, pp. 654-666, p. 655.

³¹⁷ Horwich says:

The distinction has to do with the complexity involved in Horwich's reductionist theory of meaning properties. As was explained earlier Horwich denies that the use-property of an expression *strongly determines* (DETERMINES) its truth-theoretic properties; namely, he denies that we can derive 7) from 6) in the following:

6) A predicate x has the use-property $U(x)$.

7) x is true all and only of dogs.³¹⁸

Neither can we derive 9) from 8):

8) A sentence x has the use-property $U^1(x)$.

9) x has the truth-condition that snow is white.

On the other hand, Horwich claims that the use-property of an expression *weakly determines* (determines) its truth-theoretic properties—if two expressions are synonymous, they have the

However, as we saw in Chapter 2 (section 2) and Chapter 3 (response to Objection 4), the expectation that one will be able to read off (and hence to explain) which particular meaning is engendered by a given meaning-constituting property is misguided. It has its origin in the Constitution Fallacy—the assumption that constituent structure is preserved under reductive analysis. (P. Horwich, *Meaning*, p. 220.)

Horwich makes the same point over and over again throughout his book *Meaning*. But he doesn't say explicitly that he distinguishes the notion of meaning from the notion of meaning-constituting properties (meaning properties). I think he should have had—that would have made it much easier to understand his deflationist theory of meaning. In a sense, I can understand why he didn't do so—one of his deflationist theses of meaning is that the notion of meaning, when it is properly understood by his deflationary account, is no different from the notion of meaning-properties. That is to say, Horwich's point is that we'd better understand ' x means <dog>', in terms of ' x means DOG'. Nevertheless, to say this is different from saying that ' x means <dog>' is reduced to ' x means DOG'. To say the latter isn't consistent with his claim that the meaning-properties of a predicate don't strongly determine its extension or the claim that we cannot read off which particular meaning is engendered by a given meaning-constituting property. This is the reason why I say that Horwich's deflationist theory of meaning has subtleties.

³¹⁸ As was just pointed out, Horwich denies that we can read off which particular meaning is engendered by a given meaning-constituting property. A meaning-constituting property is a use-property. Therefore, we are not supposed to read off 7) from 6).

same truth-theoretic property.³¹⁹ In Horwich's view, two expressions are synonymous if and only if they have the same use-properties. So, on this view, if two names have the same use-property, they are co-referential; if two predicates have the same use-property, they are co-extensional; and if two sentences are associated the same use-properties, they have the same truth condition. But these truth-theoretic properties are not themselves reduced to the use-properties of expressions. For example, the property of being true of all and only dogs in 7) isn't reducible to $U(x)$ in 6); the property of having the truth condition that snow is white in 9) isn't identified with $U^1(x)$ in 8).

Now, consider conditionals of the forms 10) and 11) in the following:

10) x means $\langle \text{dog} \rangle \rightarrow x$ is true of all and only dogs.

11) x means that snow is white $\rightarrow x$ has the truth condition that snow is white.

Both 10) and 11) are conceptual truths; they are trivially true in the sense that anyone who has mastered English would accept 10) and 11).³²⁰ On Horwich's account, however, the antecedents of 10) and 11) are not reducible to use-properties. For if ' x means $\langle \text{dog} \rangle$ ' or ' x means that snow is white' were reducible to use-properties, then we should be able to read off truth-theoretic properties such as ' x is true of all and only dogs' or ' x has the truth condition that snow is white' from the use-properties of x alone. Given that we accept 10) and 11) as conceptual truths, if the

³¹⁹ This is one of the subtleties that make it hard to understand Horwich's view of meaning at first glance. It is like saying that although we cannot read off which particular disease someone has from his symptom, if two people have exactly the same symptom, they must have the same disease. This statement makes perfect sense, but the question is why; why can't we read off a disease from a set of symptoms? Is it because of our epistemic limitations or because we are not supposed to read off meaning from meaning-constituting properties in the first place? The answer is the latter—it is, so to speak, committing a *category mistake* to think that we can read off the meaning and truth-theoretic properties of an expression from its use-properties. To think so is to assume that truth is a naturalistic property. It isn't, however, from a deflationary point of view. Therefore, it is committing a category mistake to expect that we should be able to read off the meaning and truth-theoretic properties of an expression from its use-properties.

³²⁰ Horwich accepts the triviality of these meaning-to-truth conditionals. See P. Horwich, *Meaning*, p. 7.

antecedents of 10) and 11) were reducible to use-properties such as 6) and 8) above, we should be able to read off the consequents of 10) and 11) from the use-properties 6) and 8). As was pointed out above, however, Horwich denies that we can do so—we cannot read off the truth-theoretic properties 7) and 9) from use-properties 6) and 8).

The conclusion that we can draw from this is that in Horwich's account ordinary meaning-attributions of the forms

x means $\langle \text{dog} \rangle$

x means that snow is white

are not reducible to use-theoretic statements of the forms 6) and 8) above—'A predicate x has the use-property $U(x)$ ' and 'A sentence x has the use-property $U^1(x)$.' Indeed, Horwich himself makes this point clear; he says,

And finally there is Kripke, whose argument *against* the naturalistic reduction of Meaning-facts is based implicitly on the Strong Relationality Constraint. He assumes that in order for some property of a word to constitute its meaning, it would have to be possible to, as he puts it, *read off* from that property exactly which meaning it constitutes. And this 'reading off' requirement is tantamount to the view that a reductive theory of meaning properties would have to have the form: x means $F = T(x, f)$.³²¹

Here, Horwich says that in order to read off the *meaning* of an expression from its use-properties, we need a principle such as ' x means $F = T(x, f)$.' To read off the meaning of an expression amounts to reading off its truth-theoretic properties from its use-properties. If we can read off the meaning of an expression from its use-properties, then, via meaning-to-truth conditionals such as 10) and 11) above, we should be able to read off the truth-theoretic properties of an expression

³²¹ P. Horwich, *Meaning*, p. 24.

from its use-properties. To do so, we need an inflationary principle such as ‘ x means $F = T(x, f)$.’ A theory of meaning, however, needs this principle only when it assumes an inflationary conception of truth. Since Horwich has already freed a theory of meaning from an inflationary conception of truth, his theory of meaning isn’t required to show how we can read off the meaning or truth-theoretic properties of an expression from its use-property. So, in Horwich’s theory of meaning, meaning-attributions of the forms ‘ x means $\langle \text{dog} \rangle$ ’ and ‘ x means that snow is white’ are not reducible to facts about use-properties.

The reason why I don’t think the notion of meaning in Horwich’s account is straightforwardly factual has to do with this feature of his theory—that ordinary meaning-attributions (e.g. ‘ x means $\langle \text{dog} \rangle$ ’ and ‘ x means that snow is white’) are not reducible to facts about use-properties. As Gupta says above, Horwich’s use theory doesn’t aim to reduce *meaning* to use; it is only concerned with reducing *meaning-properties* such as ‘ x means DOG’ and ‘ x means SNOW IS WHITE’ to use-properties. The difference between meaning and meaning-properties is that whereas meaning is an intrinsically truth-theoretic property, meaning-properties are not. As is indicated by 10) and 11) above, we can derive, without an auxiliary hypothesis, ‘ x is true all and only of dogs’ and ‘ x has the truth condition that snow is white’ from ‘ x means $\langle \text{dog} \rangle$ ’ and ‘ x means that snow is white.’ We cannot derive, however, ‘ x is true all and only of dogs’ and ‘ x has the truth condition that snow is white’ directly from ‘ x means DOG’ and ‘ x means SNOW IS WHITE’. For this reason, the epistemological status of 12) and 13) seems to be different from that of 10) and 11):

12) x means DOG $\rightarrow x$ is true all and only of dogs.

13) x means SNOW IS WHITE $\rightarrow x$ is true iff snow is white.

Horwich says that his use theory of meaning involves a series of reductive analyses such as

- x means DOG = $U(x)$
- x means ELECTRON = $V(x)$
- . . . and so on.³²²

Given that meaning-properties are reduced to use-properties in the way above, if we could derive truth-theoretic properties directly from meaning-properties, we should be able to *read off* truth-theoretic properties from use-properties. As was discussed above, however, this is denied in Horwich's deflationary theory of meaning. Therefore, the consequents of 12) and 13) cannot be derived directly from their antecedents—we need some auxiliary hypothesis such as 14) and 15):

- 14) 'dog' is true all and only of dogs.
- 15) 'Snow is white' is true iff snow is white.

Horwich says that although we cannot read off the meaning or truth-theoretic property from a use-property, if two expressions have the same use-property, they should have the same meaning and truth-theoretic property. In Horwich's theory, the concept, DOG is by definition the meaning-property of the English word 'dog'; and the concept, SNOW IS WHITE is the meaning-property of the English sentence 'Snow is white.' 14) and 15) are trivially true in a deflationary theory of truth; they are instances of the disquotational schemas. So, unlike 10) and

³²² P. Horwich, *Meaning*, p. 68.

11), 12) and 13) can be made sense of only when they are supplemented with auxiliary hypotheses such as 14) and 15).

To summarize, in Horwich's account, meaning is an intrinsically truth-theoretic, but not straightforwardly factual notion—the meaning of an expression cannot be reduced to a use-property. A meaning-property, however, isn't an intrinsically truth-theoretic, but is straightforwardly factual notion—it is reduced to a use-property. Such a discrepancy between the ordinary notion of meaning and a meaning-property is also found in Field's theory of meaning. Although Field doesn't take meaning and synonymy to be straightforwardly factual notions, he takes a *meaning-characteristic* to be a straightforwardly factual notion. It is a straightforwardly factual matter whether an expression has a certain meaning-characteristic. It isn't, however, a straightforwardly factual matter which *meaning* we are to ascribe to an expression on the basis of its meaning-characteristics. Neither Horwich nor Field, however, considers the irreducibility of *meaning* an obstacle in the way of developing an adequate theory of meaning. Given that the ordinary notion of meaning is intrinsically truth-theoretic, the irreducibility of meaning to a factual notion would engender a problem only if truth-theoretic properties were to play important roles in an account of meaning and content. The whole point of deflationism about meaning and truth is to deny it. But then, the irreducibility of meaning to a straightforwardly factual property isn't a problem at all.

4.4 Field's Response to The Success Argument: A Different Version

In this section, I will discuss Field's response to the success argument against a deflationary theory of truth—the criticism that a deflationist cannot accommodate the role of

truth in an account of the success of behavior. I discussed Horwich's response to the success argument in Ch. 3. Field's response to the success argument, however, is a little different from Horwich's—whereas Horwich's response is focused on explaining the role of *truth* in a success-explanation, Field's is focused on explaining the role of *truth conditions*. To be more precise, Horwich responds to the success argument by showing that the role of truth in a success-explanation of the form 1) below can be explained within the framework of a deflationary theory:

- 1) *X* succeeds in achieving his desire because, among other things, he believes that *p*, and his belief is true.

The role of truth in 1) is simply that it enables us to infer 2):

- 2) *X* succeeds in achieving his desire because, among other things, he believes that *p*, and *p*.

In order to explain how 1) entails 2), we need to assume nothing more or nothing less than the equivalence schema (E) ($\langle P \rangle$ is true iff *P*). To be sure, Horwich's point isn't that the notion of truth is redundant in a success-explanation—it serves the generalization function.³²³ A deflationist recognizes this function, but she argues that the generalization function can also be fully accommodated by what is implied by the equivalence schema.

Field's concern with the success argument, in a nutshell, is whether a deflationist can explain the role of *truth conditions* in a similar way—that the truth conditions of an agent's belief play only an expressive or logical function in an account of the success of an agent's

³²³ We appeal to the notion of truth as a device of generalizations over beliefs when we cannot articulate an agent's belief. For example, we may say 'Warren Bennett succeeds in making money because most of his beliefs about the stock market are true.' Here, the notion of truth cannot be eliminated because we don't know what Warren Bennett believes.

behavior. If a deflationist can show this, she can also show that no inflationary theory of truth conditions is needed—the role of truth conditions in a success-explanation can fully be explained by assuming nothing more or nothing less than the disquotational schema (DS): ‘ p ’ is true iff p .

How can a deflationist show that truth conditions play only an expressive role in an account of the success of behavior? This is not an easy task, since in order to show this, a deflationist needs to replace an ordinary success-explanation of the form 2) with an explanation that doesn’t invoke ‘that’-clauses. As Field says, to say that an agent believes that p is to say that he is in a belief-state that has the truth condition that p . Suppose we explain the success of a pilot’s behavior in the way below:

- 3) A pilot manages to land a plane safely with regularity because she slows down her plane when she believes that the airspeed is too high, and her belief is true.

In the explanation 3), the truth condition of the pilot’s belief plays an important role, since it characterizes her belief by invoking a ‘that’-clause—i.e., *that* the airspeed is too high. To say that she believes that the airspeed is too high is to say that her belief is true if and only if the airspeed is too high. So, according to 3), the truth-conditional content of the pilot’s belief plays an important role in an account of the success of her behavior. The truth-conditional content of the pilot’s belief is its representational content—to say that the pilot’s belief has the truth condition that the airspeed is too high is to say that it represents the external condition that the airspeed is too high.³²⁴ Field’s concern is whether a deflationist can recognize the explanatory role of the truth-conditional or representational content of a belief. Field says,

³²⁴ Presumably, the truth-conditional content of a belief-state isn’t always equivalent to its representational content. In this example, though, it is.

A detailed explanation of how the pilot functions would involve the existence of some class *C* of internal representations, intuitively representing airspeed, such that when she believes a representation in *C* that represents too high an airspeed she slows the plane and when she believes one that represents too low a speed she speeds it up; and the representations in *C* that she believes tend to be true. The last part can of course be rephrased as a generalization: she tends to believe a representation in *C* that represents too low a speed when the plane is in fact too slow, etc. But is this restatement of the claim that she has true beliefs about airspeed ‘innocent’ from the deflationist viewpoint? What may well seem a problem is that it is put in terms of what the pilot's internal representations ‘represent’.³²⁵

To make a long story short, Field's answer is that a deflationist cannot recognize the explanatory role of truth-conditional or representational content in a success-explanation. A deflationist can recognize an expressive or logical function served by the truth-conditional or representational content of a belief-state, but she cannot recognize a role that goes beyond that. But then, in order to show that the role of truth-conditional or representational content is exhausted by its expressive or logical function, a deflationist needs to replace a success-explanation such as 3) with an explanation that doesn't explicitly invoke truth-conditional or representational content. This in turn implies that a success-explanation such as 3) should be replaced with an explanation that doesn't invoke 'that'-clauses in the characterization of beliefs. Thus, Field suggests that we replace the success-explanation with one that goes as follows:

An obvious strategy for responding to this argument is to say that talk of representation is serving a merely heuristic role in the explanation of the pilot's ability; put without the heuristic, the explanation involves the existence of some class *C* of internal representations in the pilot and two subclasses *C*₁ and *C*₂ of *C* such that (i) when she believes a representation in *C*₁ she slows the plane and when she believes one in *C*₂ she speeds it up, (ii) there is a 1-1 function *f* from *C* to a certain set of real numbers such that (a) *C*₁ is that subclass of *C* that is mapped into numbers above a certain threshold and *C*₂ is that subclass of *C* that is mapped

³²⁵ H. Field, "Postscript to Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact*, pp. 153-154. Underlining was done by me.

into numbers below a certain (slightly lower) threshold, and (b) she tends to believe a representation r in C when the airspeed in knots is approximately $f(r)$.³²⁶

What Field suggests here is that we consider the truth-conditional or representational characterization of a belief as serving only a “heuristic role”—that is, we characterize a belief-state by invoking a ‘that’-clause (e.g. ‘believes that the airspeed is too high’) only when we don’t have detailed knowledge of an agent’s computational psychology. When we do, we replace the truth-conditional or representational characterization of a belief-state with a purely computational characterization of the pilot’s internal representations.³²⁷ A purely computational characterization of a belief invokes only the computational roles (causal-inferential roles) and indication relations of an agent’s internal representations. In order to explain the success of the pilot’s behavior only by invoking the computational roles and indication relations of the pilot’s internal representations, we first identify the class C of her internal representations in terms of their computational roles. The class C has two subclasses C_1 and C_2 which are defined as follows: when the pilot believes an internal representation belonging to the subclass C_1 , she slows down her plane, and when she believes an internal representation belonging to the subclass C_2 , she speeds up her plane. Second, we identify the class C of the pilot’s internal representations in terms of their indication relations. The internal representations belonging to the subclass C_1 indicate airspeeds above a certain threshold, and those belonging to the subclass C_2 indicate airspeeds below a certain threshold.

Once we identify the computational roles and indication relations of the pilot’s internal representations, we explain the success of her behavior by invoking the internal representations.

³²⁶ H. Field, “Postscript: Deflationism about Meaning and Content,” in *Truth and the Absence of Fact*, p.154.

³²⁷ It may be a little confusing to use the term ‘internal representation’ while Field adopts a negative attitude toward the role played by representational content. I thought about using the term ‘internal symbol’ instead of ‘internal representation’, but not all internal representations are symbols.

That is to say, the pilot succeeds in landing her plane safely because she slows down her plane when she believes an internal representation belonging to the subclass C_1 ; and she speeds up her plane when she believes an internal representation belonging to the subclass C_2 . Unlike the intentional (truth-theoretic) explanation 3) above, this explanation doesn't invoke a 'that'-clause in characterizing the pilot's belief-state. Instead of saying that the pilot succeeds because she believes that the airspeed is too high when she slows down her plane, we say that she succeeds because she believes an internal representation that belongs to the subclass C_1 . As a result, we avoid invoking the truth-conditional or representational content of the pilot's belief-state in an account of the success of her behavior.

Basically, what Field tries to show is that a deflationist can replace an intentional explanation with a non-intentional explanation just in the same way Horwich replaces an explanation of the form 1) with an explanation of the form 2). By doing so, Horwich shows that the role played by truth in a success-explanation can be accommodated by a deflationary theory of truth. In the same way, Field tries to show that the role played by a truth-condition in a success-explanation can be accommodated by a deflationary theory. Surely, we cannot always replace a success-explanation invoking the notions of truth and truth-conditions with the one that doesn't. When we don't have detailed knowledge of an agent's computational psychology, we cannot replace an intentional explanation such as 3) with a non-intentional explanation. Even when we have detailed knowledge of an agent's computational psychology, we still need an intentional explanation because of its convenience—it is simply too inconvenient to give a computational explanation of an agent's behavior in everyday life. So, we cannot eliminate an intentional explanation—we cannot entirely avoid truth-conditional characterizations of belief or desire-states. When we invoke them, however, we use them as a merely heuristic device. Just as

we invoke the notion of truth when we cannot or don't want to articulate a proposition or sentence, we invoke truth-conditions when we cannot or don't want to give a computational explanation of an agent's behavior.³²⁸

So, Field's deflationist view of truth conditions doesn't advocate an eliminativist view of truth conditions or truth-conditional content. It only advocates a deflationist or minimalist view of the role of truth-conditional content. In his review of Field's *Truth and the Absence of Fact*, Loewer says:

The most serious challenges for the deflationist program rather come from issues concerning explanation. Can psychological explanations be understood or reconstructed or replaced without loss by explanations in which the only semantic notions that figure are D-notions? . . . it is not at all obvious that there would be a way of picking out the relevant class of computational roles without appealing to R-reference and R-truth conditions.³²⁹

Loewer is right here—we need to invoke intentional (truth-conditional) content in order to pick out relevant class of computational roles and indication relations of internal representations. In the pilot example above, we appeal to the truth-conditional content of the pilot's belief in order to pick out the relevant class of internal representations—namely, we appeal to the 'that'-clause characterization 'The pilot believes that the airspeed is too high when she slows down her plane.' This doesn't show, however, that the truth-conditional content of the pilot's belief plays an "explanatory role" in an account of the success of her behavior. The success of her behavior is exhaustively explained by the computational roles and indication relations of her internal

³²⁸ In this sense, Field says that a truth-conditional characterization of a belief-state (i.e., a 'that'-clause characterization) loosely encodes the computational roles and indication relations of internal representations. Namely, we use a truth-conditional characterization of a belief-state in order to pick out a class of internal representations with certain computational roles and indicational relations. See H. Field, "Reply to Loewer," *Philosophical Studies* 124, 2005: 105-128.

³²⁹ Here, 'D-notions' means deflationary notions of truth and truth conditions', and 'R-notions' mean non-deflationary notions of truth and truth conditions. Underlining is done by me. See B. Loewer, "On Field's *Truth and the Absence of Fact*—Comment," *Philosophical Studies*, 124, 2005, pp. 59-70, p. 67.

representations. There is simply no room for truth-conditional content to play an additional explanatory role. The only way to argue for the explanatory role of truth-conditional content is to give a reductive or quasi-reductive account of truth-conditional content—an account that systematically connects the truth-conditional content of a belief- or desire-state with the computational roles and indication relations of internal representations.³³⁰ Field is skeptical about the prospect of such an account, which I think has to do with his conversion to deflationism in the first place. This is not the right place, however, to discuss this issue. Assuming that a reductive or quasi-reductive account of intentional content isn't likely to be forthcoming, Field's non-eliminative and deflationist view of truth conditional content seems to have plausibility. It recognizes the important role played by the concepts of intentionality (e.g., the concepts of truth-conditions, representational content, etc.). In doing so, however, it doesn't inflate these concepts—their roles are purely heuristic or expressive. So, we have arrived at an all-around deflationist view of truth-theoretic concepts—they are purely expressive devices.

³³⁰ Field explains at length why he doesn't think that the truth-conditional content of a belief- or desire-state can systematically be connected to the computational roles and indicational relations of internal representations. One main reason is that the truth-conditional content of a belief-state cannot be explained in terms of the indicational relations of internal representations. An internal representation can have several indicational relations, and so, in order to explain the truth-conditional content in terms of indicational relations, there should have a theory that explains why a certain indicational relation constitutes the truth condition of a belief-state. Field is skeptical about the prospect of such a theory. See H. Field, "The Deflationary Conception of Truth," in G. McDonald and C. Wright, eds., *Fact, Science, and Value*, Oxford: Blackwell, 1986; "Deflationist Views of Meaning and Content," in his *Truth and the Absence of Fact* (originally published in *Mind*, 1994).

Bibliography

- Austin, J. L. (1950). 'Truth'. In M. P. Lynch (ed.), *The Nature of Truth: Classic and Contemporary Perspective*. MIT Press.
- Ayer, A. J. (1946). *Language, Truth, and Logic*. 2nd Ed. London: Gollancz.
- Barrett, R. B. & Gibson, R. (eds.) (1990). *Perspective on Quine*. Oxford: Blackwell.
- Benacerraf, Paul. (1965). 'What Numbers Could Not Be'. *Philosophical Review* 74: 495-512.
- Block, Ned. (1980). 'Functionalism'. In N. Block (ed.), *Readings in the Philosophy of Psychology*. Cambridge: Harvard University Press.
- _____ (1996). 'What is Functionalism'. *The Encyclopedia of Philosophy Supplement*.
- Boghossian, Paul. (1989). 'The Rule-Following Considerations'. *Mind* 98: 507-549.
- Boyd, Richard. (1983). 'On the Current Status of Scientific Realism'. *Erkenntnis* 19: 45-90.
- Brandom, Robert (1998). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass: Harvard University Press
- _____ (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass: Harvard University Press.
- Buchanan, A. E. (2003). 'Are Truth and Reference Quasi-Disquotational?' *Philosophical Studies* 113: 43-75.

- Church, Alonzo. (1950). 'On Carnap's Analysis of Statements of Assertion and Belief'. *Analysis* 10: 97-99.
- Churchland, Paul. (1979). *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.
- Damnjanovic, Nic. (2003). 'Deflationism and the Success argument'. *The Philosophical Quarterly*, Vol. 55: 53-67.
- _____ (1997/2010). 'The Deflationary Theory of Truth'. *Stanford Encyclopedia of Philosophy*.
- David, Marian. (1994). *Correspondence and Disquotatation*. Oxford University Press.
- _____ (2002/2009). 'The Correspondence Theory of Truth.' *Stanford Encyclopedia of Philosophy*.
- Davidson, Donald. (1967). 'Truth and Meaning'. *Synthese* 17: 304-323.
- _____ (1984). 'On Saying That'. In D. Davidson, *Inquires into Truth and Interpretation*. Oxford University Press.
- _____ (1990). 'The Structure and the Content of Truth'. *The Journal of Philosophy* Vol. 87, No. 6: 279-328.
- _____ (1996). 'The Folly of Trying to Define Truth'. *The Journal of Philosophy*, Vol. 93, Issue 6: 263-278
- Davis, Martin and Stone, Tony. (eds.) (1995). 'Introduction'. In *Mental Simulation: Evaluation and Applications-Readings in Mind and Language*. Wiley-Blackwell.
- Dennett, Daniel. (1987). *The Intentional Stance*. Cambridge, Mass: MIT Press.

Devitt, Michael. (1984). *Realism and Truth*. Princeton, NJ: Princeton University Press.

_____ (1991). 'Deflationist Truth: A Critical Notice of Paul Horwich's *Truth*'. *Mind and Language*, 6: 273-283.

_____ (1991). 'Review: Naturalistic Explanation'. *The British Journal for the Philosophy of Science*: 425-443.

_____ (1997). 'On Determining Reference'. In Alex Burri (ed.) *Sprache und Denken/Language and Thought*, New York: Walter de Gruyter: 112-121.

_____ (2001). 'The Metaphysics of Truth'. In Michael Lynch (ed.) *The Nature of Truth*, Cambridge, MA: MIT Press: 579-611.

_____ (2003). 'Theories of Reference' In *Encyclopedia of Cognitive Science*, Vol 3.

L. Nagel, ed. London: Nature Publishing Group: 904-07.

_____ (2010). 'Still Against Direct Reference'. In R. Schantz (ed.) *Prospects for Meaning*. Walter de Gruyter.

Dodd, Julian. (2000). *An Identity Theory of Truth*. McMillan Press Ltd.

Donnellan, Keith. (1966). 'Reference and Definite Descriptions'. *The Philosophical Review* 75: 281-304.

Dretske, Fred. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

Dummett, Michael. (1959). 'Truth'. *Proceedings of the Aristotelian Society* 59 (1): 141-162.

_____ (1973). *Frege: Philosophy of Language*. Duckworth.

_____ (1982). 'Realism'. *Synthese* 52 (1): 145-165.

Field, Hartry (1977). 'Logic, Meaning and Conceptual Role', *Journal of Philosophy* 74: 379-409

_____ (1986). 'The Deflationary Conception of Truth'. In G. McDonald and C. Wright (eds.), *Fact, Science and Value*. Oxford: Blackwell, 55-117

_____ (1990). 'Narrow Aspects of Intentionality and the Information-Theoretic Approach to Content'. In Villanueva (ed.) , *Information, Semantics and Epistemology*. Oxford: Blackwell, 102-116.

_____ (1992). 'Critical Notice: Paul Horwich's *Truth*'. *Philosophy of Science* 59, 321-330.

_____ (1972/2001a). 'Tarski's Theory of Truth'. In Field (2001*i*).

_____ (1974/2001b). 'Quine and the Correspondence Theory'. In Field (2001*i*)

_____ (1978/2001c). 'Mental Representation'. In Field (2001*i*).

_____ (1986/2001d). 'Stalnaker on Intentionality'. In Field (2001*i*).

_____ (1994/2001e). 'Deflationist Views of Meaning and Content'. In Field (2001*i*).

_____ (1994/2001f). 'Disquotational Truth and Factually Defective Discourse'. In Field (2001*i*).

_____ (2000/2001g). 'Apriority as an Evaluative Notion'. In Field (2001*i*).

_____ (2001*h*). 'Apriority as an Evaluative Notion'. In Field (2001*i*).

_____ (2001*i*). *Truth and the Absence of Fact*. Oxford: Clarendon Press.

_____ (2005). 'A. Reply to Anil Gupta and Jose Martinez-Fernandez'. *Philosophical Studies* 124: 105-128.

_____ (2005). 'B. Reply to Loewer'. *Philosophical Studies* 124: 105-128.

Fine, Arthur. (1984). 'The Natural Ontological Attitude'. In R. Boyd, P. Gasper, & J.D. Trout (eds.) (1991), *The Philosophy of Science*. Cambridge, Mass: MIT Press.

_____ (1991). 'Piecemeal Realism'. *Philosophical Studies* 61: 79-96.

Fodor, Jerry. (1975). *The Language of Thought*. Cambridge, Mass: Harvard University Press.

_____ (1980). 'Methodological Solipsism'. *The Behavioral and Brain Science* 3: 63-109.

_____ (1987). *Psychosemantics: The Problems of Meaning in the Philosophy of Mind*. The MIT Press.

_____ (1990a). 'Semantics, Wisconsin Style'. In J. Fodor (1990d).

_____ (1990b). 'A Theory of Content I: The Problem'. In J. Fodor (1990d).

_____ (1990c). 'A Theory of Content II: The Theory'. In J. Fodor (1990d).

_____ (1990d). *A Theory of Content and Other Essays*. Cambridge, Mass: MIT Press.

_____ (1994). *The Elm and the Expert*. Cambridge, Mass: MIT Press.

_____ (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, Mass: MIT Press.

Fodor, Jerry & Pylyshyn, Zenon. (1988). 'Connectionism and Cognitive Architecture'. *Cognition* 28: 3-71.

- Frege, Gottlob. (1892). 'On Sense and Reference'. *Translations from the Philosophical Writings of Gottlob Frege*. P. Geach and M. Black (eds., and trans.), 3rd ed., 1980. Oxford:Blackwell
- _____ (1915). 'The Fundamental Laws of Arithmetic,' *The Monist* 25 (4): 481-494.
- _____ (1956). 'The Thought'. In P. F. Strawson (ed.) *Philosophical Logic*. Oxford university Press, 1973.
- Godfrey-Smith, Peter. (1989). 'Misinformation'. *Canadian Journal of Philosophy*, Vol.19, No. 4: 533-550.
- Gordon, Robert. (1986). 'Folk Psychology as Simulation'. *Mind and Language* 1: 158-171.
- Grover, Dorothy, Camp, Joe, and Belnap, Nuel (1975). 'A Prosentential Theory of Truth'. *Philosophical Studies* 27: 73-125.
- Gupta, Anil. (1993). 'A Critique of Deflationism'. *Philosophical Topics* 21: 51-83.
- _____ (2007). 'Deflationism, the Problem of Representation, and Horwich's Use Theory of Meaning'. *Philosophy and Phenomenological Research* Vol. 67: 654-666.
- Hacking, Ian. (1982). 'Experimentation and Scientific Realism'. *Philosophical Topics* 13: 71-87.
- Harman, Gilbert. (1973). *Thought*. Princeton: Princeton University Press.
- _____ (1982). 'Conceptual Role Semantics'. *Notre Dame Journal of Formal Logic* 23: 242-256.
- _____ (1999). 'Immanent versus Transcendent Theories of Meaning and Mind'. In G. Harman, *Reasoning, Meaning, and Mind*. Oxford University Press.

Hill, Christopher. (2002). *Thought and World: An Austere Portrayal of Truth, Reference and Semantic Correspondence*. Cambridge: Cambridge University Press.

Horwich, Paul (1990/1999). *Truth* 1st edition/2nd edition. Oxford: Blackwell

_____ (1994). 'The Essence of Expressivism'. *Analysis* Vol. 54. No.1: 19-20.

_____ (1999). *Meaning*. Oxford: Clarendon Press.

_____ (2001). 'A Defense of Minimalism'. *Synthese* 126: 149-165.

_____ (2004a). 'Three Forms of Realism'. In Horwich (2004e).

_____ (2004b). 'Realism and Truth'. In Horwich (2004e).

_____ (2004c). 'How to Choose between Empirically Indistinguishable Theories'. In Horwich (2004e).

_____ (2004d). 'Meaning, Use, and Truth'. In Horwich (2004e).

_____ (2004e) *From a Deflationary Point of View*. Oxford: Clarendon Press.

_____ (2010a). 'Semantics: What's Truth Got to Do with it'. In Horwich (2010b).

_____ (2010b). *Truth—Meaning—Reality*. Oxford University Press.

Jackson, Frank and Pettit, Philip. (1990). 'Program Explanation: A General Perspective'.
Analysis 50 (2): 107-117.

_____ 'Causation in the Philosophy of Mind'. *Philosophy and Phenomenological Research*
50: 195-214.

Jackson, Frank, Oppy, Graham, and Smith, Michael. (1994). 'Minimalism and Truth Aptness'.

Mind 103: 287-302.

Kim, Jaegwon. (1993a). 'Concept of Supervenience'. In Kim (1993).

_____ (1993b). 'The Myth of Nonreductive Materialism'. In Kim (1993).

_____ (1993c). 'Dretske on How Reasons Explain Behavior'. In Kim (1993).

_____ (1993d). 'Multiple Realization and the Metaphysics of Reduction'. In Kim (1993).

_____ (1993e). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge University Press.

_____ (2001). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press.

Kitcher, Philip. (1981). 'Explanatory Unification'. *Philosophy of Science* 48: 507-531

_____ (2001). 'Real Realism: The Galilean Strategy'. *Philosophical Review* 110 (2): 151-197.

_____ (2002). 'On the Explanatory Role of Correspondence Truth'. *Philosophy and Phenomenological Research* 64: 346-64.

Kripke, Saul. (1972). *Naming and Necessity*. Cambridge, Mass: Harvard University Press.

_____ (1982). *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.

Kunne, Wolfgang. (2002). 'Disquotationalist Conceptions of Truth'. In R. Schantz (ed.) *What is Truth?* Berlin" New York; de Gruyter.

_____ (2003). *Conceptions of Truth*. Oxford University Press.

Lauden, Larry. (1981). 'A Confutation of Convergent Realism'. *Philosophy of Science* 48:19-48.

Leeds, Stephen (1978). 'Theories of Reference and Truth', *Erkenntnis* 13: 111-129.

_____ (1979). 'Church's Translation Argument', *Canadian Journal of Philosophy* 81: 567-578.

_____ (1995). 'Truth, Correspondence, and Success', *Philosophical Studies* 79: 1-36.

_____ (2007). 'Correspondence Truth and Scientific Realism'. *Synthese* 159 (1): 1-21.

Leplin, Jarrett. (ed.) (1984). *Scientific Realism*. University of California Press.

Levin, Michael. (1984). 'What Kind of Explanation Truth?' In J. Leplin (ed.) *Scientific Realism*.
University of California Press: 124-139.

Lewis, David. (1970). 'General Semantics'. *Synthese* 22: 18-61.

_____ (1970). 'How to Define Theoretical Terms?' *The Journal of Philosophy* 67 (13): 427-446.

_____ (1975). 'Languages and Language'. In K. Gunderson (ed.), *Minnesota Studies of Philosophy of Science*. University of Minnesota Press.

_____ (1979). 'Attitudes de Dicto and de se'. *Philosophical Reviews* 88 (4): 513-543.

_____ (2001). 'Forget about the 'Correspondence truth''. *Analysis* Vol. 61. No.4: 275-280.

Loar, Brian. (1988). 'Social Content and Psychological Content'. In D. Rosenthal (ed.), (1991).
The Nature of Mind. Oxford University Press.

Loewer, Barry. (1987). 'From Information to Intentionality' *Synthese* Vol. 7 No. 2: 287-317.

- _____ (2005). 'On Field's *Truth and the Absence of Fact*—Comment'. *Philosophical Studies* 124: 59-70.
- Lynch, Michael. P. (ed.) (2001). *The Nature of Truth: Classic and Contemporary Perspective*. MIT Press.
- McGee, Vann. (1993). 'A Semantic Conception of Truth?' *Philosophical Topics* 21 (2): 83-111.
- McDonald, G & Wright, C. (eds.) (1986). *Fact, Science and Value*. Oxford: Blackwell.
- McDowell, John. (1978). 'Physicalism and Primitive Denotation: Field on Tarski'. *Erkenntnis* Vol. 13. No. 1: 131-152.
- McMullin, Ernan. (1984). 'A Case for Scientific Realism'. In J. Leplin (ed.) *Scientific Realism*. University of California Press.
- Miller, Alexander. (2000). 'Horwich, Meaning, and Kripke's Wittgenstein'. *The Philosophical Quarterly* 50 (199): 161-174.
- Millikan, Ruth. (1989). 'Biosemantics'. *Journal of Philosophy* 86: 281-297.
- Moore, G. E. (1899). 'The Nature of Judgment'. *Mind* 8. Reprinted in T. Baldwin (ed.), *Selected Writings* (1993). London & New York: Routledge.
- _____ (1901-2). 'Truth and Falsity'. In T. Baldwin (ed.), *Selected Writings*.
- Neale, Stephen. (1990). *Descriptions*. MIT Press.
- Neander, Karen. (1991). 'The Teleological Notion of a Function'. *Australasian Journal of Philosophy* 69: 454-168.

_____ (1995). 'Misrepresenting and Malfunctioning'. *Philosophical Studies* 79 (2): 109-141.

Papineau, David. (1984). "Representation and Explanation". *Philosophy of Science* 51: 550-572.

_____ (1987). *Reality and Representation*, Blackwell.

_____ (1998). 'Teleosemantics and Indeterminacy' *Australasian Journal of Philosophy* Vol. 76.
No.1:1-14.

Prior, Arthur. (1971). *Objects of Thought*, Oxford: Clarendon Press.

Putnam, Hilary. (1975). 'The Meaning of "Meaning"'. In K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, vol. 7.

_____ (1975). *Mind, Language, and Reality: Philosophical Papers* Vol. 2. Cambridge: Cambridge University Press.

_____ (1978). *Meaning and the Moral Sciences*. London: Routledge.

_____ (1985). 'A Comparison of Something with Something Else'. *New Literary History* 17: 61-79.

Pylyshyn, Zenon. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press.

Ramsey, Frank. (1927). 'Facts and Propositions'. In D. H. Mellor (ed.) (1990). *Philosophical Papers*. Cambridge University Press.

Raatikainen, P. (2003). 'Is Quine a Verificationist?' *The Southern Journal of Philosophy* 41: 399-409.

- _____ (2005). 'On How to Avoid Indeterminacy of Translation'. *The Southern Journal of Philosophy* 43: 395-413.
- Resnik, Michael D. (1990). 'Immanent Truth'. *Mind* 99 (395): 405-424.
- Russell, Bertrand. (1903). *The Principles of Mathematics Vol. 1*. Cambridge: University Press.
- _____ (1905). 'On Denoting'. *Mind* 14: 479-493.
- _____ (1911). 'Knowledge by Acquaintance and Knowledge by Description'. In his *Mysticism and Logic*. (1917). London: George Allen and Unwin, 152-167.
- _____ (1912). *The Problems of Philosophy*, London: Williams and Norgate; New York: Henry Holt and Company. Repr. New York and Oxford: Oxford University Press, 1997.
- _____ (1919). 'The Philosophy of Logical Atomism', Lecture 1-8. *The Monist*.
- _____ (1919). 'On Propositions: What They Are and How They Mean'. *Proceedings from Aristotelian Society, Supplementary Volumes, Vol. 2*: 1-43.
- Quine, Willard. (1953). *From a Logical Point of View*. Harvard University Press.
- _____ (1960). *Word and Object*. Cambridge, Mass: MIT Press.
- _____ (1969). 'Ontological Relativity'. In his *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- _____ (1970). *Philosophy of Logic*. Englewood Cliffs: Prentice-Hall.
- _____ (1981). *Theories and Things*. Cambridge: Harvard University Press.
- _____ (1990). *Pursuit of Truth*. Cambridge, Mass: Harvard University Press.

_____ (1990). 'Comment on Hintikka'. In R. B. Barrett & R. Gibson (eds.) *Perspectives on Quine*. Oxford: Blackwell.

Schantz, Richard. (2002). *What is Truth?* Berlin; New York; de Gruyter.

Shapiro, Stewart. (2003). 'The Guru, the Logician, and the Deflationist: Truth and Logical Consequence'. *Nous* 37(1): 113-132.

Sher, Gila. (1999). 'What is Tarski's Theory of Truth'. *Topoi* 18: 149-166.

Schiffer, Stephen. (1981). 'Truth and the Theory of Content'. In H. Parrett & J. Bouveresse (eds.), *Meaning and Understanding*. Berlin and New York: de Gruyter.

_____ (1987). *Remnants of Meaning*. Cambridge, Mass: Bradford Books.

_____ (2003). *The Things We Mean*. Oxford: Clarendon Press.

Soames, Scott. (1984). 'What is a Theory of Truth?' *The Journal Of Philosophy*, Vol. 80. No. 8. 1984: 411-429.

_____ (1998). *Understanding Truth*. Oxford University Press.

Stampe, D. W. (1977). 'Toward a Causal Theory of Linguistic Representation'. *Midwest Studies in Philosophy*, 2: 42-63.

Stich, Stephen. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. MIT Press.

_____ (1993). *Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. MIT Press.

_____ (1996). *Deconstructing the Mind*. New York: Oxford University Press.

Strawson, Peter. (1949). 'Truth'. *Analysis* Vol. 9, No. 6: 83-97.

_____ (1950). 'Truth'. *Aristotelian Society Supplementary Volume* 24.

Tarski, Alfred. (1944). 'The Semantic Conception of Truth and the Foundation of Semantics'.
Philosophy and Phenomenological Research 4: 341-376.

_____ (1956a). 'The Concept of Truth in Formalized Languages'. In *Logic, Semantics, and Metamathematics*. Oxford: Clarendon Press.

_____ (1956b). 'The Establishment of Scientific Semantics'. In *Logic, Semantics, and Metamathematics*. Oxford: Clarendon Press.

Van Fraassen, B.C. (1976). 'To Save the Phenomena'. *Journal of Philosophy* 73: 623-632.

Vision, Gerald. (1997). 'Why Correspondence Truth Will Not Go Away'. *Notre Dame Journal of Formal Logic* 38 (1): 104-131.

Williams, Michael. (1986). 'Do We (Epistemologists) Need a Theory of Truth?'. *Philosophical Topics* 14 (1): 223-242.

_____ (1999). 'Meaning and Deflationary Truth'. *Journal of Philosophy* 96: 545-564.

_____ (2001). 'Contextualism, Externalism, and Epistemic Standards'. *Philosophical Studies* 103 (1): 1-23.

Wittgenstein, Ludwig. (1922). *Tractatus Logico-Philosophicus*. Pears and McGuinness (trans.) side-by-side-by-side edition, version 0.24 (August 1, 2011).

_____ (1953). *Philosophical Investigations*. G. E. M. Anscombe and R. Rhees (eds.) trans. G. E. M. (trans.) Oxford: Blackwell.

Wright, Crispin. (1992). *Truth and Objectivity*. Cambridge, Mass: Harvard University Press.