

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

+

Solving Kripke/Wittgenstein's Rule-Following Paradox

By

Kai-Yuan Cheng

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

2002

UMI Number: 3063818

**Copyright 2002 by
Cheng, Kai-Yuan**

All rights reserved.

UMI[®]

UMI Microform 3063818

**Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.**

**ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346**

© 2002

Kai-Yuan Cheng

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.

6/11/02
Date

Jonathan Adler
Chair of Examining Committee *etc*

6/11/02
Date

Michael Deuth
Executive Officer

Professor Michael Levin

Professor Simon Blackburn

Professor David Rosenthal
Supervisory Committee

The City University of New York

Abstract

Solving Kripke/Wittgenstein's Rule-Following Paradox

By

Kai-Yuan Cheng

Advisor: Professor Michael Levin

The rule-following paradox of Kripke's Wittgenstein posits that there is no fact of the matter about an individual that can determine whether he means one thing or another by a term, such as "+". The paradox thus renders the existence of meaning illusory. The objective of this thesis is to examine the paradox and try to offer a version of a dispositional account that can counteract Kripke's skeptics.

Gaining insights from previous dispositionalist accounts of meaning and rule-following, including those of Dretske, Fodor, Millikan, Martin & Heil, Shogenji, and Yallowitz, my project is to put all those insights together to formulate a functional view of human dispositions to rule-following which will resolve the error, finitude, and first-person epistemic problems that Kripke raises against the dispositionalist approach. Specifically, I argue that rule-following or meaning consists in one's possessing a disposition, realistically construed. A rule-following disposition is real in the sense that it is responsible for, and hence not to be equated with, its manifestations under various conditions. A person acquires a real disposition to rule-following through a learning history which is constrained by biologically innate human endowments of learning and cognition. The error problem then becomes the problem of how to identify the disposition without vacuity or circularity. My claim is that identifying a rule-following

disposition is a task of scientific hypothesizing, which is epistemically and methodologically sound. The finitude problem is in my view a form of Humean inductive skepticism. My response to it is to adopt a reliabilist theory of justification and its treatment of the induction problem. A distinctive contribution of mine is to offer a functionalist account of dispositions and introspection in the resolution of the first-person epistemic problem. My aim is to do full justice to the phenomenology and epistemology of meaning and rule-following from the first-person point of view.

My conclusion is that a satisfactory dispositional account such as the one I offer not only vindicates the reality of rule-following and meaning but also increases our understanding of the nature of meaning and human rule-following phenomenon.

Acknowledgement

About a decade ago, I made a major switch from engineering to philosophy. A few years later I encountered Simon Blackburn's book *Think*, in which he views philosophy as a kind of engineering, namely, *conceptual engineering*. I realized then that the shift of my major perhaps is not as abrupt as it might seem, and that both engineering and philosophy are important in their own ways. I am happy that I have embraced a bit of both.

During the marathon journey of writing the dissertation, I am exceedingly lucky to have Professor Michael Levin to "run" with me. Professor Levin encouraged me to take on Kripke's rule-following paradox, and showed me where the core problem lies. As an advisor, Professor Levin has the wisdom, elegance, and patience to give me an abundance of free space to develop my own ideas (often slowly). But at the same time, he is always most critical, encouraging, and amazingly efficient in terms of giving me extremely helpful comments and advices. It is with his guidance and his genuine concern for students that a thesis generously described by Professor Simon Blackburn as "a genuine *tour de force*" could have taken the shape.

Professor David Rosenthal has exerted a very positive influence on me throughout my CUNY years. I was fortunate to have taken several excellent courses with him. He generously agreed to serve on all my major oral exams. His compassion for philosophical discussion and for helping a junior like me is extraordinary and truly admirable.

In the difficult process of writing my dissertation, Professor Simon Blackburn's involvement was a very important and pleasant part. His written comments on the final

draft of my thesis were so tremendously encouraging to me that I indeed aspire to be a better philosopher more than ever. I am deeply indebted to Professor Blackburn's kindness and generosity.

I am also very grateful to Professor Claudine Verheggen and Professor Jonathan Adler. They agreed to read the final draft of my thesis in a very short time and yet came up with very enlightening comments. Their generosity is greatly admired.

I have to thank many friends who have given me intellectual and emotional nourishments that are necessary for me to survive the graduate training. Robert Buckley kindly offered help on proofreading my thesis, which significantly improved my English. Rob is also my best companion in my CUNY life. I am also grateful to Rob's Mom and Grandma, who used to spoil me as one of their Italian boys in the family. Those experiences have become part of my sweetest memories. My dear friend Lara Kusnetzky kindly shared with me her love and support throughout the years. My artistic sensibility and sense of humor were greatly enhanced by her. Inna Ivanchenko added the most wonderful fragrance to my New York experiences. Inna helped me see that philosophy is only a small part of life. I also enjoyed numerous philosophical conversations with Jun-Yu Wu. He always made me feel free to talk about my immature ideas with him and in the meantime always refined and deepened my thoughts. My Taiwanese fellows at the Graduate Center, Pei-Yao Chen, Chia-Wen Fu, Pei-ting Hsu, Wei-Li Jao, Chih-Ping Ma, Dwan-Dwan Ou-Yang, Lisa Tsay, Chun-Ping Yen, Hong-Tzu Yhan, helped me one way or another. I am very grateful to them.

The completion of the dissertation is made possible by the generous financial support of the Chiang-Ching Kuo Foundation, from which I received a scholarship in

2000-2001. It allowed me to get rid of the library job and all the teaching responsibilities in New York, and to move back to Hsinchu, Taiwan, in order to concentrate on my writing. In Hsinchu, the professors of Institute of Philosophy of National Tsin-Hua University had offered me the opportunities to attend their colloquium talks and to give a talk on the rule-following paradox. Moreover, in the spring semester of 2001, when Professor Wan-Shan Zhan of the Institute was on his sabbatical leave to Germany, he kindly offered his office to me. I worked productively in Professor Zhan's office pleasantly surrounded by trees and birds. I could not thank Professor Zhan and all other faculty more.

I am also greatly indebted to Academia Sinica in Taiwan, which generously awarded me a dissertation fellowship in 2001-2002. I moved from Hsinchu to live in Taipei, and enjoyed the quiet and decent campus of Sinica. The Institute of European and American Studies of Academia Sinica sponsored my visit, and offered me a spacious and well-equipped office. I am very grateful to the dean of the Institute, Professor Cheng-Yi Lin, and to the scholars and staffs there for the most congenial and pleasant environment they provided. Among all colleagues there, I am especially indebted to three very fine philosophers, Professors Wang-Chuan Fang, Jih-Chin Ho, and Norman Ten. They generously gave me their time and ideas whenever I came to them with problems. I presented part of my thesis at the Institute, and would like to thank the audience for their very helpful feedback.

Among those that have contributed to my doctoral pursuit, Ruey-Yun Horng, a Professor of Psychology of the National Chiao-Tung University, has a very unique and continuous influence. Ruey was the one who encouraged me to change my major from

engineering to philosophy. Without her support, it would have been a much harder decision to make. I can discuss with her issues in psychology and philosophy any time, and she always showed her interest in ferment response. She read the first draft of my manuscript and gave me numerous invaluable suggestions. Her specialty in cognitive psychology helped me enrich the empirical studies consulted in my thesis. Under her influence, I even began to take beginners' Ballet lessons while I was in Hsinchu. Dancing on my toes with a pair of tight pants turned out to be the most refreshing experience to balance the complexity of Kripke/Wittgenstein's rule-following paradox.

Lastly, my greatest indebtedness goes to my family. My mother, grandparents, my brother, Ying-Chung and sister, Chun-Yun, all have given me unconditional support and love, and shared with me all my joys and pains along the way. My thesis, and philosophical pursuit, would mean nothing without them. This dissertation is dedicated to them, and to the memory of my father.

Table of Contents

Copyright	ii
Abstract	iv
Acknowledgements	vi
Table of Contents	x
1. Introduction	1
2. Kripke/Wittgenstein's Skeptical Paradox on Meaning and Rule-Following: Its Significance	9
3. The Anti-Realist Solution to the KW's Skeptical Paradox: A Very Unlikely Approach	29
A Communitarian Account	31
An Eliminativist Account	39
4. The Anti-Reductionist Solution to the Skeptical Paradox: An Unlikely Approach	55
McDowell's Wittgensteinian approach	57
Wright's Judgment-Dependence Account	75
5. The Dispositional Solution to the Skeptical Paradox: A More Likely Approach (I)	103
Shogenji's initial dispositional account	117
Martin & Heil's realist dispositional account	125
Yalowitz's expressivist dispositional account	140
6. The Dispositional Solution to the Skeptical paradox: A More Likely Approach (II)..	153
Millikan's teleological account	154
Fodor's asymmetric dependence account	164
Dretske's distinctive learning stages account	168
7. Proposing a Plausible Dispositional Account of Rule-Following (I): Solving the Error and the Finitude Problem	176
Old Plagues and New Lights in the Dispositional Account of Meaning and Rule-Following	176
A Counterfactual Analysis of Dispositions and the Error Problem in Martin & Heil's Realist View of Dispositions	182

A Solution to the Error Problem in Martin & Heil's Realist Dispositional Account of Rule-Following	193
Realism for a Disposition to Follow a Rule	215
Solving the Finitude Problem	236
8. Proposing a Plausible Dispositional Account of Rule-Following (II): Solving the First- Person Epistemic Problem	243
A Functionalist Account of Dispositions	246
First-Person Access to a Functionally Characterized Disposition of Rule-Following	250
Resolution of the First-Person Epistemic Problem and Explanation of Some Puzzling Phenomena of Rule-Following	260
9. Conclusion and the Future Directions	271
Bibliography	282

Chapter I. Introduction

When we mean something by a term, we may be certain that our applications of the term are correct in many cases, and we may truly believe that the correctness of those performances are guaranteed by the meaning of the term that we firmly grasp. For example, we may use the “+” sign to mean the *addition* function. We thus answer “4” when queried with the question “ $2 + 2 = ?$ ”. Moreover, we normally feel certain when giving the answer. When pressed to justify our answer, we are often able to cite the rule or meaning of the sign that we firmly grasp in the act of justifying our meaning or rule-following behavior. Kripke (1982), however, reconstructs from Wittgenstein’s *Philosophical Investigation* (1953) a powerful argument to the effect that meaning something by a term or rule-following is merely an illusion. According to Kripke, there is no fact of the matter that can determine whether or not we follow the plus rule or any other rules at all. This skeptical argument is widely known as Kripke/Wittgenstein’s (or “Kripkenstein”, to be abbreviated as “KW” in what follows) rule-following paradox.

KW’s rule-following paradox has attracted much attention, partly because KW’s skeptical argument is clearly and forcefully presented, and partly because the rule-following considerations touch some of the most central problems in contemporary analytical philosophy. For example, is meaning real? If yes, what kind of real entity or property is it? Is meaning reducible to other naturalistic properties such as causal, dispositional, or even physical properties? It seems that we know the meanings of our words in a direct, non-inferential, and authoritative manner. How can this special way of knowing be explained? Does meaning exist only in a communal context, rather than

belonging to an individual considered in isolation? Many philosophers have responded to KW's skeptical paradox, and tried to provide answers to the questions mentioned above.

If KW's meaning skepticism cannot be refuted, its consequence would be extremely counter-intuitive to our ordinary view of meaning and rule-following. It is also very destructive to our commonsense understanding of agency and rationality. The excitement, and maybe anxiety, of contemporary philosophers stirred up by KW's skepticism can be compared to Kant's reactions to skepticism about the existence of an external world. That is, if the skeptic cannot be refuted, the "meaning" we so firmly adhere to for our existence and pride may simply evaporate. Even though many philosophical positions have been proposed in the last two decades to resolve KW's skeptical paradox, no consensus has yet been reached as to which position or approach is likely to succeed.

The primary purpose of my dissertation is to argue that although all the positions or approaches to solve KW's skepticism that have been tried so far are not satisfactory, there is still room for a dispositional approach to account for human rule-following and meaning. In a dispositionalist view, an agent's following a rule or meaning something by a term is said to consist in his being disposed to behave in a certain way. For example, a person meaning plus by "+" consists in nothing but his being disposed to give the sum r when queried with " $m + n = ?$ " on numerous occasions, when $m + n = r$. This account, however, faces KW's three major objections. The first objection is known as the normativity problem. Meaning is normative, that is, meaning something by a term carries with it a condition under which only applying the term in a certain way counts as correct. On the other hand, however, a disposition does not seem to possess the element of

normativity. For according to the dispositional account, whatever manner one is disposed to apply a term is counted as correct. There will then be no distinction between a correct use and an incorrect use of a term. Errors in meaning and rule-following will hence be impossible. In KW's view, the dispositional view is simply "an equation of performance and correctness."¹ This normativity problem is also known as "the error problem."

The second objection KW's skeptic raises against the dispositional account of rule-following is based on the fact that meaning or a rule that one grasps or follows has an infinite number of implications. For example, when I use "+" to mean *plus*, the *plus* rule dictates infinitely many instances that my calculations ought to conform in order for me to qualify as a competent adder. However, KW argues that my disposition to perform addition calculations is only finite. I am capable of performing only finitely many calculations in my life-span. As a result, meaning/rule-following and a disposition cannot be the same, for one is infinite, and the other merely finite. This is known as "the finitude problem."

The third objection against the dispositional account of rule-following is derived from the intuitive idea that the meaning or rule that we grasp *guides* and *justifies* our applications of a term or rule-following behaviors. A machine might be able to add, by behaving in a way that exhibits a certain rule-conforming pattern. What distinguishes the machine and a competent adder like me is that when I add, I seem to be *directly* guided by the rule that I follow, and I am capable of citing the rule that I *consciously* grasp to justify my performances. A disposition, on the other hand, is something whose identity condition can only be known indirectly through inference from its behaviors manifested

¹ Kripke (1982), p. 24.

under various conditions. How then can a disposition be identified as the fact that constitutes meaning/rule-following? This is “the first-person epistemic problem.”

I will point out that though various dispositional accounts of rule-following have failed previously, there are genuine insights inherent in them that are worth further pursuing. I will propose a plausible dispositional account of rule-following that integrates all those insights and show that it can solve the error, finitude, and first-person epistemic problems. The major idea of my dispositional account is based on the presumption that the disposition to follow a rule or to mean something by a term is real. It is not reducible to a counterfactual analysis as KW thinks it is. All the three major problems about a dispositional account of rule-following are unfortunate derivations of KW’s treating a disposition in a counterfactual manner. I will show that by adopting a realist view of a disposition, all the three problems will be resolved. The result of my work, hopefully, would vindicate as well as help shed light on the reality of meaning and rule following.

The plan of my dissertation is as follows. In chapter II, I will begin by presenting KW’s sceptical argument as originally formulated by Kripke. After unfolding how the paradoxical conclusion about the non-existence of meaning/rule-following facts is derived, I will discuss its philosophical significance.

In response to the threats posed by KW’s skeptic, two major approaches to solve KW’s paradox have been proposed. One of them is based on acceptance of KW’s skeptical conclusion and then from there tries to propose a method to rescue human meaning-ascribing. In Chapter III, I will review two basic accounts that take this line of approach. The first account is a communitarian account, which is a “skeptical solution”

attributed to Wittgenstein by Kripke himself (1982, Chapter III). Its main suggestion is that we look at language and understanding in an anti-traditional way so that the disastrous consequence of the skeptical conclusion may be accommodated. The second account is an eliminativist account, which suggests that we simply accept the extremely implausible skeptical conclusion and try, though it is very hard, to live with it. I will argue that neither of these two positions is tenable.

The other approach to KW's skeptical paradox is to propose a positive account of rule-following. In Chapter IV, two anti-reductionist accounts will be reviewed. They are Wright's Judgment-Dependence Account and McDowell's Wittgensteinian Approach. Wright's account assimilates meaning and rule-following to secondary qualities, such as the color red. Its major appeal is that it accounts for the special way in which we come to know meanings and rules in our minds, i.e., by a way that is immediate, non-inferential, and authoritative. However, I will argue that Wright's account violates some of our most fundamental notions of meaning and rule-following and is hence implausible. On the other hand, McDowell views all constructive attempts toward solving KW's paradox as futile and instead seeks a Wittgensteinian approach to "dissolve" the problem. I will argue, however, that genuine philosophical puzzles about meaning and rule-following remain unresolved in his account. Besides, McDowell's account of meaning and rule-following is no less counter-intuitive than some rival theories that he criticizes as being too counter-intuitive to be true.

In Chapter V, I will argue that a dispositional account of rule-following is a more promising approach to KW's paradox. Three different versions of dispositional account of rule-following/meaning will be reviewed there. The first one is Shogenji's initial

dispositional account, which helps bring it sharply into focus that underlying the error problem is an epistemological difficulty that persists to plague a dispositional account of rule-following. The second one is Martin & Heil's realist dispositional account, which brings our attention to KW's holding certain misconceptions about the metaphysical nature of a disposition. They argue that the error as well as the finitude problems can be solved if a correct, realist view of dispositions is adopted. The third one is Yalowitz's expressivist dispositional account, which is directed at solving the first-person epistemic problem by showing how meanings or rules can be known directly by an agent who grasp them within a dispositional account of meaning. I will show that although each of them has shed some new light on the nature of human disposition to rule-following, each nevertheless suffers from certain specific difficulties in responding to the error problem, the finitude problem, and the first-person epistemic problem.

In Chapter VI, I will further review three major functional theories of the mind with specific attention to how they solve the error problem. The error problem continues to constitute a core problem for a causal/dispositional theorist because there does not seem to be a non-circular way of specifying the correct or incorrect behavior a disposition disposes a person to produce. Millikan's teleological account proposes that the biological function of an agent's cognitive mechanisms that are responsible for producing rule-following behaviors provide the correctness condition. Fodor's Asymmetric Dependent Account holds that it is possible to pinpoint one rather than another word/world dispositional correlation that determines the meaning of a term because the existence of the former is a pre-requisite for the existence of the latter. Dretske's Distinctive-Learning Stage Account states that what one means by a term is determined by what he learned

during the learning stage. I will argue that none of these attempts is successful in solving the error problem.

In Chapter VII, I will present my own version of dispositional account of rule-following which is based on the insights we gained from previous efforts. In brief, I will treat a human disposition as a real property, which is not to be reduced to or identified as a counterfactual as KW claims it is. The error problem will then become a problem of how to identify which real disposition to rule-following an agent possesses, rather than a problem of offering a criterion that can enable us to non-circularly distinguish between correct and incorrect rule-following behaviors. I will suggest that the task of identifying which real disposition an agent possesses be equated with an enterprise of hypothesizing, essentially like scientific theorizing. KW's skeptic would contend that hypothesizing about a rule-following disposition is hopelessly vacuous and will inevitably lead to circularity. I will claim that this charge is premature and ungrounded. I will argue that hypothesizing a disposition is both methodologically legitimate and epistemologically sound. I will show that the error problem that has been consistently threatening a dispositional account of meaning and rule-following can be solved with this reformulation of a human disposition. I will also point out that the finitude problem is based on a sort of Humean inductive skepticism, which can be coped with successfully by adopting a reliabilist view of justification.

In Chapter VIII, I will further explicate the first-person epistemic problem by showing how the meaning of a term or the rule that we follow can be introspected and provide us with guidance within the dispositional framework of meaning and rule-

following. The account I offer can also be applied to explain a Wittgensteinian puzzle about why a rule seems to be self-interpreting but is not fixed by interpretation.

In Chapter IX, I will conclude my study on the dispositional account of rule-following and present possible directions for further research into human rule-following.

Chapter II. Kripke/Wittgenstein's Sceptical Paradox and Its Significance

According to our common sense, we know what we mean by our words. Kripke (1982), however, reconstructs from the later Wittgenstein, especially from those passages that are concerned with the topic of rule-following--*Philosophical Investigations* (##138-242) and *Remarks on the Foundations of Mathematics* (section VI)--a skeptical argument that seriously challenges such a commonsense. The conclusion of the skeptical argument Kripke finds in Wittgenstein is that there is no fact about an individual that constitutes his meaning one thing rather than another by using a term, whether the term is a linguistic expression, a mathematical sign, or other. Put in another way, the skeptical conclusion says that meaning-ascribing sentences such as "I mean dog by 'dog'" or "John understands the '+' sign to mean plus" do not have truth conditions, and hence are neither true nor false. This is Kripke-Wittgenstein's (henceforth "Kripkenstein" or "KW") "skeptical paradox". My plan in this chapter is to lay out KW's skeptical argument, and discuss the extent of its philosophical significance.

The Skeptical Paradox

KW uses addition as the central example to present the skeptical paradox about meaning. Suppose that I am presented with an addition problem, say, " $68 + 57$ ", that I have never encountered before. I shall certainly answer "125", after some careful calculations. I shall also naturally suppose that "125" is the correct answer, the answer I must give if I want to be in accord with what I have all along meant by "+". KW, however, has me encountering a "bizarre skeptic" who challenges me: why it is "125", rather than "5"? I should answer "5" if I meant quus by "+" in the past, where quus refers to the following rule:

$$\begin{aligned}
 x \text{ quus } y &= x + y \text{ if } x, y < 57 ; \\
 &= 5 \quad \text{otherwise.}
 \end{aligned}$$

Supposing it is a fact that I meant plus by "+" in the past, I should be able to find such a fact, which constituted my having meant plus by "+" in the past, and justify now my answering "125" rather than "5". Note that there are two senses in which "125" is the correct answer to "68 + 57". First, there is the arithmetical sense in which "125" is correct, because $68 + 57 = 125$ is an arithmetic fact. Second, there is the metalinguistic sense in which "125" is correct, because the "+" sign does mean the *addition* function.

Explicated a bit more precisely, to say a sentence expressing a proposition with certain arithmetic symbols such as "+" is correct in the arithmetical sense means that the expressed proposition is an instance implied by some particular arithmetic function, such as *addition*. It is another distinct question, however, that concerns what arithmetical symbols such as "+" refer to. In a meta-linguistic sense, the correctness of the sentence depends on the reference of its linguistic or symbolic expressions. These two senses of correctness may come apart when the "+" sign means, say, the *subtraction* function and the correct answer to "68 + 57" then becomes "11". KW's skeptic does not doubt that there are truths of mathematical fact, nor does he challenge my present understanding of "+", that it means the mathematical function *plus*. Rather, KW's skeptic doubts whether I used the sign "+" to mean the plus function in the past, so that I should answer "125" rather than "5" now. The skeptical challenge is thus this: cite a fact that constituted my meaning plus by "+" in the past to justify my answering "125" now. Note that KW's skeptic makes an assumption in presenting his challenge, namely, that one ought to speak the truth. Only when this assumption is made will it be the case that I *ought to* or *should*

give a certain answer to " $68 + 57$ ".

KW's skeptic argues that there is no such meaning-constituting fact. The strategy of his argument is as follows. Meaning something by a word or understanding what one means by a word--let the word be a linguistic expression, a mathematical sign, or other--is presumably a matter about one's mental states. That is, there is something going on inside one's head that constitutes one's understanding. If there is such a fact that I meant plus by "+" in the past, then the most likely and suitable areas to search for this fact are features or contents of my previous mental history, linguistic or non-linguistic behavioral or dispositional facts about me, and so on. We are then allowed by KW's skeptic to have complete and omniscient access to all mental, behavioral, and dispositional facts about me in the search for the meaning-constituting fact. As a consequence, our knowledge-acquiring capacities are idealized with respect to the areas in question. The idea is that if there is such a fact, then given unlimited epistemic access, we must be able to find it. If the sought-after fact remains beyond our grasp under this epistemic condition, then the best conclusion to draw, according to KW's skeptic, would be that there is no such fact. It is worth noting here that the argument of KW's skeptic is ontological in character, rather than epistemological. KW's skeptic aims to show that there exists no things of a certain sort, i.e., meaning-facts, rather than that meaning-facts exist but we cannot know them. This is clearly revealed by KW's skeptic's presupposition that we are cognitively ideal, that we have access to all facts that are likely to constitute meaning. The correct way to view KW's skeptical argument, therefore, is that it is only a device for him to derive a metaphysical thesis about meaning.

KW considers several promising meaning-constituting candidates and rejects them in turn. His argument relies on formulating a normativity requirement of meaning and on claiming that none of the proposed candidates can meet this requirement. The formulation of the normativity requirement is derived from our most fundamental intuition about meaning, i.e., meaning is normative. The idea is that there is a matter of correctness and incorrectness regarding how I apply a term such as "+" that I understand, that is meaningful to me. As a result, when I mean "+" in a certain way, meaning it that way determines how I should use it in order for my usage to be correct. Boghossian expresses the idea as follows:

Suppose the expression "green" means *green*. It follows immediately that the expression "green" applies *correctly* to these things (the green ones) and not to those (the non-green ones). The fact that the expression means something implies, that is, a whole set of *normative* truths about my behavior with that expression: namely, that my use of it is correct in application to certain objects and not in application to others.²

What is being said is that if I mean something by a term, then on every occasion and at any time I apply that term there is always a matter of correctness and incorrectness concerning my application of the term. As a result, evaluative claims like the following ones are true of me: e.g., *it is correct that I answer "68 + 57 = 125"* (if I mean plus by "+") or *it is incorrect for me to say that "My blood is green"* (if I mean green by "green").

² Boghossian (1989a), p. 513.

Based on this idea of the normative nature of meaning, the normativity requirement can be formulated and appropriately imposed on any alleged account of meaning-constituting fact. Boghossian continues to write:

(B) The normativity of meaning turns out to be, in other words, simply a new name for the familiar fact that...meaningful expressions possess conditions of *correct use*. Kripke's insight was to realize that this observation may be converted into a condition of adequacy on theories of the determination of meaning: any proposed candidate for the property in virtue of which an expression has meaning, must be such as to ground the 'normativity' of meaning--*it ought to be possible to read off from any alleged meaning constituting property of a word, what is the correct use of that word...*(emphasis mine)³

With the normativity requirement (B) in mind, we can now see how each of the possible meaning-constituting facts discussed by KW founders on this requirement.

The Past Usage Theory

The past usage theory proposes that an agent's actual patterns of addition-problem solving behavior constituted or determined what he meant by "+" in the past. To see whether this theory is tenable, we can ask the question, based on the normativity requirement (B), Can we read off from my past usage of "+" that I was following the plus rather than the quus rule and hence ought to now answer "125" rather than "5"? The answer is, obviously, no. For one's actual calculations are finite while a meaning-constituting fact has an infinitary nature (We are here ignoring the fact that some of my past performances may even be erroneous. The issue about "errors" will be discussed in

³ Boghossian (1989a), p. 511.

more detail when we come to dispositional accounts). Our intuition is that the fact that constituted what I meant by "+" in the past determines how the term "+" is to be correctly used in an infinite number of cases, including the new case " $68 + 57$ " I have never encountered so far. On the other hand, however, the extent of my past usage of the term "+" is finite. We can then envisage that my past usage falls under the quus rule whose instances are the same as those of the plus rule but diverge from the plus rule at the point $\langle 57, 68 \rangle$. In other words, the pattern of my finitely many past usage may match not only implications of the plus rule but also those of the quus rule (and, actually, an infinite number of many other rules). Consequently, merely knowing that my past usage is such and such does not enable us to tell which rule I followed in the past and hence which answer is the correct one to give now. KW concludes that my past usage of a term cannot be what constituted what I meant by that term in the past or constitutes what I mean now.

This objection to the past usage theory is based on the tension between the finitude of the past usage and our intuitive notion of the infinite nature of a meaning-constituting fact. The implication is then that any plausible account has to accommodate the fact that a meaning-constituting fact possesses a characteristic of infinity of some sort.

The Counting Proposal

It is very natural and tempting, alternatively, to appeal to a set of instructions or algorithm to explicate what constitutes my following the plus rule. When we do addition, aren't we guided by a set of instructions which tells us what to do step by step to get a correct sum? For example, consider the counting proposal discussed by Kripke:

Take a huge bunch of marbles. First count out x marbles in one heap. Then count out y marbles in another. Put the two heaps together and count out the number of marbles in the union thus formed. The result is $x + y$.⁴

The "counting" proposal claims that a more basic rule--the "counting rule"--than the plus rule specifies a finite set of simple procedures which I explicitly followed in guiding my counting behavior in the past. These procedures will continue to guide me in every new case, and justify my answering "125" to " $68 + 57$ ".

To test the adequacy of this proposal, we again apply Boghossian's normativity requirement (B) as a heuristic device: Can we read off from my following basic "counting" procedures in the past that I meant plus by "+" and hence that " $68 + 57 = 125$ " is the correct answer I now ought to give? The result is that not much progress is made. The problem with this account is that it presupposes the meaning of "counting". By "counting" I could have meant *quounting* rather than *counting*, where "quounting" refers to the procedures by which I should give the sum "5" when one of the heaps counted exceeds 57. If I was really following the counting rather than the quounting rule, there must be some fact that determined it. What fact, however, Kripke's sceptic asks, can we appeal to such that my following the quounting rule in the past can be excluded? Maybe we can appeal either to more basic linguistic rules to interpret the counting rule or to past performances that were guided by the counting rule. The problem with the former is that the sceptic can come up again with non-standard interpretations for the terms involved in the more basic rules that we appeal to and hence a vicious regress would seem unavoidable. The problem with the latter is that my past usage of "counting" is finite.

⁴ Kripke (1982), p. 15.

and hence my past usage is compatible with my having meant quounting by "counting" as well as counting. In either case, the fact that determined that I meant counting by "counting" (and hence meant plus by "plus") remains undisclosed. The appeal to more basic rules or algorithms, therefore, does not give us an appropriate meaning-constituting fact.

This line of approach fails because its account is circular. Appealing to more basic rules would involve linguistic expressions such as "counting" whose meaning would again be open to various interpretations and remain unfixed. KW's skeptic turns this observation into a condition for any plausible solution, namely, any adequate account of the meaning-constituting fact must be non-circular.

Dispositional Accounts

Much of Kripke's discussion of the plausibility of various candidates focuses on the dispositional approach. A crude dispositional account of meaning is stated by Kripke as follows:

(D) To mean addition by '+' is to be disposed, when asked for any sum ' $x + y$ ' to give the sum of x and y as the answer (in particular, to say '125' when queried about ' $68 + 57$ '); to mean quus is to be disposed when queried about any arguments, to respond with their quum (in particular to answer '5' when queried about ' $68 + 57$ ').⁵

A dispositional approach appears to be more promising in meeting KW's skeptical challenge, for it attempts to characterize a disposition in a way that does not already presuppose what the function is that is meant by an adder when he operates with the "+" sign. According to a dispositional account, what function an agent means by "+" shall be

⁵ Kripke (1982), p. 23.

directly read off from how he is disposed to employ the sign "+". When presented with the question " $68 + 57 = ?$ ", he means plus by "+" if he is disposed to utter "125"; whereas he means quus by "+" if he is disposed to utter "5". It is a disposition that constitutes what one means by a sign. Although my past usage or a set of instructions may not distinguish my having meant plus or quus by "+", the dispositionalist contends that there was a dispositional fact about me that determined what I meant by "+" such that in the past if I had been queried about " $68 + 57$ ", I would have responded with "125". As a matter of fact, there are dispositional facts about me not only in the past, and present, but also in the future. The disposition I have now to say " $68 + 57 = 125$ " has been the same as the one that I had in the past, and it will also determine my response in the future. Moreover, in principle, I would be disposed to respond to any pair of numbers (x, y), no matter how large, with their sum z if I were queried about them under ideal conditions.

The dispositional account thus appears to avoid the drawback of the past usage theory, i.e., finiteness, because dispositions determine future and also counterfactual usage of a term. It also does not use semantic or intentional terms in its analysis so that the problem of circularity is avoided.⁶ How well does the dispositional account fare? Very badly, in Kripke's view. Though attempting to avoid the drawback of the past usage theory, the dispositional account also fails to meet the infinity condition. KW argues that my relevant dispositions with respect to calculating addition problems are presumably

⁶ Strictly speaking, to characterize a disposition as a disposition to *answer* or *respond* with a sum uses intentional notions because to answer with something is to assert or believe it. So a proper characterization of a disposition that Kripke's sceptic can accept has to be a disposition to produce merely verbal or physical outcome. But see Yallowitz's rejection of an outright reductivist stricture imposed on a dispositional account in his "A Dispositional Account of Self-Knowledge", *Philosophy and Phenomenological Research*, 2000.

finite, because all my capacities are finite. On the other hand, a rule can, intuitively, be instantiated by infinitely many cases. There are pairs of numbers so large that I will not and have not calculated them before I die. Who is to say what my semantic disposition is with respect to that pair? My semantic disposition can be a disposition to follow the plus rule, but it can equally be a disposition to follow the quus rule, because it is capable of only finitely many manifestations, which are compatible with the quus-rule-disposition as well as the plus-rule-disposition (actually, as well as infinitely many other dispositions). KW thus concludes that a dispositional account cannot determine unambiguously which rule one is following. We may call this problem that faces the dispositional account "the finitude problem".

Sophisticated dispositionalists may try to deal with the finitude problem by appeal to the *ceteris paribus* notions of dispositions. For example, a *ceteris paribus* clause may be fleshed out as follows: if my brain had been stuffed with sufficient extra matter and if my life were prolonged enough, then given an addition problem involving two large numbers (presently too large for me to grasp) I would respond with their sum. KW argues that two problems face this kind of attempt. First, we are not sure what would happen to me if my brain were stuffed with extra matter. Perhaps I would respond in a quus way or I might go insane. Failing to offer further specification of how to carry out the operation (which seems anyway so speculative that the task would better be left to fiction writers or futurologists), the result of the brain surgery would seem to be indeterminate. Second, and the most fundamental difficulty, is that it is hard to see how to specify the *ceteris paribus* clause without circularity. For any such attempt would have to specify the ideal condition in the following way: my brain is to be operated in

such a way that it will enable me to add rather than quoad. It is evident that this clause is circular, for it presupposes which function I meant by "+" in the past. We may equally conceive another ideal condition under which my brain is so modified that I will give quum rather than sum to arithmetic problems involving large numbers. Why should we favor one condition over another? It is hard not to agree with KW that we cannot favor one of them without circularity. The prospect for a sophisticated dispositional account thus seems to be dim.

KW argues that there is another major difficulty that faces a crude dispositional account. The failure is derived from the fact that people occasionally make mistakes. The dispositional account of meaning claims that one's disposition to answer addition problems constitutes what one means by a term. Thus, given the normativity requirement (B), it entails that what an agent meant by "+" is to be directly read off from his dispositions. Consequently, it follows that, when an agent gives a "deviant" answer, to say that the agent miscalculates presupposes that we have identified the function he meant, and is hence blatantly circular. The dispositional account, therefore, makes it impossible to claim that the outcome of a calculation is incorrect.

The problem that faces that dispositional account may be presented as a dilemma. Either the account is circular (when certain performances are pointed out as incorrect, the function being meant is presupposed), or it has to say that correctness and incorrectness cannot be applied to rule-following behaviors (because whatever I do can be made out to be right, to be in accord with any rule) and thus that errors are impossible. Either horn of the dilemma renders meaning impossible. We may call this difficulty of accommodating errors "the error problem". The challenge to solving the error problem is to provide a

sophisticated dispositional account which non-circularly specifies a normal or ideal condition under which a suitably selected set of dispositions captures the extension of an expression's meaning accurately. It is not clear at all, KW argues, how such a normal condition can be specified without circularity.

Even if a sophisticated dispositional account can non-circularly specify a normal condition under which a selective set of dispositions determines the correct extension of a term (and even if it also solves the finitude problem), Kripke seems to think that there is a crucial element of normativity that would be missing in such a dispositional account. The identification of this missing element of normativity is both important (for it concerns whether a satisfactory dispositional account can be found or not) and complicated (due to Kripke's obscure presentation and inadequate explication in the subsequent literature), so I will leave the task of locating the missing element to a later chapter (chapter V). It suffices at this point to note that no dispositional account can be satisfactory unless the error and finitude problems are solved.

The Quale Account

The quale account identifies the meaning-constituting fact with the qualitative character of an agent's experience when he is introspectively sure that he is indeed meaning something by his linguistic expression. A *sui generis* quale is an inner state known directly to an agent by introspection, such as a headache, a tickle, a mental image, and so on. A quale that constituted my having meant plus by "+" in the past tells me what I ought to do in a new case now; it carries information in my consciousness regarding how I should act in order to be in accord with the rule I am trying to follow. It thus seems intuitively plausible to identify the meaning-constituting fact with a quale.

Kripke regards the quale account as giving the most natural and fundamental response to the skeptical challenge, because it captures the basic requirement on facts about what I meant in the past: when I meant plus by "+", I performed a calculation unhesitatingly and with a feeling of certainty that my calculation was correct. The phenomenological character of an introspectible quale guided and justified my application of a term when I tried to put my understanding of the term into practice. The first-person epistemic dimension of meaning is what a dispositional account apparently fails to accommodate and what the quale account is directed at.

However, following Wittgenstein's discussions of the classical empiricist picture of meaning, i.e., an image associated with a word constitutes a word's meaning, KW argues that a quale is neither necessary nor sufficient for my meaning plus by "+". It is not necessary because it is doubtful that a special quale invariantly accompanies each of my calculations. Moreover, we can imagine that there may not be any qualitative experience at all when I perform a calculation. A quale is not sufficient either for my meaning plus by "+". Even if there is a particular quale invariantly accompanying each of my calculations, no internal qualitative impression could tell me in itself how a term is to be applied in future cases. For how can an impression make one rule salient out of the presumably infinite number of rules? My having a special quale when I took myself as meaning plus by "+" may be interpreted in a way that is compatible with my meaning quus by "+". We can imagine that a competent adder like me, while trying to mean plus by "+" but giving an incorrect answer "5" to "68 + 57" in some peculiar occasions, may have one qualitative inner state that is introspectively indistinguishable from another qualitative state that accompanies my correctly answering "125". Thus, the qualitative

experience of an introspective quale can justify answering "5" as well as "125" to "68 + 57." In short, a quale can guide me correctly, but it can also guide me incorrectly. Given that a quale is neither necessary nor sufficient for meaning something by a term, the quale account cannot supply an appropriate meaning-constituting fact.

An Anti-Reductionist Response

KW's skeptic further considers a response which suggests that my having meant plus by "+" in the past may be considered as a primitive state, a state of a unique kind of its own, which is not reducible to any other states. This move is a reaction against the reductionist assumption implicitly held by KW's skeptic. All the proposed candidates discussed earlier try to reduce meaning properties to naturalistic properties, such as past usage, dispositions, phenomenological characters of experiences, etc. They are found inadequate to constitute meaning on the ground that intentional or semantic properties are appealed to in the determination of meaning and hence are circular. There are also the infinity and error problems that prove difficult to be handled. But why should we hold, an anti-reductionist asks, that meaning is a state reducible to any other states? An anti-reductionist thus proposes that meaning is a primitive fact, a more *sui generis* inner state than a qualitative introspectable quale that we discussed earlier. Given such a primitive state, the state of one's meaning something by a term is irreducible to any other naturalistic properties.

KW's response to this anti-reductionist move is that it is too "desperate" and leaves meaning wholly "mysterious". For such a primitive state is not introspectible like a quale, which a possessor is supposedly aware of with a feeling of certainty whenever it occurs. Moreover, according to KW, such a state is supposedly a finite object existing in

one's mind. As a finite being, a primitive state would not contain a whole table of the infinitely many ordered triples implicated by the plus function. How then does it instruct me how to go in a future case, which is presumably not contained in my mind at the present? KW thus concludes that an anti-reductionist proposal cannot offer us a satisfactory response to the skeptical challenge.

So, there seems to be no fact that constituted what I meant by "+" in the past. All the facts of suitable sorts, such as the mental, dispositional, behavioral, etc., have been explored but found inadequate to ground what the notion of meaning requires-- normativity. The best conclusion to draw then, according to KW, is that there is no fact constitutive of what I meant by "+" in the past.

This skeptical argument can be repeatedly applied to any other individual and to any other linguistic expression and any other content bearer, on the presumption that there is nothing special about me and the "+" sign. Hence, there is no fact as to what anyone meant by any term in the past. Moreover, the state of one's understanding at the present will become past in the next moment. Consequently, KW's skeptical argument can then be applied again and show us that there is no fact constitutive of what one means by a term at present. By the same token, there will be no fact that constitutes what one will mean by a term in the future. Meaning something by a word is merely an illusion. This is KW's "skeptical conclusion", or "skeptical paradox".

Meaning Skepticism and Skepticism about Mental Content

KW's skeptical argument is primarily addressed to the meaning of a linguistic expression. However, it is not just linguistic terms that have meaning. My propositional attitudes, such as beliefs, desires, hopes, etc., can also possess content that is meaningful

to me, and understood by me. The content of my belief that it's rainy is that *it's rainy*; the content of my hope that the rain will stop soon is that *the rain will stop soon*. The question then arises whether or not KW's skeptical argument carries over to mental content. The question is, if KW's skeptical argument against meaning for linguistic terms were effective and showed that there is no such thing as meaning anything by any term, could we unproblematically continue to hold that there exist contentful psychological states? The answer to this question, as Boghossian⁷ points out, depends on how the relation between language and thought is construed.

Philosophers of language, for example, Sellars, Grice, Davidson, and others, have divided opinions over how to explain the relation between language and thought; in particular, whether language or linguistic meaning should be explained in terms of mental content, or vice versa? Three major approaches have been proposed that seem to exhaustively occupy possible positions regarding this issue. The Gricean approach explains the semantic in terms of the psychological, so that linguistic expressions derive their meaning from mental properties. On the other hand, according to the Sellarsian/Dummettian views the notion of linguistic meaning is primary, and the content of propositional attitudes is derivative and only to be explained in terms of the former. The Davidsonian approach argues that linguistic meaning and mental content either have to be fixed together, or not at all.⁸

⁷ Boghossian (1989a), especially, p. 509-14.

⁸ For the Gricean approach, see H.P. Grice, "Meaning", *Philosophical Review*, 1957; S. Schiffer, *Meaning*, 1972; For the Sellarsian and Dummettian view, see Sellars "Empiricism and the Philosophy of Mind", in his *Science, Perception, and Reality*, 1963; Dummett, *The Logical Basis of Metaphysics*, 1991, *The Seas of Language*, 1993; for the Davidsonian approach, see Davidson, *Inquiries into Truth and Interpretation*, 1984.

Our concern here is how propositional attitudes would fare if KW's skepticism about linguistic meaning were effective: if there is no fact of the matter that determines what one means by a term, could there be a fact of the matter that determines the content of one's propositional attitude? If one adopts the Sellarsian/Dummettian approach, then it follows immediately and obviously that the status of propositional attitudes is in serious doubt. For on this view, the content of a propositional attitude is explained in terms of linguistic meaning. If the latter is proved to be fictive, there is little to make sense of the former.

The destruction of the status of propositional attitudes is even more clear if one adopts the Davidsonian view. On this view, neither mental content nor linguistic meaning has explanatory priority. As a result, if we cannot explain one, we cannot explain the other either. On the Gricean view, however, the matter is not as straightforward. As Boghossian notes: "Since the Gricean holds that linguistic items acquire their meaning from the *antecedently* fixed content of mental states, an argument to the effect that nothing non-mental fixes linguistic meaning would leave the Gricean unmoved."⁹ A Gricean, whose view is that a linguistic expression possesses its meaning in virtue of a speaker's mental states, would simply respond to KW's skeptic that one means plus by "+" because the concept of *plus* (and not quus) figures in his thought or belief that, say, $12 + 15 = 27$. KW's skepticism about linguistic meaning would seem powerless and much less interesting if mental content remains intact under the threat of KW's skeptical argument.

⁹ Boghossian (1989a), p. 510.

However, a deeper explication of KW's normativity requirement will seriously undermine the claim that KW's skeptical argument would leave the Gricean view of language and meaning intact. The key to understanding KW's normativity requirement is that: it has little to do with the *bearers* of meaning and content--be they sentences of a natural language, mental states, neural structures in the brain, items in a language of thought, etc.--but everything to do with something's *possessing a content*. Although KW may not directly challenge the determinacy of the content of a mental state in his skeptical argument, he does implicitly adopt a strategy of doubting anything with content. This can be clearly seen in the case where he rejects the algorithm or set of instructions proposal. There KW considers the counting proposal in which one's meaning plus by "+" is constituted by one's being guided by a set of instructions that involves a term "counting", which is defined as such and so. Based on the normativity requirement, KW's skeptic asks: does knowing that one meant counting by "counting" and hence plus by "+" provide a sufficient basis for us to conclude that he should answer "125" rather than "5" to " $68 + 57 = ?$ " The answer is surely, "yes", if he meant counting by "counting", but "no", if he meant quounting by "counting" in the past. KW's skeptic would then press the point: what could determine that he meant *counting*, rather than *quounting*, by "counting" in the past? If there is such a past fact, then there must be something that determined it. Remember, we are allowed unlimited epistemic access to all facts about my physical, behavioral, and mental history (the list of mental facts can include qualitative experiences, sensations, representations, what have you, as far as their contents or interpretations, if any, are not taken for granted) in the search for the content-

determining fact. The same strategy can be applied to any other items, such as concepts, that are claimed to possess contents.

McGinn¹⁰ denies that KW's skeptical argument can be applied at the level of thought and concepts, but this denial is based on a fundamental misconstrual of KW's idea of normativity. Suppose that an agent takes himself to have employed the concept (the essential component of which is presumably a content) *plus* in his thought in the past, and, attempting to be in accord with this earlier employment of the concept *plus*, the agent *should* now entertain the thought that $68 + 57 = 125$. KW's skeptic, however, would press the agent with the following question: "Surely $68 + 57 = 125$ is a correct thought, if you are applying the same concept as in the past and do not want to change your intention of applying the same concept; however, if you applied the concept *plus* rather than *quus* in the past, then there must be a fact that determined it..." The skeptic's argument then continues as before and the skeptical problem recurs¹¹. By the same token, KW's normativity argument can be applied as well to other propositional attitudes such as beliefs, desires, hopes, etc., which possess concepts and hence contents. As a result, KW's skeptical argument applies to meaning in particular, and content in general.¹²

¹⁰ See McGinn's *Wittgenstein on Meaning*, 1984.

¹¹ See Boghossian (1989a), Wright (1989b), and Satorelli (1991) for convincing criticisms of McGinn and lucid discussions on KW's normativity requirement.

¹² KW's argument apparently relies on a reductionist assumption where only non-semantic and non-intentional properties are regarded as constituting meaning/content. This reductionist assumption is not universally accepted. Some philosophers have adopted a non-reductivist approach. A discussion of the plausibility of this approach has to wait until we get to Wright's judgment-dependent account of meaning and McDowell's Wittgensteinian approach in chapter IV. The purpose in this section is to see the nature of Kripke's normativity argument and the extent of its destructive power.

KW's skeptical argument thus leads not only to a paradox about linguistic meaning, but also to a paradox about mental content. There is no such thing as meaning something by a term, and no such thing as thinking, believing, hoping, desiring, and so on. Moreover, once the content of propositional attitudes is proved to be nonexistent, our ordinary notions such as rationality, cognition, agency, etc., have to be abandoned, for the content of propositional attitudes plays a central role in the explanation of our practical-syllogistic reasoning, intentional actions, and so on. The consequence of accepting KW's skeptical conclusion is indeed not only paradoxical, but also very destructive.

Chapter III. The Anti-Realist Solution to the Skeptical Paradox: A Very Unlikely Approach

How could the skeptical paradox be resolved? There are, in general, two ways in which KW's skeptical paradox may be responded to. One is to find a "straight" solution, i.e., to find fault with KW's skeptical argument and propose a meaning fact that meets the criteria set up by the skeptic. There are in turn two general ways in which a straight solution can be given, depending on how the reductionist assumption implicitly held in KW's skeptical argument is regarded. A dispositionalist views the reductionist assumption as legitimate, and therefore tries to solve KW's skeptical paradox by offering a non-circular specification of the condition of correct use of a term. An anti-reductionist finds KW's reliance on a reductionist methodology questionable, and thus argues that the derived skeptical conclusion is hardly convincing. In chapter IV, I will discuss two major anti-reductionist accounts and argue that they are unsatisfactory in responding to KW's skeptical paradox. I will discuss in chapters V and VI the dispositional approach which I find more promising.

Another way in which KW's argument can be responded to, other than a straight solution, is to accept KW's skeptical paradox, i.e., there is no fact of the matter as to whether I mean one thing rather than another by a term. On this approach, we would regard meaning ascriptions like "I mean plus by '+'" as not corresponding to any fact, thus as being neither true nor false. If we take a realist view about a given region of discourse as holding that "the world is furnished to provide states of affairs of the kind which such statements are apt to represent"¹³, then accepting KW's skeptical paradox amounts to

¹³ This characterization of realism is due to Wright, "Can There be a Rationally Compelling Argument for

rejecting the realist position about meaning and content. As we have seen, the resulting anti-realist position about linguistic meaning entails an anti-realist view about propositional attitudes, and this consequence is very paradoxical and destructive.

There are two ways in which an anti-realist position may be proposed. The first one is to offer a "skeptical solution", of the sort attributed to Wittgenstein by Kripke (in chapter III, 1982). KW's skeptical solution tries to offer an anti-traditional way to look at language and meaning such that we can live with the "insane and intolerable"¹⁴ skeptical conclusion. However, various attempts have been made, by Blackburn (1984), McGinn (1984), Hoffman, (1985), and others, to show that KW's skeptical solution cannot salvage our ordinary linguistic practice from the damage done by accepting the skeptical conclusion. Another way to deal with the paradoxical skeptical conclusion is simply to accept it, without trying to accommodate its extremely paradoxical and implausible implications. After all, KW's argument is solid and it may be impossible to get around it. However, this position is unstable, too. For the conclusion of the skeptical paradox may be incoherent and hence self-refuting. The debate with regard to such an anti-realist position, i.e., simply accepting the skeptical conclusion, has been centered on whether it is internally coherent or not (See Boghossian, 1989, 1990; Devitt, 1990; Soames, 1997, 1998; Wright, 1984, 1992; etc.) If the skeptical conclusion is not even coherent, then not only is it very likely to be false, but it *must* be false. Besides, if the skeptical conclusion is incoherent, then it would also show that KW's skeptical solution is definitely wrong, since KW's skeptical solution has the skeptical conclusion as one of its main components. In what follows, I will review and criticize both versions of the anti-realist position.

Anti-Realism about Ordinary ("Folk") Psychology?" 1995, p. 198.

¹⁴ Kripke (1982), p. 60.

KW's Skeptical Solution: A Communitarian Account

KW's skeptical solution consists of two parts: the first part accepts the skeptical conclusion that there exists no fact as to what one means by a term and thus that meaning ascriptions do not correspond to facts, and the second part nevertheless claims that meaning ascriptions are legitimate because there are assertability-conditions for ascriptions of meaning that are sometimes met. In other words, the skeptical solution's goal is to render intact our ordinary practice of ascribing meaning to others and ourselves even though these meaning ascriptions do not correspond to any facts. One useful way to understand what the skeptical solution purports to do is by way of comparing it with Hume's "projectivist"¹⁵ view of morality and causality. In Hume's view, sentences that we use to talk about moral issues such as "stealing is wrong" or causal events such as "a thrown baseball causes a glass window to break" do not, despite their superficial grammatical and syntactic similarities to other types of declarative sentences, report facts of the matter in the world. Rather, those sentences are expressions used to *project* our attitudes and sentiments onto the world. For example, we use the declarative sentence "stealing is wrong" simply to express our attitude of disapproval toward stealing behavior, rather than report an objective state of affairs in the world. There is no objective property of "wrongness" that we pick out in an act of stealing. Nor do there exist any objective causal relations over and above observed regularities in Hume's view. Our ordinary conceptions merely mistake observed regularities for "superlative"¹⁶ causal

¹⁵ See Kripke's (1982, p. 68-69) own comparison of KW's skeptical solution with Hume's skeptical solution. For Hume's skeptical solution, see Hume, *An Enquiry Concerning Human Understanding*, 1952. See also Blackburn's (1984) point on this.

¹⁶ This term comes from Kripke (1982), p. 69.

facts. It is our minds that project causal relations onto observed regularities. KW's skeptical solution is "skeptical" in exactly the sense Hume's treatment of moral and causal facts is: it denies the existence of any meaning fact, which is ordinarily conceived of as real and objective, belonging to the realm of an individual's mental, neuro-physiological, or behavioral properties. In KW's skeptical solution, meaning-ascribing sentences such as "John means plus by '+'" correspond to no facts. KW's skeptical solution is intended to be a "solution" to accommodate our ordinary meaning-attributing practice from the damage done by accepting the paradoxical skeptical conclusion. The problem, however, is how this can be achieved.

The key, according to the skeptical solution, lies in having the anti-traditional idea of language that a sentence does not have to be fact-stating to be meaningful and true. The traditional Frege-*Tractatus* thesis of language says that "a declarative sentence gets its meaning by virtue of its *truth conditions*, by virtue of its correspondence to facts that must obtain if it is true."¹⁷ Citing Dummett's work, Kripke argues that the later Wittgenstein replaces the traditional conception of language with a new approach in which a sentence is meaningful insofar as it has a suitable assertibility or justification condition. Asserting a sentence under such a condition has a certain role to play in our lives. Notice that this new conception of language applies to all declarative sentences, and is adopted by KW's skeptical solution to apply to meaning-ascribing sentences, which are considered among declarative sentences. The following paragraph describes how KW's skeptical solution, by adopting the new approach to language, is able to respond to the skeptical challenge:

¹⁷ Kripke (1982), p. 72.

Now the replacement of truth conditions by justifications has a dual role in the *Investigations*. First, it offers a new approach to the problems of how language has meaning, contrasted with that of the *Tractatus*. But second, it can be applied to give an account of assertions about meaning themselves, regarded as assertions *within* our language... Recall Wittgenstein's skeptical conclusion: no facts, no truth conditions, correspond to statements such as "Jones means addition by '+'"... Now if we suppose that facts, or truth conditions, are of the essence of meaningful assertion, it will follow from the skeptical conclusion that assertions that anyone ever means anything are meaningless. On the other hand, if we apply to these assertions the test suggested in *Philosophical Investigations*, no such conclusion follows. 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.¹⁸

It is important to note that KW's skeptical solution does not attempt to offer an account of what meaning "consists in". For a constitutive account of meaning that tries to find a necessary and sufficient condition of applying a term correctly has been rejected by the skeptical argument. KW's skeptical solution is very limited in nature in that it only aims to provide a description of assertability conditions, which cannot be taken as providing the content of the meaning-attributing sentences. The description in question does no more than to record the circumstances under which a speaker actually makes and assesses, as warranted or unwarranted, ascriptions of meaning to a sentence.

¹⁸ Kripke (1982), p. 77-78.

It turns out that referring to a linguistic community is essential, according to KW, to describing the assertability condition of a sentence, such as "I mean addition by '+'". Moreover, it is only in virtue of referring to a linguistic community, rather than to a person considered in isolation, that the possibility of error, which is essential to the normativity of meaning, can be captured. For, suppose we consider a person in isolation, who speaks a solitary language which includes the sign "+". We would be unable to distinguish between his genuinely meaning plus by '+' and his merely believing himself to mean plus by '+'. In such a case, whatever seems right to the person is right, "and that only means that here we can't talk about 'right'" (Wittgenstein, PI, section 258). Matters are quite different if we take into consideration a community in which the sign "+" is used. A speaker's use of "+" would be correct if it accords with that of most other competent speakers in the community and incorrect if it does not. In such a picture, in which agreement and disagreement on a response is possible only in a community, but impossible in the case of an individual considered in isolation. Only a public language can possibly exist, while a solitary language¹⁹ is impossible. Consequently, the skeptical solution amounts to an argument against a solitary language. This result, however, is highly problematic, as we will see below.

Two Objections to KW's Skeptical Solution

¹⁹ Here, following Goldfarb (1985), "Kripke on Wittgenstein on Rules", we use a "solitary language" to refer to the concept of a language constituted by properties of an individual considered in isolation. This concept is in Goldfarb's view, somewhat different from Wittgenstein's concept of a "private language. The latter is a more general idea.

Two objections, as far as I know, have been raised against KW's skeptical solution. The first objection²⁰ claims that in describing the assertibility conditions of a meaning ascription, we don't have to refer to agreement and disagreement of responses between an agent and other members in his community. Contrary to the communal depiction in KW's skeptical solution, this objection holds that in our actual practice, only an individual himself is referred to in the justification of his meaning-ascribing behavior. The second objection²¹ challenges the idea that there is genuine agreement and disagreement of responses, once we accept the skeptical conclusion (contrary to what the sceptical solution asserts). The consequence of making it clear that the contents of agreement and disagreement of responses among speakers are vacuous is that a public and a solitary language would be indistinguishable. Both objections are in my view convincing. I will review each of them in more detail below.

McGinn/Goldfarb/Boghossian's objection. Suppose that the skeptical solution is true in that no substantive assertibility conditions can be given without reference to the agreement and disagreement of a community of speakers. Then, according to McGinn (1984), the following case should be excluded, because its formulation refers to no one other than an individual considered in isolation:

- (A) We can assert with warrant that a speaker means plus by "+" when we see him responding to arithmetic problems with *sums* in most of the cases posed thus far.

²⁰ This objection can be found in McGinn (1984), Goldfarb (1985), and Boghossian's (1989a) summary of the former two's views.

²¹ Blackburn (1984) and Hoffman (1984) offer similar considerations along this line of objection.

McGinn argues that as a matter of fact, *our* practice of ascribing understanding of a concept to others is one that seems to adopt (A). To bring out this point, consider KW's communitarian proposal of meaning-attributions in the skeptical solution, as described in (B) below:

(B) Smith will judge Jones to mean addition by "plus" only if he judges that Jones's answers to some particular addition problems agree with those he is inclined to give. Or, if they occasionally disagree, he can interpret Jones as at least following the proper procedures...If Jones consistently fails to give responses in agreement...with Smith's, Smith will judge that he does not mean addition by "+". Even if Jones did mean it in the past, the present deviation will justify Smith in judging that he has lapsed.²²

In (B), the meaning plus is ascribed to a fellow speaker, say Jones, when his performances conform to what another competent speaker, say Smith, is inclined to perform in most of the cases. This view fits our practice of meaning-ascriptions at first glance and differs from (A) in that it relies on (other) members of a community, in this case, Smith, in describing Jones as a user of "+". However, a moment's reflection of our practice reveals that Kripke's proposal (B) is not strictly true of our actual practice. In our practice, Smith's inclinations to do arithmetic won't be regarded as a good measure of Jones's arithmetic performances unless Smith, when giving those calculations, is not drunk, tired, distracted, and so on. That is, Smith has to be a *reliable* "+-user. Smith's performances have to conform in most cases to the sums of arithmetic queries in the first place to anchor Jones being a "+-user. It turns out, thus, that our practice does not lend

²² Kripke (1982), p. 91.

any support to (B). Rather, it seems that (A) is the principle underlying our meaning-ascribing practice. Although the aim of the skeptical solution is to achieve descriptive adequacy, this aim is apparently not met. We can conclude therefore that the skeptical solution fails.

It may be argued, however, that adopting (A) in our actual practice involves a reliance on our *knowledge* about what counts as a correct or incorrect use of a term and such knowledge has been denied by the skeptical argument. (A) thus has to be excluded as a candidate for describing our meaning-ascribing practice. This is perhaps why (B) is introduced as a plausible solution to describing our practice of meaning ascription, even though (B) is not true of our actual practice. (B) has the prospect of success, not because of its descriptive adequacy, but because it provides normativity in our practice. (B) depicts normativity as grounded in agreement or disagreement in responses among speakers of a community

This defense of a communitarian account of our meaning-ascribing practice as described in (B) is blocked by the second objection to KW's skeptical solution as follows.

Blackburn/Hoffman's Objection. KW's skeptical solution holds that only by appeal to agreement or disagreement in responses between a speaker and other members of his community can the ground of normativity be provided. Objectors to this aspect of KW's skeptical solution, notably Blackburn (1984) and Hoffman (1985), contend that there is no genuine agreement or disagreement in responses, upon which the skeptical solution rests. To have genuine agreement or disagreement in responses, according to these objectors, we have to know first whether two responses to " $68 + 57 = ?$ ", such as "125" and "5", are different responses or not. But the skeptical challenge has shown us that it's

not a factual matter as to whether it is the *quus* function that is meant by "+" and hence "5" is a correct answer, or it is the *plus* function that is meant and hence "5" is a mistake. Similarly, responding "125" may be correct if the agent means plus by "+", but incorrect if he means quus by "+". Consequently, when two speakers both utter "5", there is no way to determine whether they agree or disagree in their responses, for there is no fact to determine whether or not one of them may mean a different function but have simply made a mistake in his response. KW's skeptical conclusion shows, therefore, that there can't be genuine agreement or disagreement in responses among speakers.

To the above objection, KW's skeptic may reply that even though there is no genuine agreement or disagreement in responses, it suffices if we *judge* the responses to agree or disagree. So far as there are *judgments* of agreement and disagreement in responses, there is consensus about what is to count as a correct or incorrect response and this renders meaning possible. The problem with this reply is that KW's skeptical argument can be applied again to those judgments in question, and there will be no way to differentiate judgments of agreement and disagreement. For example, suppose a speaker Smith judges that another speaker Jones' answer "125" is correct. KW's skeptic would argue that there is no fact that can settle the issue whether or not Smith is making an error in judging that Jones' answer is correct. That amounts to giving up the idea of having consensus in our judgments of agreement or disagreement. We merely seem to make judgments of agreement or disagreement while in fact there is no genuine differentiation between these two kinds of judgments. The crucial consequence is that we would have to give up the distinction between a public and a solitary language. For once we acknowledge that a public language is possible as long as it merely seems to us that

we make judgments of agreement or disagreement, then the fact that my present response seems to be in accord with my past responses should be sufficient to ensure the possibility of a solitary language. The upshot is then that if proceeding blindly is good enough for public language, it is puzzling why it is not good enough for solitary language. KW's skeptical solution, therefore, cannot be right.

KW's Skeptical Conclusion: An Eliminativist Position

Both McGinn/Goldfarb/Boghossian's and Blackburn/Hoffman's objections against KW's skeptical solution focus primarily on pointing out the defects of the ways in which the skeptical solution tries to accommodate the consequence of accepting the skeptical conclusion. This leaves us with an alternative option: why not give up the attempt, such as KW's skeptical solution, to repair the matter, and simply accept the paradoxical skeptical conclusion? The skeptical conclusion states that there is no fact that can determine whether one means one thing rather than another by a term, and thus implies an eliminativist claim about meaning, namely that there exists no such thing as meaning something by a term. As we have seen earlier, this position is both very paradoxical and destructive: it is paradoxical because we do seem to mean something by the words we use; it is destructive because it threatens ordinary notions of agency, pragmatic reasoning, and rationality that are central to our everyday lives. Nevertheless, it might be claimed that KW's skeptical argument is so powerful that the derived conclusion, though extremely counterintuitive and implausible, is what we have to accept and learn to live with this consequence.

Are we forced to accept the skeptical conclusion? Is there some way to resist taking this position, the consequence of which is so undesirable? Some philosophers, notably

Wright and Boghossian,²³ offer arguments that show that the skeptical conclusion is not even an option to take, because it involves incoherence. If their incoherence claim in question is right, then not only is the skeptical conclusion very likely to be false, but it *must* be false. Can such a strong thesis be sustained?

Wright's Incoherence Claim

Wright's argument (1984) tries to capture a very intuitive idea about the skeptical conclusion, that is, that the skeptical conclusion appears to be self-refuting. For if the skeptical conclusion were right that no one means anything by using a word, then the sentences used in constructing the skeptical argument would not be meaningful either. The skeptical conclusion could not be derived or even understood. As a result, the skeptical conclusion cannot, at least *prima facie*, be right. There seems to be something very wrong about the skeptical conclusion. Wright (1984) shows us where the mistake lies by pointing out that the skeptical conclusion, which amounts to meaning skepticism, leads to global skepticism and the latter is incoherent.

Wright's argument relies on the idea of the meaning-truth platitude. That is, the truth-value of a sentence is, presumably, determined by what the sentence means and what the state of the world is in relevant respects. If the determination of the meaning of a sentence is a non-factual matter as the skeptical conclusion states, then the truth-value of the sentence will turn out to be non-factual as well. Using Wright's analogy, if whether it's worthwhile to see an exhibition is determined by how well presented the

²³ Paul Boghossian, "The Status of Content", *The Philosophical Review*, 1990a. "The Status of Content Revisited", *Pacific Philosophical Quarterly*; 1990b, Crispin Wright, "Kripke's Account of the Argument Against Private language", *Journal of Philosophy* 81, 1984; *Truth and Objectivity*, 1992, esp. chapter 6 and appendix.

leading exhibits are, then if what determines a good exhibit is not a factual matter, it would not be factual either regarding whether it's worthwhile or not to go to see the exhibition. As Wright writes, "a projectivist view of meaning is thus, it appears, going to enjoin a projectivist view of what is for a statement to be true."²⁴ The consequence is that no sentence has a factual matter. It is then a surprising consequence of localized skepticism about the factuality of meaning that it entails global skepticism about the factuality of all sentences, i.e., sentences of all regions of discourse are not factual. Wright then puts into doubt the coherence of the global non-factualism in question:

...it is doubtful that it is coherent to suppose that projectivist views could be appropriate quite globally. For, however exactly the distinction be drawn between fact-stating and non-fact-stating discourse, the projectivist will presumably want it to come by way of a *discovery* that certain statements fail to qualify for the former class; a statement of the conclusion of the skeptical argument, for instance, is not *itself* to be projective.²⁵

Wright's worry is that KW's skeptic would presumably want it to be the case that the skeptical conclusion is *true* and thus is a genuine factual matter. However, the entailed global non-factualism applies effectively to the skeptical conclusion itself and renders it a non-factual matter. The skeptical conclusion thus appears, according to Wright, unstable.

But, is there really an incoherence in the skeptical conclusion? It seems not. As Boghossian points out, KW's skeptic would simply admit that "it is no more than *assertible* that no sentence possesses a truth condition."²⁶ Consistent with KW's skeptic's

²⁴ Wright (1984), p. 769.

²⁵ Wright (1984), p. 770.

²⁶ Boghossian (1989a), p. 525.

own position, KW's skeptic can adopt some sort of communitarian account as in the skeptical solution and claim that the skeptical conclusion is assertible under certain circumstances and asserting it under those circumstances has certain role to play in our lives. Thus, there is no incoherence involved in holding the skeptical conclusion.

Miller²⁷ raises and discusses a different line of argument for the incoherence of the skeptical conclusion that may be lurking in Wright's argument. The claim of the skeptical conclusion--that meaning statements are not factual--seems to rely on a *distinction* between factual and non-factual discourses such that meaning statements can be classified as belonging to the latter discourse. Could an account of this distinction be given by a proponent of the skeptical conclusion, asks the objector? The answer has to be no, because the consequences of the skeptical conclusion--global non-factualism--as Wright's objection shows, imply that no such a distinction exists. The objection is then that the skeptical conclusion depends on a distinction that its consequences seem to obliterate.

Is there an incoherence involved in the skeptical conclusion? Miller's own responses to this objection are quite compelling. A proponent of the skeptical conclusion, such as KW's skeptic, does not have to burden himself with providing a distinction between the factual and the non-factual. He can simply take our favored notion of factuality, whatever it is, and argue that under that notion not only is meaning discourse non-factual, but the discourse on factuality is also non-factual. The result then seems to be that, according to Miller, KW's skeptic can consistently reply that not only

²⁷ Miller (1998), Ch 5., p 171.

does the notion of meaning vanish into thin air but so does the notion of factuality.

Again, no transparent incoherence is involved in holding the skeptical conclusion.

Boghossian's Incoherence Claim

Boghossian's argument for the claim that the skeptical conclusion is incoherent consists in the following two claims: i) for a proponent of the skeptical conclusion to claim his anti-realist position, an irrealist, meta-linguistic formulation is necessarily called for, and ii) an irrealist formulation of the skeptical conclusion unavoidably leads to a contradiction. However, both of the above two claims are disputable. Let's examine each of them in turn.

Presumption of Semantic Formulations for The Skeptical Conclusion. It may first strike us as odd that a proponent of the skeptical conclusion has no coherent way to state his position, as Boghossian argues. Hasn't the skeptic's position been stated as that there is no fact constitutive of what one means by a term and hence that meaning statements correspond to no facts? Isn't this position, so stated, perfectly coherent? Why does a proponent of the skeptical conclusion necessarily need an irrealist thesis to state it?

To answer this question, we should first ask why we need a semantic formulation to express anti-realist positions and what semantic formulations are available to an anti-realist, such as a proponent of the skeptical conclusion. The answer is this. It is very likely that terms such as "phlogiston", "witch", and so on, refer to no real entities or properties in the world. A question then arises as to how we should treat sentences of our discourse that employ such predicates, whose extensions are empty. Traditionally, as Boghossian notes, two semantic theses have been proposed to formulate these anti-realist positions: one is the error theory, the other is non-factualism, depending on whether or

not predicates in a region of discourse express properties. An error theorist for a region of discourse holds that predicates in the region express properties, and hence that sentences containing those predicates with empty extension are normally false, for nothing in the world instantiates the properties in question. Eliminativist talk about witches is a good instance of an error theory: there are witch properties expressed by the "witch" predicate, but no human beings instantiate those properties; as a result, there exist no witches. In contrast, non-factualism about a region of discourse is a more radical view. A non-factualist holds that predicates in the region do not even express properties, and hence sentences containing those predicates do not have truth conditions and are neither true nor false. Classical ethical emotivism--roughly, the view that the predicate "is wrong" does not express moral properties and hence a sentence like "stealing is wrong" does not have a truth condition--is a good example of non-factualism.

Major anti-realist views on contentful entities, according to Boghossian, have been formulated either in the error-theoretic or in the non-factualist version. For example, Churchland's eliminativism on folk psychology is claimed by Boghossian to defend an error-theoretic thesis. Consequently, Churchland argues that all our talk about mental states is false and should eventually be abandoned. Dennett's instrumentalism on folk psychology is a non-factualist thesis. It claims that "attributions" of beliefs and desires to something do not aspire to represent real properties of that thing. Nonetheless, those "attributions" serve as useful tools for predicting behavior. Similarly, KW's meaning/content skepticism is interpreted by Boghossian as holding a non-factualist view on meaning or content discourse. Statements involving the notion of meaning or content have no truth conditions but only conditions of warranted use.

The question, however, as Devitt²⁸ points out, is that whether to hold a metaphysical position such as anti-realism that denies certain entities is quite different from how to construe talk that employs predicates whose extensions are empty. It may be the case that talk becomes rather difficult to make sense of from the old theoretical point of view, once we accept a new theoretical position. Nonetheless, this difficulty can be taken as a pressure for a theorist to find an alternative way of talking about the things whose existence is denied. Consider the following dilemma for an eliminativist such as Churchland. If an eliminativist believes his theory--which claims that there are no beliefs--to be true, then his theory must be false. For his believing it to be true shows that there exists a belief. Or, an eliminativist can hold no belief about whether his theory is true or not. Either way, his theory can in no way be accepted as true. It is obvious that this dilemma uses a question-begging strategy, for its formulation saddles an eliminativist with exactly what he denies--that an eliminativist *believes* his theory to be true. An eliminativist may owe us a new way of talking about folk psychology, but the difficulty of providing one does not show that his metaphysical position is intrinsically wrong. It would therefore seem that, unless Boghossian makes a strong case that a semantic construal is indispensable for certain metaphysical positions, Boghossian's charge that the sceptical conclusion is incoherent due to its semantic characteristics cannot be sustained.

Boghossian indeed offers an argument for the need to formulate an anti-realist position with semantic tools. When an anti-realist about P wants to make an ontological claim that there is no such property as being P, a problem occurs, according to

²⁸ Devitt (1990), "Transcendentalism about Content", *The Philosophical Quarterly*.

Boghossian, as to whether this claim can be coherently made without relying on meta-linguistic operations. If one tries to make the following negative existence claim:

There is no such property as that of being P,

he is saying, according to Boghossian, nothing different from

The property of being P is such that there is no such property,

which is apparently incoherent. Boghossian maintains that this problem can be avoided if meta-linguistic tools are introduced and used as the following:

The predicate "P" does not refer to a property.

This turns out to be the way in which an irrealist would formulate an anti-realist metaphysical position.²⁹

However, as Devitt & Rey³⁰ point out, Quine has convincingly shown us in his famous paper "On What There Is"³¹, that Boghossian's worry, like McX's, about the incoherence of making a negative existence claim is unnecessary. We do not have to presuppose, as Boghossian does, the existence of an entity whose existence is denied by a negative existence claim. As a result, Boghossian's intent to equate eliminativism and irrealism is ill-warranted. Devitt (1990) and Devitt & Rey (1991) contend that a metaphysical position should not be confused with a semantic thesis. When a metaphysical position finds no proper semantic tools to express itself, it only shows that the position is implausible. The most we can say, according to Devitt (1990), is that the position in question owes us a way of talking about things if our ordinary semantics and folk psychology turn out to be nonexistent.

²⁹ See Boghossian's discussion on this in "Status of Content Revisited", p265-6, 1990b.

³⁰ In "Transcending Transcendentalism", n18, p 99.

³¹ In Quine, *From a Logical Point of View and Other Essays*, 1953.

I will not discuss this matter further, since it hinges on the disputable issue of whether eliminativism and irrealism are inseparable³². It suffices to show that Boghossian's first claim--that a proponent of the skeptical conclusion has to express his position in an irrealist semantic formulation--is questionable.³³

Contradiction in the Skeptical Conclusion, Given Irrealism Presumption. Even if Boghossian's first claim, i.e., that KW's skeptical conclusion needs to be expressed in a meta-linguistic formulation, is granted, his second claim, that no meta-linguistic formulation of the skeptical conclusion can be coherent remains highly doubtful. In what follows, I will examine whether or not Boghossian's second claim is right. I will grant that KW needs a non-factualist semantic tool to express his metaphysical position about meaning and content, and see if it leads to a contradiction.

Non-factualism about any subject matter consists of the following two claims: in a declarative sentence of the form "x is P".

- (1) The predicate "P" does not denote a property
- and (hence)

³² Boghossian accuses Devitt of making a "major confusion" in which Devitt takes Boghossian's irrealist formulation of content eliminativism as a *distinct* thesis from the metaphysical thesis of content eliminativism; Boghossian intends his irrealist formulation to be an *expression* of content eliminativism. Boghossian's claims seem to obliterate a distinction between semantics and metaphysics, a distinction being challenged by by Devitt as leading to serious errors in the realism debate. See Devitt's *Realism and Truth*, 1991); "Aberrations of the Realism Debate", *Philosophical Studies*, 1991.

³³ Boghossian's argument that meaning/content eliminativism is incoherent relies on an intricate maneuver with the truth predicate. We will touch on this shortly. It may be worth pointing out that Miller's argument using Jane Heal's criteria of factuality for the coherence of KW's skeptical conclusion seems to presuppose that there is no need to further express the skeptical conclusion with meta-linguistic tools. See Miller, *Philosophy of Language*, 1998, chapter 5, section 3, and Heal, *Fact and Meaning*, 1990.

(2) The overall (atomic) declarative sentence in which it appears does not express a truth condition.³⁴

Boghossian then brings to our attention the fact that *non-factualism about any subject matter presupposes a notion of truth richer than the deflationary conception*. On a deflationary understanding of truth, "is true" is a predicate that does not refer to any real, language-independent property that a sentence or thoughts might possess. The predicate "is true" is merely a linguistic device for "semantic ascent" in Quine's phrase on a disquotational theory³⁵ or "a compliment we pay sentences we are prepared to assert" in Rorty's terms on performative theories³⁶. What sentences then are qualified for possessing truth conditions on a deflationary construal of truth? According to Boghossian, the following two minimal requirements must be fulfilled: first, the sentence must be significant, and second, it must be declarative in form. Clearly there is a tension³⁷ between non-factualism and a deflationist conception of truth. The tension is that a sentence fulfilling the two minimal requirements seems perfectly truth-conditional, for it is ready to be asserted, and in the meantime to be assertible is no more than being asserted as true on a deflationary understanding of truth. But non-factualism denies that the declarative and significant sentences of some targeted region of discourse are truth-conditional. What underlies this tension is the fact that non-factualism about a given

³⁴ Boghossian (1990a), p. 161.

³⁵ See W. V. Quine, *The Philosophy of Logic* (1986), pp. 11-12.

³⁶ See R. Rorty, "Pragmatism, Davidson, and Truth", in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. Lepore (1986), p. 334.

³⁷ Boghossian notes that many people also notice the tension between deflationism and non-factualism. For example, C. McGinn, *Wittgenstein on Meaning*, 1984, p71 and Simon Blackburn, "The Individual Strikes Back", *Synthese* 58 (1984), pp 281-302.

subject matter presupposes a conception of truth richer than the deflationary: it is committed to holding that the predicate "true" stands for some real property of truth and that this property is not enjoyed by certain sentences which are merely declarative and significant.

According to Boghossian, Kripke takes a radical route--a non-factualist approach--in formulating the skeptical conclusion. In such an approach, meaning-predicates like "means that P" do not express properties; and sentences like "S means that P", where S may be a sentence of English, a sentence of a language of thought, and so on, do not have truth-conditions. Meaning-ascribing sentences are therefore neither true nor false.

So if we apply the standard claims of non-factualism about any subject matter stated in (1) and (2) mentioned earlier, and assume that the essential core of meaning (and content as well) is truth condition, the skeptical conclusion that meaning-ascribing sentences do not correspond to facts can be formulated as consisting of the two following claims:

(3) The predicate "has the truth condition that P" does not refer to a property.

and

(4) For any sentence S and propositional content P, "S has truth condition P" is not truth-conditional.³⁸

However, (3) entails (5):

(5) "is true" does not refer to a property.

The entailment (3) $\rightarrow$ (5) holds because, according to Boghossian, the truth-value of a sentence is presumably a function of the sentence's truth condition. If a sentence's

³⁸ Boghossian (1990a), p. 175.

possession of a truth condition is a non-factual matter, then the truth-value of that sentence is also non-factual. Since the predicate "has the truth condition that P" does not refer to a property, as (3) states, the predicate "is true" does not refer to a property, as is stated in (5). What (5) amounts to is that "is true" is not what Boghossian calls "robust".

It becomes clear now, Boghossian points out, that a contradiction arises from the joint claims of (4) and (5). For on the one hand, claim (4) presupposes, as we saw earlier in the discussion of non-factualism in general, a robust notion of truth. The reason is that in (4), the significant and declarative sentence "S has truth condition P" is claimed as non-truth-functional. But since any significant and declarative sentence possesses a truth condition on a deflationary notion of truth, (4) presupposes a notion of truth richer than a deflationary. On the other hand, however, (5) amounts to a claim that denies exactly such a notion of truth. As a result, the non-factualist thesis about meaning consisting of (3) and (4) leads to a contradiction and hence an incoherence in the skeptical conclusion.

Is Boghossian's argument successful? It seems not. The problem, pointed out by Wright³⁹, is rooted in the assumption underlying (4). Boghossian claims that (4) presupposes that the predicate "...has truth condition P" or "is true" is robust. This is because the predicate in question refers to a property that a certain type of significant and declarative sentence (i.e., truth-condition-ascribing sentence) does not have, and that certain other types of significant and declarative sentences may have. What underlies the presupposition in question is the assumption that making a distinction between significant and declarative sentences that possess a *robustness* property (let's call them "truth-apt" sentences) and those that do not ("correctness-apt" sentences) is a substantial matter. In

³⁹ Wright (1992), *Truth and Objectivity*, appendix of ch6.

other words, it has to be assumed that statements that distinguish between truth-apt sentences and merely correctness-apt sentences have to be themselves truth-apt. As Wright points out, however, this assumption is inconsistent with (3). For (3) amounts to the claim that there is no *robustness* property and thus amounts to the denial that anything is truth-apt.

A proponent of the skeptical conclusion, therefore, would not accept the meaning/content non-factualist formulations of (3) and (4), unless the assumption in question, despite being very natural, is rejected. The result, according to Wright, would then be that all statements that make claims about whether a particular sentence belongs to the truth-apt category or to the correctness-apt category are themselves only correctness-apt. That is to say, it is not a substantial, factual matter whether a sentence has a truth condition or merely a correctness condition. Consequently, a proponent of the skeptical conclusion can coherently accept both (3), and its derivation (4), as far as he also holds that (4) (itself a statement that classifies a particular sentence as not truth-conditional and thus falls in the correctness-apt category) is only apt for correctness. It thus seems that there is no incoherence of the sort that Boghossian describes as inherent in meaning/content non-factualism stated by the joint claims of (3) and (4).

Wright's No-Rational-Support-for The Skeptical Conclusion Argument

It appears then that the skeptical conclusion remains an open option, since no conclusive result can be established for the incoherence charge. Where does this leave us? Remember that the philosophers who attempt to show the incoherence of the skeptical conclusion are motivated by an intuitive idea that there must be something very wrong with the skeptical conclusion, because it eliminates everything contentful, thus

seriously threatens our ordinary notions of rationality, agency, and cognition, and puts certain vigorous scientific disciplines such as cognitive psychology into danger. The incoherence claim, however, has been shown to be too strong. The skeptical conclusion may be highly implausible, counter-intuitive, and thus very likely to be wrong. But it cannot be shown that it must be wrong. Nonetheless, is there an argument that can allow us to make a somewhat weaker claim than incoherence, but still answers to our intuitive thought that there must be something wrong which renders the skeptical conclusion so implausible and counter-intuitive?

Yes, there is one such argument, offered by Wright (1995)⁴⁰. Wright's basic idea is to show that the skeptical conclusion is intrinsically incapable of receiving rational support. It takes two steps to show this. The first step is to show that the skeptical conclusion consists of a *non-factual* theoretical claim about the status of certain statements. The second step then shows that no non-factual claims have rationally compelling support. Let's go through each step more closely.

The skeptical conclusion makes the claim that meaning/content statements are not factual. As we have seen, this leads to the global claim that all statements are not factual. In other words, there turns out to be no genuine distinction between meaning sentences and sentences of other types. All claims that make distinction between factual and non-factual sentences are themselves non-factual claims. Consequently, the theoretical claim, made by proponents of the skeptical conclusion, that meaning statements are non-factual is hence itself a non-factual claim.

⁴⁰ Wright, "Can There Be a Rationally Compelling Argument for Anti-realism about Ordinary ("Folk") Psychology?", in *Philosophical Issues*, 6, 1995.

As a matter of fact, the same result can be obtained from the way in which a proponent of the skeptical conclusion states his anti-realist position in a non-factualist version. As we have seen in Wright's rejection of Boghossian's argument that (3) and (4) lead to a contradiction, a proponent of the skeptical conclusion can consistently hold (3) and (4) as far as he also holds that (4) is itself only apt for correctness. That is, as long as (4), the claim that meaning/content statements are not factual, is held as a non-factual claim.

We can now move to the second step. Wright connects the two terms "factual" and "cognitive" in such a way that they are definitive of each other: "...Roughly, factual matters are those which may be appraised just by cognitive capacities; cognitive capacities are those which enable the appraisal of factual matters."⁴¹ It follows from this construal of "factual" and "cognitive" that rejecting the factuality of a claim implies that only non-cognitive capacities are essentially involved in the appraisal of the claim. Now with the result of step one--the claim that the skeptical conclusion is non-factual--it follows that the claim in question can only be appraised by the exercise of non-cognitive abilities. Assuming then that reason is taken as a purely cognitive capacity, it would thus follow that holding the skeptical conclusion cannot be rationally compelling.

This argument seems to me to be the most convincing one to support our intuition that there must be something wrong with holding the skeptical conclusion that we do not mean anything by our words, that meaning statements are not factual. The intuition that something is intrinsically wrong with the skeptical conclusion is right. But it is not because the skeptical conclusion is internally inconsistent. Rather, it is in the nature of its

⁴¹ Wright, 1995, p 219. A more detailed treatment of "cognitive command" is given in his *Truth and Objectivity*, 1992, ch.3 and ch.4.

position that it can admit of no cognitively-based support. This may not do full justice to our intuition that KW's skeptical conclusion is self-refuting, and there might be some other way in which the skeptical conclusion is self-refuting. Nonetheless, Wright's no-rational-support argument does convincingly point out a critical reason why the skeptical conclusion cannot be sustained. This gives us strong reason to give up the skeptical conclusion. The only plausible option left for responding to KW's skeptical paradox is a straight solution.

Chapter IV. The Anti-Reductionist Solution to the Skeptical Paradox: An Unlikely Approach

In responding to KW's skeptical paradox, the two anti-realist accounts reviewed in the last chapter hold that KW's skeptical argument is sound, and hence accept the derived skeptical conclusion. These anti-realist solutions are far from satisfactory, due to the problem of how to construe our everyday meaning-ascribing practice, given the skeptical conclusion that our ordinary meaning-ascribing sentences correspond to no facts. On the one hand, the skeptical solution accepts the skeptical conclusion that a meaning-ascribing sentence does not correspond to any fact, but argues that it does not have to correspond to a fact in order for a meaning-ascribing sentence to be meaningful. As far as a meaning-ascribing sentence is assertible under some suitable condition, the meaning-ascribing sentence is perfectly legitimate and meaningful. More specifically, insofar as a speaker's responses regarding a linguistic term are in accord with those of most other fellow speakers of a same community, a sentence ascribing meaning to that speaker can be appropriately uttered and is hence meaningful. The problem, however, is that in our actual meaning-ascribing practice, agreements and disagreements with other speakers' linguistic responses are not counted in the attributing of meaning. Rather, only the speaker's own linguistic performances are relied on in our deciding whether to attribute meaning to the speaker or not. Moreover, no agreement or disagreement in linguistic responses between a speaker and other fellow speakers can be genuinely factual, once the skeptical conclusion is accepted. Consequently, the claim of the skeptical solution that a solitary language is impossible while only a public language is possible collapses. On the other hand, the skeptical conclusion itself has not been

examined critically. Once we examine the skeptical conclusion itself closely, we realize that the position in question is not internally incoherent, and is hence not definitely false, as a few philosophers claim that it is. Nevertheless, the skeptical conclusion cannot be a sustainable position, due to its being inherently incapable of receiving cognitively based support.

The remaining option left open to anyone who wants to find a satisfactory account of KW's skeptical paradox is then to find a straight solution, namely, to find faults inherent in KW's skeptical argument, so that the foundation upon which the skeptical conclusion is based could be shown to crumble.

Efforts to find a straight solution to KW's skeptical paradox fall under two rubrics--reductionists and anti-reductionists, depending on whether or not they accept the reductionist assumption implicitly held in KW's skeptical argument regarding meaning. Recall that when one claims that he meant plus by "+", KW's skeptic challenges him to cite a past fact that can determine what he meant by "+." For any attempt that proposes the meaning-fact as already possessing a determinate content, KW's skeptic persistently asks him to cite a further fact to determine which meaning his word expresses. Any further fact cited is then shown to be subject to various interpretations, and no justification can be offered in support of one's favoring one interpretation over another to fix the content of a word or thought. It is clear that KW's skeptic implicitly adopts the methodology of reductionism, which amounts to a requirement that meaning-constituting facts, if they exist at all, must be specifiable in non-semantic, non-intentional terms.

Reductionist approaches to KW's skeptical argument accept the reductionist assumption, and try to specify the meaning-constituting-fact without any use of semantic

or intentional notions. This group covers various sophisticated versions of dispositional accounts that aim to provide naturalistic accounts of meaning. In contrast, anti-reductionists simply reject the reductionist assumption implicitly held by KW's skeptic. They hold that meaning and intentional properties are real, and need not be further reducible to other naturalistic properties. According to the anti-reductionist view, meaning or intention is something with determinate content that is known directly and authoritatively by its possessor. Any requirement to go on to cite other more basic facts, such as dispositional or phenomenological facts, to determine the content of a symbol's meaning or an intention would simply amount to a bias against our ordinary idea of meaning and intention, according to which we simply know directly what we mean by our words and what we intend.

My aim in this chapter is to examine the plausibility of the anti-reductionist approach to KW's meaning/rule-following paradox. In particular, two major anti-reductionist positions will be discussed, one taken by John McDowell, and another taken by Crispin Wright. I will argue that neither anti-reductionist position is satisfactory. The reductionist approach to the same problem will be reviewed in the next two chapters.

McDowell's Wittgensteinian dissolution of KW's Skeptical Paradox

To refute KW's skeptical argument, McDowell (1984, 1992a) points out that Kripke had a seriously mistaken reading of Wittgenstein, which unfortunately leads Kripke to derive the paradoxical conclusion and to propose the ingenious but hopeless skeptical solution. Once we get the real Wittgenstein right, McDowell claims, KW's skeptical paradox would dissolve itself.

Recall that Kripke's Wittgenstein accepts the skeptical paradox, which amounts to the claim that there exists no linguistic meaning. In contrast, McDowell's Wittgenstein ("MW" for short) claims that linguistic meaning is real, and does exist. MW's position on meaning (and intention as well) can be summarized by the following three major interlocking themes: 1) meaning is real and irreducible; 2) meaning can exist only in a custom, or in a communal context; and 3) the aim of philosophy is to free us from illusions and misunderstandings caused by unnecessary theoretical constructions; hence, once certain mistaken constructive philosophies about meaning are removed, we get a clear view about meaning. The third theme, strictly speaking, is not directly about the nature of meaning or intention *per se*. Rather, it introduces a "quietist" meta-philosophy by Wittgenstein himself, i.e., the view, roughly, that "significant metaphysical debate is impossible"⁴², that lends important support to the way in which meaning and intention are treated in the first two themes. I agree that these three interlocking themes taken together give us a very convincing reading of Wittgenstein. Nevertheless, I will argue that getting an accurate reading of Wittgenstein is different from being able to satisfactorily dissolve the rule-following paradox that Kripke's Wittgenstein presents us. In what follows, I will go through each of the three themes in some depth.

The First Theme: Meaning and Intention are Real and Irreducible

McDowell (1992a) points out that KW's employment of the methodology of reductionism--cite a fact, which does not already presuppose meaning, to determine the meaning of one's word--presupposes a master thesis about meaning and intention. The master thesis states "that whatever a person has in mind, it is only by virtue of being

⁴² Wright (1992), p. 202.

interpreted in one of various possible ways that it can impose a sorting of extra-mental items into those that accord with it and those that do not."⁴³ The master thesis, embodied in Wittgenstein's rule-following considerations, assimilates word meaning (and intention as well) to a "sign post". Take a sign post with a duck-looking picture or a "duck" sign for example. What the sign post signifies is, as we may all agree, ambiguous unless an interpretation is set out to specify what it means. The problem with the view that the meaning of a sign post has to be mediated by an interpretation is that the interpretation is open to various further interpretations and none of them can be decisively determined to be the right one. By the same token, MW argues that underlying the reductionist assumption employed in the skeptical argument is the view, central to the master thesis, that meaning or intention, treated similarly to a sign post, is itself insufficient to fix its own content. The bad consequence of accepting the master thesis is an infinite regress of interpretation, which renders meaning and intention impossible. Hence, in McDowell's view, Wittgenstein's rule-following considerations, which embody the master thesis, does not purport to derive the skeptical paradox as KW does, but at most to reveal to us certain misconceptions about meaning and intention to which we might easily fall prey.

Besides, McDowell argues, the master thesis that meaning/intention is similar to a sign post is itself too counter-intuitive for a commonsense philosopher like Wittgenstein to accept. It is our common sense that we understand each other's word meaning directly, while the master thesis implies that understanding without mediation of interpretation is merely an illusion. McDowell believes that such a thesis is too counterintuitive to be true. McDowell himself offers independent reasons that we should

⁴³ McDowell (1992a), p. 45.

reject the master thesis. Consider the following case about intention that McDowell presents:

Suppose I am struck by the thought that people are talking about me in the next room. The hypothesis implies that only a state of affairs in which *people are talking about me in the next room* would be in accord with my thought. Now the master thesis implies that whatever I have in mind on this occasion, it cannot be something to whose very identity that normative link to the objective world is essential. It is at most something that *can* be interpreted in a way that introduces that normative link, although it can also be interpreted differently. ('I am free in the future to interpret it in different ways.').⁴⁴

Although McDowell here talks about intention, instead of meaning, the point he makes about intentions applies to meaning as well. For what is central to the notion of meaning is also central to the notion of intention, i.e., the idea of correctness and incorrectness. For example, suppose I have an intention to go hiking. Only my going hiking, rather than, say, going skating, will count as a correct act when I try to carry out the intention of going hiking. The key point is that my having an intention by itself sets up a standard of correctness for the evaluation of my behaviors.

Similarly, having a thought by itself brings about normative consequences. So, suppose, as in McDowell's example, that one has the thought that *people are talking about me in the next room*. There immediately follows a normative link between having the thought in question and states of affairs in the world. Namely, only if the world is in a certain state, i.e., people are talking about me in the next room, will my thought that

⁴⁴ McDowell (1992a), p. 46.

people are talking about me in the next room be true. However, this is the opposite of what the master thesis implies. What the master thesis implies is that one's thought that *people are talking about me in the next room*, never has the content that *people are talking about me in the next room*, but has at most something that can be interpreted as having that content. McDowell argues that this is too counter-intuitive to be true. For our commonsense intuition is that once a thought is tokened, its content is determined and transparent to its possessor. No further act of interpreting by oneself or others is necessary to fix the content of the thought. McDowell argues, therefore, that the master thesis that meaning/intention is equivalent to a sign post should be regarded as a target to be rejected in Wittgenstein's writing, rather than regarded as a means to create a paradox and reveal a need for a solution, such as KW's skeptical solution.

In the view of MW, rejecting the master thesis amounts to learning the following lesson: it is a misconception to think that giving an interpretation is needed to fix the content of one's meaning or intention. In McDowell's words, giving an interpretation to fix the content of an agent's thought is "over-mentalizing" the agent's thought. As shown in McDowell's own example, being in an intentional state is itself sufficient to determine the content of the thought. No interpretation from either the agent himself or other people is needed to fix the content of the agent's thought. The unwanted result of "over-mentalizing" our meaning-behavior is the generation of an interpretation regress, and the skeptical paradox.

MW further argues that in order to get a clear picture of meaning and intention, we have to avoid another major misconception. That is, it is wrong to think of meaning and intention as consisting in a dispositional fact. McDowell's Wittgenstein warns that, in

order to avoid “over-mentalizing” meanings or intentions as a consequence of accepting the master thesis, we may unwittingly swing to the opposite extreme of “under-mentalizing”. Namely, we may view meaning as simply what we are disposed to do, which is the product of training in one’s history of learning. In this dispositional account of meaning, meaning is simply how one has learned to react to a situation. For example, to mean blue by “blue” is to be disposed to call something “blue” when it is blue, or to mean plus by “plus” is to be disposed to respond “6” to the arithmetic question “ $3 + 3 = ?$ ”. McDowell argues that this dispositional depiction of meaning cannot be right, for it describes a person’s meaning something by a term as no different from issuing an involuntary and purely automatic physical movement in face of certain stimuli, such as one’s crying “help!” when he is drowning. According to MW, meanings and intentions involve something more than brute behavioral/dispositional outputs such as crying out “help!” when drowning. When we act on our understanding of a word’s meaning, we “take a commitment”, or “are faithful”, to the way we should use the term in order to be in accord with the meaning we grasp. In contrast, brute movements such as crying out “help!” when drowning are the core theoretical constituents of a causal/dispositional account of meaning. In the view of MW, this dispositional/behavioral picture may fit an automatism, but loses the core idea of meaning in which we act on the basis of our understanding of a word. Consequently, the crucial idea of norms, i.e., taking a commitment or being faithful to something, is left out in a causal/dispositional account.⁴⁵

A causal/dispositional account of meaning, therefore, cannot be right.

⁴⁵ We will have more discussions of McDowell’s ideas of a norm as “being committed to” or “being faithful to” something in the next chapter. McDowell takes Kripke’s Wittgenstein as presenting a powerful case against dispositionalism based on the normativity argument. However, McDowell’s idea of a norm.

With regard to the discussion of the dispositional approach to meaning and intention, MW and KW thus converge. They both think that norms, or correctness conditions, which are essential to meaning and intention, can not be provided by properties of an individual alone, e.g., an individual's dispositions. Nevertheless, KW embraces the skeptical paradox, whereas MW affirms the existence of meaning and intention. Section 201 of *Philosophical Investigation* (PI 201 for short) is important and can illustrate the position of MW, and how it differs from that of KW. PI 201 consists of two following paragraphs:

This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer is: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here.

It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us at least for a moment, until we thought of yet another standing behind it. What this shows is that there is a way of grasping a rule which is not an interpretation, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.

Kripke's mistake, McDowell claims, derives from merely focusing on the first paragraph of PI 201, while entirely ignoring the second paragraph, where an affirmation of the

i.e., "being faithful to something", is no more firmly fleshed out than KW's. The main concern here is to present the position of McDowell's Wittgenstein in rejecting dispositionalism. So it suffices here that we take a norm as whatever McDowell might think it involves, and that it is left out in a purely causal/dispositional account.

existence of meaning facts is apparently made. In McDowell's reading of PI 201, the second paragraph tries to say two things: 1) there is a misunderstanding in thinking that the first paragraph suggests that any proposed meaning-constituting fact is subject to interpretation regress and hence that there is no fact that constitutes meaning, and 2) there is a way of meaning something by a word that is not mediated by an interpretation. McDowell thus claims that Kripke regrettably misses these two important points when he interprets Wittgenstein.

We are now in a position to summarize the upshot of the first theme of McDowell's Wittgenstein, crystallized in McDowell's own reading of the two paragraphs of PI 201: there exists meaning and intention facts, which are irreducible to other facts.

The Second Theme: Meaning and Intention Can Only Exist in a Social Context

Having dispelled some tempting misconceptions about meaning and intention and reassured the reader that meaning and intention are real and irreducible, MW argues that meaning and intention lie in custom, or communal practice. In PI 198, a causal/dispositional account of meaning is rejected by Wittgenstein and a concept of a custom was proposed as the replacement of it:

Then can whatever I do be brought into accord with the rule?--Let me ask this: What has the expression of a rule--say sign-post--got to do with my actions? What sort of connexion is there here?--Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it. But that is only to give a causal connexion: to tell how it has come about that we go by the sign-post: not what this going-by-the-sign really consists in."--On the contrary; I have further

indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom. (PI 198)

According to MW, both the appeal to interpretations and the dispositional account fail for the same reason--no norms (i.e., standards of correctness) can be provided. Therefore, the concept of custom or a regular pattern of communal usage is proposed as an indispensable middle way of construing meaning that is neither the result of giving an interpretation nor the consequence of some training that one underwent in one's upbringing.

But, what exactly is a concept of a custom, or "a regular use of sign-posts", and how can it provide the needed norms of meaning and intention? MW claims that a custom forms the "bedrock", which provides the ultimate justification for meaning. McDowell cites the following passage of Wittgenstein in which the picture of "bedrock" is introduced:

How am I able to obey a rule?"--if this is not a question about causes, then it is about the justification for my following the rule in the way I do.

If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: "This is simply what I do. (PI 217)

McDowell points out that it may be tempting to think that at the basic level of "bedrock", there is nothing but brute behavior, coupled with feelings of being constrained by a rule. At such a basic level, norms are "leaching out" from "bedrock" and hence we simply act "blindly". However, McDowell argues that in Wittgenstein's view, norms are grounded at the basic level of "bedrock":

To use an expression without a justification does not mean to use it without right.
(PI 289)

In this "bedrock" picture, meaning and intention are already presupposed in our social practice:

Following according to the rule is FUNDAMENTAL to our language-game.
(RFM VI-28)

The difficult thing here is not, to dig down to the ground; no, it is to recognize the ground that lies before us as the ground.(RFM VI-31)

MW claims that the correct way to construe meaning and intention is to locate them at the right place. Neither is it above the "bedrock", or you would "over-mentalize" meaning and intention and hence confront a problematic skeptical paradox due to interpretation regress. Nor is it below the "bedrock", or you would "under-mentalize" meaning and intention and thus run the risk of "leaching out" norms from the picture of "bedrock". Rather, the norm lies at the "bedrock" where "normal human life is itself already shaped by meaning and understanding."⁴⁶

The picture of meaning and intention we get from the first and second themes of MW can now be presented as follows. Every competent speaker of a community knows what he means by a word and what the content of his thought is. However, it would be wrong to think that these contentful meaning or intention states, which a competent speaker is capable of being in, are properties of an individual speaker alone. For no properties of an individual, such as a disposition to act in a certain way or giving an interpretation, are capable of providing standards of correctness and incorrectness that are

⁴⁶ McDowell (1992a), p. 51. There McDowell quotes Wittgenstein: "Commanding, questioning, recounting, chatting, are as much part of our natural history as walking, eating, drinking, playing." (PI 25)

essential to the norms of meaning and intention. Only a custom, or a regularity of uses in a communal practice, can ground the much needed norms in question. As to the question, “how can a custom, or a communal practice, offer the ground of norms, i.e., standards of correctness and incorrectness?”, the answer of McDowell’s Wittgenstein is that norms of meaning and intention are already inherent in our everyday practice. Meaning or intending something is what we do all the time, and having an epistemic warrant in so doing is beyond question.

The Third Theme: A Quietist Approach to Meaning and Intention

MW posits that a custom forms the “bedrock” of meaning and intention, below which we should not dig. Nevertheless, it is clear that no attempt is given to provide a substantive analysis of correctness conditions or norms in terms of communal dispositions. Although MW makes vigorous efforts to illustrate that no properties of an individual alone are sufficient to ground the norms in question, he puts surprisingly little effort into arguing for the idea that norms are grounded in custom. Similarly, in Wittgenstein’s original text, the idea of a custom or practice is only briefly, though positively, stated as part of the correct picture of meaning and intention (usually right after Wittgenstein energetically argues that certain constructive philosophical ideas about meaning and intention are wrong). Little explication is given to make it more clear what a custom or practice consists in, let alone giving an analysis of correctness conditions by virtue of communal dispositions.

As a matter of fact, Wittgenstein regards the lack of vigorous philosophical constructions or theoretical explanations as characteristic of a legitimate and appropriate way of philosophizing:

Philosophy may in no way interfere with the actual use of language; it can in the end only describe it. For it cannot give it any foundation either. It leaves everything as it is. (PI 124)

Philosophy simply leaves everything before us, and neither explains nor deduce anything. (PI, 126)

In Wittgenstein's view, the task of philosophy is not to offer explanations of phenomena that we find puzzling, but to point out that the puzzles are often results of our being unfortunately misled by our own language:

The problems are solved, not by giving new information, but by arranging what we have always known. Philosophy is a battle against the bewitchment of our intelligence by means of language. (PI 109)

The results of philosophy are the uncovering of one or another piece of plain nonsense and of bumps that the understanding has got by running its head up against the limits of language. These bumps make us see the value of the discovery. (PI 119)

Should we then offer constructive answers to questions that we find puzzling, once we are shown why and how those puzzling questions arise? In Wittgenstein's view, clearing up possible misconceptions about a particular subject matter is the only aim of philosophy:

The real discovery is the one that makes me capable of stopping doing philosophy when I want to.—The one that gives philosophy peace, so that it is no longer tormented by questions which bring *itself* in question...There is not *a* philosophical method, though there are indeed methods, like different therapies. (PI 133)

A philosophical problem has the form: "I don't know my way about." (PI 123)

What is your aim in philosophy?—To show the fly the way out of the fly bottle. (PI 309)

The approach Wittgenstein takes to a philosophical problem is therapeutic, in which the urge to give constructive answers to philosophical problems is regarded as a form of “illness” that has to be “cured”:

The philosopher’s treatment of a question is like the treatment of an illness (PI 255) Wright describes this characteristic meta-position about philosophy of Wittgenstein’s as “quietism”, which states that no significant metaphysical debate is necessary. It is in this “quietist” spirit of doing philosophy that Wittgenstein warns us not to explain meaning and intention in sub-bedrock terms:

Our mistake is to look for an explanation where we ought to look at what happens as a “proto-phenomenon”. That is, where we ought to have said: *this language-game is played*. (PI 654)

The question is not one of explaining a language-game by means of our experiences, but of noting a language-game. (PI 655)

The most we have to do in understanding meaning and intention, in Wittgenstein’s view, is to notice a custom or practice of meaning and intention. “How is it possible” questions concerning meaning and intention are unnecessary and inappropriate. No serious account is needed to answer them.

We may now try to answer the following critical question: does the picture of meaning and intention offered by McDowell’s Wittgenstein manage to solve, or better, dissolve, the skeptical paradox of Kripke’s Wittgenstein? The answer seems to be negative.

Objection 1: The Idea That Correctness Can Only Exist in a Social Context Is Problematic

The key move that MW makes to dissolve KW's skeptical paradox is to reject the master thesis, and insist on the difference of levels in terms of "bedrock". The thesis is rejected on the ground that it is too counter-intuitive to be true. Meaning something by a term or thinking a thought in MW is a state the content of which simply has to be taken for granted, rather than fixed by interpretations. The problem, however, is that this position on meaning and intention is itself no less counter-intuitive. MW suggests that meaning and intention lie in a custom, or a communal practice, which forms the "bedrock". *On intuitive reflection, however, meaning or intention seems much more like a state of an individual than something that is embedded in a community. For when I apply a term that I understand, I perform in light of the meaning I grasp. Receiving some training in my upbringing, which takes place in a community, may be indispensable for my coming to possess such a state of understanding. But once this state of understanding is reached, the state itself seems to be sufficient for my using a term in a certain way, without having to make reference to a community.* Robinson Crusoe, a fictional character who survived a wreck on an island, remained a competent speaker according to our commonsense intuition. Crusoe's thoughts and spoken words were both understood to him to deal with things in his surroundings when he was living alone on the isolated island. The claim of MW that meaning and intention can only exist in a social context simply comes out of the blue. For neither does it follow from any positive and convincing argument, nor is it suggested by quietism.

McDowell responds to the above complaint that an individual's understanding does not seem to substantively depend on a communal practice by raising the following query:

...if the concept of a communal practice can magic meaning into our picture, should not this power be credited to the concept of a practice as such—so that the practice of an individual might serve just as well?⁴⁷

His own answer to the above query is as follows:

If, on the other hand, we reject the 'anti-realist' restriction on what counts as manifesting one's understanding, we entitle ourselves to this thought: shared membership in a linguistic community is not just a matter of matching in aspects of an exterior that we present to anyone whatever, but equips us to make our minds available to one another... This non-'anti-realist' conception of a linguistic practice gives us a genuine right to the following answer: shared command of a language equips us to know one another's meaning without needing to arrive at that knowledge by interpretation, because it equips us to hear someone else's meaning in his words... a linguistic community is conceived as bound together, not by a match in mere externals (facts accessible to just anyone), but by a capacity for a meeting of minds...⁴⁸

The essential point is the way in which one person can know another's meaning without interpretation.⁴⁹

⁴⁷ McDowell (1984), p. 351.

⁴⁸ McDowell (1984), p. 350-351.

⁴⁹ McDowell (1984), p. 351.

McDowell seems to claim that only in a social, interpersonal context can it be noticed that a speaker can understand another speaker's word meaning directly, without being mediated by interpretation. Communication among speakers of the same community simply takes place without their entertaining genuine doubt about what each other's words mean. In short, meaning is inherent and transparent in our communal practice.

Does this response of McDowell's enable us to better appreciate the claim that correctness, or meaning has to be grounded in a social practice, rather than an individual speaker? I do not think so. An individual speaker can know the meaning of his own words without interpretation as authoritatively as two speakers do in their communication. Then why does McDowell not credit an individual speaker as thoroughly as he does a community in the locating of meaning?

McDowell claims that "a linguistic community is conceived as bound together...by a capacity for a meeting of minds." What McDowell tries to say here may be a bit obscure. Nevertheless, we may well grant that it is indeed our commonsense intuition that we, as members of the same linguistic community, understand each other without needing a mediating interpretation, as if fellow speakers' minds are directly presented to us. However, isn't it equally commonsensical to say that I understand the meaning of my words without mediation of interpretation as if my mind is directly presented to myself? The insistence of MW in claiming that correctness conditions, or meaning, can only exist in a social context is, therefore, not convincing. My diagnosis of McDowell's fallacious reasoning is that he has confused a situation in which meaning is transparent with one which constitutes meaning.

*Objection 2: There Remains A Genuine Unresolved Puzzle Concerning The
First-Person Dimension of Meaning and Intention*

We may agree that there is a genuine insight in the claim of MW that an individual's state of meaning something by a word is real and irreducible. The insight in question is obtained from rejecting the master thesis—i.e., the thesis that a word meaning has to be mediated by interpretation in order for its content to be fixed on the ground that it is too counter-intuitive. Meaning, as well as intention, is real, and directly available to our mind. In contrast, KW is wrong in accepting the skeptical conclusion that there exists no linguistic meaning as a consequence of assuming reductionism, which presupposes the master thesis.

We remain, however, curious and indeed, puzzled, about what an individual's real meaning state is, if it is irreducible. In particular, in my view, in the background of reductionism that sustains the master thesis as used by KW, we discover that meaning is not an introspectible state with a given qualitative character. Yet we are aware of it with a fair degree of certainty under normal conditions. Meaning is not a dispositional fact either. Yet, it appears to have a dispositional characteristic of projecting onto the future standards by which the agent's behaviors can be deemed correct or incorrect. What can such a state be, which appears to possess both first-person epistemic accessibility and third-person dispositional projectivity? This is a thorny question that Kripke uncovers from Wittgenstein's rule-following discussions. Does the claim of MW that meaning something is a real, but non-reductive state, satisfactorily respond to KW's question just mentioned? Or does the quietist claim, i.e., that once certain misconceptions about

meaning are dispelled, the urge to find a constructive answer to a puzzling question such as KW's becomes eased, really help bring peace to our inquiring minds?

Here I side with Wright (1989b), in thinking that a non-reductionist still owes us an answer. Wright sees that there is a genuine puzzle discovered by KW, which is waiting and deserving to be solved by a constructive account. Wright characterizes this puzzle about meaning and intention as follows:

How is it possible to be effortlessly, non-inferentially, and generally reliably authoritative about psychological states which have no distinctive occurrent phenomenology and which have to answer, after the fashion of dispositions, to what one says and does in situations so far unconsidered?⁵⁰

Wright points out that it is an admirable achievement of Kripke to have uncovered one of Wittgenstein's fundamental puzzles about meaning and intention. Indeed, Wittgenstein brings out the puzzlement in question in an explicit way as follows:

"It is as if we could grasp the whole use of a word in a flash"—And that is just what we say we do. That is to say: we sometimes described what we do in these words. But there is nothing astonishing, nothing queer, about what happens. It becomes queer when we are led to think that the future development must in some way already be present in the act of grasping the use and yet it isn't present—For we say that there isn't any doubt that we understand the word, and on the other hand its meaning lies in its use. (PI 197)

Wittgenstein himself might not be very sympathetic with any constructive response to the puzzle in question about meaning and intention. However, as far as I can

⁵⁰ Wright, 1989b.

see, taking an anti-reductionist and quietist Wittgensteinian position as depicted by McDowell about this puzzle simply leaves the puzzle unresolved.

Wright's Judgment-Dependence Account of Meaning and Intention

To respond to KW's skeptical paradox about meaning, Wright⁵¹, like McDowell's Wittgenstein, takes an anti-reductionist position. Wright agrees with MW that word meaning and intention are real and directly known by a competent speaker. Nevertheless, he takes up the challenge of solving the Wittgensteinian puzzle about first-person phenomenology of meaning, while MW leaves it untouched. More specifically, Wright views the challenge in question as primarily involving how to reconcile the first-person authority of one's knowledge about the meaning of his own speech and the content of his own intention on the one hand, and the disposition-like projectivity of the meaning of a word and the content of an intention on the other.

Before entering the discussion of the plausibility of Wright's solution, it is important to note that the starting point of Wright's Wittgenstein is somewhat different from that of Kripke's Wittgenstein. Kripke's Wittgenstein is highly skeptical about the reality of meaning and intention of an individual considered in isolation. KW raises the question: how do we know which function I (used to) mean by "+"? Having found that nothing appropriate about me is constitutive of my meaning any function by "+", KW concludes that I don't mean any function by "+" at all. In contrast, Wright's Wittgenstein ("WW" for short) does not question that I do mean something by "+". WW assumes from the start that I know which function I mean by "+", but questions what the function that I grasp requires of me here and now. The main target of KW is hence the ontological

⁵¹ Wright, 1989a.

status of meaning, while the primary concern of WW are epistemological puzzles of meaning. However, this does not mean that Wright is not concerned with the metaphysical nature of meaning. On the contrary, Wright's strategy is that, by focusing on answering the question about the epistemology of meaning and intention, i.e., how do we come to know our intention and meaning in a way that is non-inferential, immediate, and authoritative, we can reveal the metaphysical status of meaning. KW's skeptical paradox would then be solved satisfactorily.

Wright proposes a judgment-dependence account of meaning and intention as a metaphysical solution to KW's skeptical paradox. In what follows, I will first review Wright's judgment-dependence account, and then argue that it does not solve KW's skeptical paradox.

The Metaphysical Nature of Meaning and Intention as Secondary Properties

Wright's metaphysical solution to KW's skeptical paradox begins with an epistemological question: Given that we know our intentions and meanings, how do we know them? More critically, how do we know them in a non-inferential, direct, and generally authoritative way? Wright's account of meaning and intention consists in claiming that the metaphysical status of meaning and intention is like that of secondary qualities. The best way to illuminate what secondary qualities are metaphysically is to contrast them with primary qualities. A property of an object is a primary quality insofar as the property in question is determined *independently* of anyone's judgment about or response to it. In other words, if the constitution condition of an object's quality is *independent* of the cognitive apparatus of any potential perceivers, then the quality of the object is a primary quality. Shapes, for example, are usually taken to be primary qualities

of objects, for their constitution conditions are regarded as independent of anyone's judgment about them. In contrast, a quality of an object is a secondary quality, as far as the quality's constitution condition is *dependent* on the cognitive mechanisms of any potential perceivers. For example, colors are often regarded as secondary qualities of objects, for their constitution conditions are thought to be dependent upon judgments or responses of those who perceive them.

Wright claims that meanings of words and intentions are best construed as secondary qualities. Wright argues that truths about our intentions and meanings of words are dependent, rather than independent, on our judgments, made under certain cognitively ideal conditions about them (let's call this kind of judgments "best judgments"). For example, whether or not I have the intention to go hiking is dependent on whether or not I judge, under cognitively ideal conditions, that I have the intention to go hiking. Similarly, whether or not I mean plus by "+" is dependent on my best judgment concerning whether or not I mean plus by "+". Therefore, in Wright's view, meaning and intentional properties are assimilated to secondary properties. On this view, meaning and intention are essentially "judgment-dependent."

The Judgment-Dependence Model of Explanation and Extension-Determining Concepts

In this section, we discuss how Wright explicated the idea of "judgment-dependence," which is central in the characterization of secondary qualities such as red, by introducing the pair of concepts: *extension-determining* and *extension-tracking concepts*. In next section, we will discuss how Wright applies the judgment-dependence account, which models in particular the concept *red*, to the concept of meaning and

intention. After that, we will evaluate whether Wright's attempted application is successful or not.

To see what a judgment-dependence account amounts to, and what it can serve to explain, let's first consider the following biconditional (P):

(P) For any act x : x is pious if and only if it is judged as pious by the gods.

Suppose (P) is true, and hence that there is a covariance between what is a pious act (or, the extension of the concept *pious*) and what is judged as a pious act by gods. According to Wright⁵², there are then two broad ways in which the truth of (P) can be explained.

The first is to treat the piety of an act as a state of affair that exists independently of the gods' judgments toward it but can nonetheless be unfailingly *tracked* or *detected* by the gods, due to the gods' perfect cognitive power. In such case, the priority is put on the left-hand side of the biconditional (P), such that facts about the extension of the concept *pious* are judgment-independent. This explanation of the truth of a biconditional like (P) is called the *judgment-independence model* of explanation. According to this model, the concept *pious* is an extension-tracking concept, which means that the extension of the concept *pious* is independent on the gods' judgments about the concept's extension.

The second way to treat the truth of (P) is to view the truth about a pious act as dependent on the gods' judgments. In such case, the priority is put on the right-hand side of the biconditional (P), such that it is the gods' judgments about an act x alone, rather than any independently constituted fact, that *determines* whether or not the act x falls under the concept *pious*. According to this model of explanation, facts about what falls under the concept *pious* are *judgment-dependent*. This model of explanation is then

⁵² Wright, 1992.

called the *judgment-dependence model* of explanation. According to this model, the concept *pious* is an extension-determining concept, i.e., the extension of the concept *pious* is dependent on the gods' judgments about the concept's extension.

With both the judgment-dependence and judgment-independence models of explanation at hand to explain the truth of a biconditional like (P), Wright proceeds (1989a) to reflect on a biconditional, (R), which involves the color concept *red*:

(R) For any x and any S, x is red if and only if S judges that it is red.

Suppose that (R) is true, namely, that there is a covariance between one's judgment about what falls under the concept *red* and the extension of the concept *red*. Wright claims that 1) it is plausible that (R) is *a priori* true, and 2) that the best explanation for (R)'s being an a priori truth is the judgment-dependence account, according to which the concept *red* that figures in (R) is an extension-determining concept. Let's examine both claims a little more closely. First consider another biconditional (S) that involves the concept of *square*:

(S) For any x and any S, x is square if and only if S judges that it is square

Wright argues that, based on Lockean intuition, (S), if true, is more likely to be an a posteriori truth. Although it seems that there is an a priori and constitutive connection between the extension of the concept *red* and one's judgments about objects as red, the link between the correct extension of the concept *square* and one's judgments about the shapes of objects seems to be contingent and a posteriori.

We may grant Wright's claim that (R), if true, is a priori true (we will have more discussion of the contrast between R and S shortly). It is still not right for Wright to make the claim that the judgment-dependence account can be offered as an explanation to

the a priori truth of (R). For not every judgment of an agent could unfailingly determine the extension of the concept *red*. For instance, judgments made under deceptive lighting or made by those who are color-blind have to be disregarded. Therefore, judgments about the extension of a concept that figures in a biconditional like (R) have to be suitably constrained in order for the judgment-dependence account to serve as an explanation of the a priori truth of (R).

To incorporate suitable constraints on judgments, Wright (1989) replaces a biconditional like (R) by what he calls a *provisional equation* (R*):

(R*) For any x and any S, C (S) $\rightarrow$ (x is red if and only if S judges that it is red)

(R*) is a conditional whose antecedent involves the condition of an observer S. Only when the condition C is suitably specified can the agent S's judgments play an extension-determining role. One important constraint to note on the specification of the C-condition is that it must be non-trivial, for any provisional equation like (R*) would be an a priori truth if the specification of the C-condition includes whatever it takes to ensure the correctness of S's judgments. For example, if the C-condition were specified as "S did not err in making his judgments concerning the color of an object", then (R*) would be trivially true. In other words, unless the C-condition is non-trivially specified, the covariance between an agent's best judgments and the extensions of the concept such as *red* involved in the provisional equation (R*) would need no explanation. Wright⁵³ offers a plausible non-trivial specification of the C-condition as follows. Roughly, S's visual equipment is statistically normal in human kind. S grasps the concept of redness. S is

⁵³ Wright (1988), "Moral Values, Projection and Secondary Qualities", *Proceedings of the Aristotelian Society*, supp. Vol. 62. p. 15-16.

viewing objects in the conditions which normally prevail at noon on a summer's day out of doors and out of shadow.

Wright points out that there is another important constraint on the specification of the C-condition. That is, the question as to whether or not the C-condition is satisfied has to be answerable without relying on any truth concerning the details of the extension of the concept, such as *red*, under investigation. That is to say, there has to be an *independence* relation between the satisfaction of the C-condition and the details of the extension of the concept *red* occurring in a provisional equation like (R*). The idea of adding this "independence" constraint is that when C is actually obtained, non-trivially specified judgments serve to determine which objects fall under the concept *red*. Otherwise, this cannot be done if the satisfaction of C itself depends on truths concerning what falls under the concept *red*. Indeed, without the "independence" constraint, Wright's judgment-dependence account modeling on secondary qualities such as red would crumble.

Finally, there comes the last constraint for a provisional equation like (R*) to meet, in order for the judgment-dependence account to explain the a priori truth of (R*). That is, there has to be no better account available that can explain the a priori truth of (R*), even though the C-conditions in (R*) are non-trivially specified and meet the independence constraint. We can infer that judgments made under the C-condition (i.e., best judgments) constitutively determine the extension of the concept under scrutiny insofar as there is no better explanation of the a priori and non-trivial covariance between the best judgments about the extension of the concept and truths about the extension.

To summarize, the following four conditions have to be met by a provisional equation like (R^*) ⁵⁴, so that an agent's judgments can determine the extension of the concept that figures in (R^*) :

- i) Apriority: (R^*) is a priori true;
- ii) Substantiality: (R^*) is substantially specified. In particular, the C-condition does not take a "whatever it takes" formulation;
- iii) Independence: the satisfaction of C-condition is, in any particular case, logically independent of the details of the extensions of the concept *red*;
- iv) Extremal: there must be no better explanation of why (i)–(iii) are satisfied than the claim that an object is red because it is judged to be red in the C-condition.

Wright argues that (R^*) meets the above four conditions, and, consequently, the best account for the truth of (R^*) is the judgment-dependence account.

In contrast, the provisional equation (S^*) that involves the shape concept *square* will not meet the conditions (i) and (iii) simultaneously:

(S^*) For any x and S , $C(S) \rightarrow (x \text{ is square if and only if } S \text{ judges that it is square})$

The reason, according to Wright (1989a), is as follows. A subject's best judgment about three-dimensional shape, if visually grounded, has to be the product of several observations from a variety of suitable spatial positions. In order for it to be a priori true that the subject's best judgment about the shape of the observed object is correct, it has to be guaranteed that the shape of the object remains *unchanged* during the period through which those several observations are conducted. Hence, the C-condition of (S) has to be

⁵⁴ Wright (1989a), p. 247-8.

specified in a way such that the shape of the object during the observation period is *stable*. The consequence of this is as follows. To know whether the C-condition is satisfied or not, we have to know first whether or not the shape of the object remains unchanged when those several observations are conducted. In order to do that, however, we have to know some prior facts about the shape of the object independently of the best judgments formed by the observer. The independence condition (iii) is thus violated, because the C-condition cannot be specified independently of facts about the shape of the observed object. Furthermore, Wright argues that it is highly doubtful that there is any prospect for (S*) to be an a priori truth, even if the stability problem in the specification of the C-condition is ignored. The reason is that it is a “deep fact of experience”⁵⁵ that our best judgments about the shape of an object based on visual observations of the object are evaluated for their correctness according to more refined operational techniques, such as taking measurements with rulers. (S*) is therefore best construed as an a posteriori truth, such that the covariance between truths and judgments is explained by the fact that the best judgments merely serve to track, rather than determine, the independently constituted extensions of the concept *square*. As a result, the judgment-dependence model of explanation is not applicable to the *square* concept.

The Judgment-Dependence Account of Intention and Meaning

Having explicated the judgment-dependence account modeled on secondary qualities such as red, Wright applies the same account to intention. Wright claims that the judgment-dependence account he offers provides a convincing model of explanation of what our concept of intention and meaning is. Namely, our concept of intention and

⁵⁵ Wright (1988), p. 20.

meaning is an extension-determining concept, just like our concept of colors such as red. Wright does not explicitly apply the account in question to meaning. Nevertheless, it should be clear that if the account is appropriate in terms of explaining intention, it can explain meaning equally well, because meaning and intention both share the two important characteristics, i.e., first-person accessibility and third-person disposition-like projectivity.

In order for Wright's judgment-dependence account to be applicable to the concept of intention, however, Wright has to show that the four conditions (i) to (iv), which we mentioned earlier when we discussed the provisional equation (R*), are met by the provisional equation (I*):

$$(I^*) \ C(S) \rightarrow (S \text{ intends } P \text{ if and only if } S \text{ judges that he intends } P)$$

(I*) is an provisional equation that has the concept *intention* figuring in it. Wright's attempt is to substantiate the claim that, if (I*) is a priori true and meets the independence condition, and if the C-condition in (I*) is non-trivially specified, S's judgments about the contents of his own intentions and the truth about the contents of his own intentions match perfectly, supposing that there is no alternative account available.

That (I*) is a priori true seems plausible, for all that is required to know the truth of (I*) is to understand the concept of intention. Wright specifies the C-conditions roughly as follows: an agent S has mastery of the appropriate concepts requisite for the expression of the intention, is not prey to material self-deception, is appropriately attentive, and so on. Wright claims that this specification of the C-conditions is substantial, and the satisfaction of the C-conditions is logically independent of the truth about S's intentions. Moreover, according to Wright, there is no better account

forthcoming that can explain the a priori and substantial covariance between S's best judgments about his intentions and facts about his intentions, unless "Cartesianism" is plausible, and Wright claims that Cartesianism is implausible.⁵⁶

For (I*) to meet the above four conditions the major difficulty lies in condition (ii), i.e., (I*) has to be specified non-trivially. Wright himself notices that the no-self-deception condition cannot be fleshed out except by trivially specifying something like "a condition to the effect that the subject be free of any condition which might somehow impede his ability reliably to certify his own intentions."⁵⁷ Wright does not deal with this problem by looking for new formulations of the no-self-deception condition. Rather, he argues that the no-self-deception condition is "positive-presumptive", which means that we are a priori justified in holding that, in the practice of an agent self-ascribing intentions, the no-self-deception condition is satisfied in the absence of positive evidence to the contrary. We can then modify the provisional equation (I*) by deleting the no-self-deception condition. The modified provisional equation will be *a priori credible*, rather than *a priori true*. Similarly, other parts in the C-condition, such as attentiveness, and possession of relevant concepts, can be construed as holding the "positive presumptive" status as well. Unless there is evidence to the contrary, it is a priori reasonable to believe that a subject is attentive and has mastery of relevant concepts. So now we have the a priori reasonableness of the supposition that one's judgments about one's own intentions are best. What then can explain the covariance between these best judgments and truths

⁵⁶ See Wright's discussion of a Cartesian account of intention and meaning and of the great difficulties that face it. My main concern regarding Wright's judgment-dependence account will not hinge on his reasons for rejecting the Cartesian account, So I will grant his refutation of Cartesianism.

⁵⁷ Wright (1989a), p. 251.

about our own intentions? Wright argues that the best explanation is the judgment-dependence account, according to which the best judgments under the C-condition "play a *defeasible* extension-determining role, with defeat conditional on the emergence of evidence that one or more of the background, positive-presumptive, conditions are not in fact met."⁵⁸

The First-Person Epistemic Accessibility and The Third-Person Dispositional Projectivity of Intention and Meaning

If Wright's judgment-dependent account modeled on secondary qualities such as red is successfully applied to explain the substantial and a priori covariance between our best judgments about our own intentions and facts about our intentions, then ontologically the intentional property is secondary. That means that the constitution conditions of one's intention are dependent on one's own best judgment about this intention. One has an intention that P, insofar as one judges that he has the intention that P, under suitable conditions. Note that the judgment-dependence account of intention is a non-reductive account, for the item that has to be explained, i.e., one has an intention that P, appears in the item that explains, i.e., one's best judgment that he has an intention that P. In Wright's account, one's having an intention that P is not explained in terms of something that does not already employ intentional or semantic content, as a reductive account typically does. Rather, one's best judgment about his own intention that serves to explain what it is for one to have an intention already contains the notion of intention. Wright thus offers a non-reductive and constitutive account of intention.

⁵⁸ Wright (1989a), p. 252.

The benefit of such a non-reductive and constitutive account of intention is that, according to Wright, it both reveals to us the metaphysical nature of intention as secondary, and explains to us the peculiar ways in which we know our own intentions, i.e., in a non-inferential, immediate, and authoritative way. The whole idea behind Wright's account is that there is a constitutive link between one's best judgments about one's own intentions and truths about one's intentions: our intentions are simply what we say they are, unless there is evidence to the contrary.

But how can Wright's judgment-dependence account of intention explain the disposition-like projectivity of intention? Projectivity is said to be a feature of intention such that once an intention is in place, it sets a standard for evaluating correct and incorrect performances that have not yet come into existence. The difficulty of accommodating the projectivity feature of intention derives from the observation that when one forms an intention, one may not have considered what behavior in the future will count as conforming to the intention. But our intuitive idea about intention is that having an intention inherently possesses such a projective capacity. In short, the problem concerning dispositional projectivity is that those so far unactualized behaviors are already explained by or contained in my present intention. But how those unactualized behaviors do so without being entertained by me now remains unclear.

Wright's solution to the above problem was that whether or not a person's future behavior is in accord with his original intention is determined by his future best judgment about his own future behaviors. Edwards nicely illustrates Wright's idea as follows:

Suppose a subject believes that he has formed an intention to continue the series--2, 4, 6--in a certain way and he writes '8, 10, 12.' He need have given

no thought as yet to what he will write after '12', but when he comes to that point he confidently continues--14, 16, 18--and judges that he is following out his former intention, that he is, in Wittgenstein's phrase, "going on in the same way." Wright's idea is that just as his original judgment that he had that particular intention constitutively determined that he did indeed have that intention, so his later judgment, that writing '14, 16, 18' implemented rather than frustrated that intention, constitutes the fact that such behavior does implement that particular intention.⁵⁹

It is important to note that such treatment of intention deviates from our intuitive notion of intention that having an intention is regarded as dictating a certain infinite pattern of behaviors that exist independently of human judgments. According to Wright's judgment-dependence account of intention, what intention an agent has is not entirely determined until his best future judgments about his own future behaviors set in. In other words, there is no pre-existing pattern of correct behaviors waiting ahead of an agent which was dictated by the agent's current intention. What the agent should do in the future in order to conform to his past intention has to be determined by his future judgments about his own intention and behaviors.

Although Wright does not directly apply the judgment-dependence account for the meaning of a word, the same treatment Wright offers to intention can be adopted in dealing with meaning. The reason is that meaning and intention share two of the most essential features, i.e., first-person authority and disposition-like projectivity. Just as in the case of intention where a subject's best judgments about his own intentions, both past

⁵⁹ Edwards, "Best opinions and intentional states", *Philosophical Quarterly*, 1992, p. 24.

and present, defeasibly determine what his intentions are, Wright claims that an agent's best judgments concerning what he means by a word, both past and present, are also analogously defeasibly extension-determining. As Wright writes:

Challenged to justify the claim that I formerly meant addition by '+', it will not be necessary to locate some meaning-constitutive fact in my former behavior or mental life. A sufficient answer need only advert to my present opinion, that addition is what I formerly meant, and still mean, and to the a priori reasonableness of the supposition, failing evidence to the contrary, that this opinion is best.⁶⁰

Is Wright's judgment-dependent account of meaning and intention plausible? In particular, does it adequately solve KW's problem? It seems not. The reasons are as follows.

Objection 1: Wright's Account Is Circular

Several philosophers⁶¹ have pointed out that Wright's judgment-dependence account of intention faces difficulties concerning how a provisional equation such as (I*) that involves the concept of intention can meet the independence condition. The independence condition requires that whether the C-conditions are satisfied or not has to be logically independent of any truth about the extensions of the concept under examination such as *red* or *intention*. That is to say, to know whether the C-conditions are satisfied or not, we must not already know some prior facts about what the agent's intention is. Otherwise, the claim that it is an agent's best judgment about his own intention that fixes the fact about his intention would be seriously undermined. The idea

⁶⁰ Wright (1989a), p. 254.

⁶¹ For example, Miller, 1989; Boghossian, 1989a, etc.

of the independence condition, therefore, is to ensure that the fact about an agent's own intention is constituted by the agent's best judgment about what the fact is, rather than any independently constituted fact.

Boghossian⁹² raises an interesting and fundamental difficulty for Wright's judgment-dependence account with regard to satisfying the independence condition. Consider the following provisional equation (M*) that involves the concept of *meaning add 2* by "add 2":

$$(M^*) \quad C \leftrightarrow (S \text{ judges that he meant } \textit{add 2} \text{ by "add 2" if and only if he meant } \textit{add 2} \text{ by "add 2"})$$

According to the independence requirement, to determine the content of the right-hand side of the biconditional in (M*), i.e., to determine what the term "add 2" refers to, no prior facts about the content of the right-hand side can be presupposed when we specify the C-conditions. Indeed, as Boghossian claims, anything else in the antecedent of (M*) that serves to constitute the content of the right-hand side is precluded from presupposing any prior facts about the content of the right-hand side of (M*). Boghossian then argues that the independence requirement could in no way be satisfied by Wright's judgment-dependence account. For the content of S's best judgments on the left-hand side of (M*), said to constitute the content of S's meaning on the right-hand side, has to be presupposed by the account itself. That is to say, we already have to know what S meant if we rely on S's judgment about what he meant. The consequence is that, according to Boghossian, any such provisional equation like (M*) "will always presuppose a

⁹² Boghossian, 1989a.

constitution of mental content quite independently of constitution by best judgment.”⁶³

Wright’s judgment-dependence account thus appears unable to get off the ground, for the account is apparently circular.

Objectionn 2: Wright’s Account Is Highly Counter-intuitive and Incapable of Solving the Normativity Problelem

The circularity problem for Wright’s account presented above by Boghossian may somehow strike us as odd, for it is clear from the start that Wright does not intend to provide a reductive account of meanings and intentions. Wright’s attempt is to offer a non-reductive account of meanings and intentions such that it could explain the peculiar ways by which we know them. The contents of meanings and intentions are indeed already presupposed by Wright’s judgment-dependence account. What then shall we conclude from Boghossian’s attack on Wright’s account? Shall we hold that Wright’s account remains intact, for Boghossian’s attack is totally off-target? After all, not all explanations have to be of reductive form. As Pettit⁶⁴ points out, there can also be genealogical or conceptual accounts that are non-reductive but have an important explanatory role to play. Therefore, Wright could well claim that as far as his non-reductive account offers an interesting and significant story about the epistemology (and ontology) of meanings and intentions, his account should not be rejected on the ground that it is circular. Alternatively, shall we say that Wright’s account is fundamentally flawed, as was shown by Boghossian’s criticism that facts about content are presupposed by Wright’s account, and hence should better be construed as being constituted independently of best judgments?

⁶³ Boghossian (1989a), p. 547.

⁶⁴ Pettit, 1990b, “Affirming the reality of rule-following”, *Mind*, p. 435.

My diagnosis of Wright's judgment-dependence account under the attack of Boghossian's charge of circularity is that Boghossian's attack does raise a problem to Wright's account. In my view, there is actually a deeper and more fundamental problem that underlies Boghossian's charge of Wright's account's committing to circularity. Below, I will try to show what the deeper problem in question is. After uncovering what the deeper problem is, I will show that Wright may have a way to evade Boghossian's circularity problem. But, I will then argue that Wright's way out only encounters more serious problems.

Boghossian's circularity charge against Wright's account arises from the fact that Wright's account identifies a meaning fact (or intention fact) as constituted by something, e.g., a judgment, of a *propositional* form. The approach of appealing to something of a propositional form is doomed to fail in the determination of semantic (or mental) content, for as Wittgenstein has energetically and convincingly shown us, semantic (or mental) knowledge cannot consist in anything of a propositional form. The reason is that meaning-constituting candidates of a propositional form could always be interpreted in one way or another, and an infinite regress would arise in the process of giving interpretations. For example, someone might judge that he means the function *add 2* by "add 2", and the function *add 2* is explicitly specified by him as the mathematically more precise function $f(x) = x + 2$. However, what function the signs "=" and "+" refer to remain indeterminate, for non-standard interpretations of the sign in question are logically possible and have not been excluded. Appealing to other rules or instructions in propositional forms does not help determine what "+" or "=" means, for it only generates the need for further interpretations to fix the sign's meaning. As a result, whether or not

the person means *add 2* by “add 2” remains undecided. if meaning is conceived as being constituted by something of a propositional form. Or more simply, the person might merely judge that he means *add 2* by “add 2”, without appealing to any other more explicitly formalized rules to explicate his meaning. In such cases, it still remains undecided what function the person really means by “add 2”. For there may still exist a gap between a person’s judging that he means *add 2* by “add 2” and his really meaning *add 2*. Imagine that the person, under cognitively ideal conditions, consistently gives some wild results, such as 2, 5, 167, 38, ..., and meanwhile judges that he means *add 2* by “add 2”. Our intuition would be that either the person does not know what the function *add 2* means or the person means by “add 2” some non-standard function other than *add 2*. In either case, we would simply disregard the person’s judgment about what he means by a term. What this imaginary case shows is that the propositional content contained in the “that clause” of a judgment is not sufficient to constitute semantic knowledge. Semantic knowledge, if it exists, has to be constituted by something else.

What is this something else, which is not a judgment of propositional form, that could constitute semantic knowledge? It appears that particular performances of a person, though they might not be sufficient, are necessary constituents of semantic knowledge. As was shown in the above “add 2” case, a person’s actual performances, such as giving the responses, 2, 5, 167, 38, ..., or the responses, 2, 4, 6, 8, ..., are indispensable in the determination of which function he means by “add 2”.

Wittgenstein’s discussion (e.g., PI 208-211) of one person’s (let’s call him “A”) trying to teach or communicate with another person (let’s call him “B”) what he means by a term such as “add 2” seems to suggest exactly this point. The point in question is that A can

convey what he means by “add 2” to B only by showing B those particular performances A is disposed to give. Those particular performances given by A may not be sufficient to exhibit or convey to B what A means by “add 2”, but beyond them, nothing seems capable of doing the trick of getting A himself understood.

This reflection, that semantic knowledge cannot consist in judgments of a propositional form but must be exhibited in particular performances, helps reveal a way by which Wright’s judgment-dependence account might evade Boghossian’s criticism of circularity. The way out of Boghossian’s attack is to notice that there is an ambiguity inherent in Wright’s judgment-dependence account regarding how to conceive a best judgment made about one’s own meaning (or intention). Consider the provisional equation (M*) that involves the concept of *add 2*. There are two ways in which an agent’s judgment about what he meant by “add 2” can be expressed:

- a) I judge that I mean(t) *add 2* by “add 2”.
- b) I judge that I mean(t) *add 2* by “add 2”, namely, that 14 is the correct number when the last numerical input is 12.

The first part of (b) consists of a judgment that is the same as the judgment expressed in (a). The only difference between (a) and (b) is that (b) has a “namely” clause which describes another judgment (we may thus say that (b) comprises two judgments).

Boghossian takes the judgment in (a) to fix the meaning of one’s term “add 2”. The problem with this interpretation is that, Boghossian claims, the content of the judgment in (a) has to be presupposed in order for us to know what the agent means by “add 2”. For example, the content expressed in the avowal (a) is typically presupposed as that *I mean add 2 by “add 2”*, rather than *that I mean quadd 2 by “add 2”*. In contrast, we could

take the “namely” clause that appears in the second part of (b) as expressing a judgment that serves to fix or specify the content of the judgment in the first part of (b). As a result, the agent’s judging that the number 14 falls under the function *add 2* informs us as to which function the agent really meant by the term “add 2”.

Thus, by suitably distinguishing between (a) and (b), which were originally conflated in Wright’s judgment-dependence account, we have an alternative interpretation of Wright’s account other than Boghossian’s. In this disambiguating version of Wright’s account (let’s call it “Wright’s modified account”), it is an agent’s best judgments about his own particular performances, rather than his best judgments about what he means, that determine what he means by a term such as “add 2”. There is an interpretation regress problem in the case where best judgments about what one means by a word such as “add 2” constitute one’s meaning. In contrast, no problem of interpretation regress arises as to what one means by judging that a particular performance is correct, e.g. judging that 14 comes after 12. In Wright’s modified account, it is the best judgments about one’s own particular performances that collectively determine what one means by a term.

In this way, Wright’s judgment-dependence account (the modified version) could avoid Boghossian’s attack. Moreover, it would fit our ordinary practice better. Imagine that an agent makes the following claim under cognitively ideal conditions:

(c) I judge that I mean(t) *add 2* by “add 2”, namely, that 0.38579 is the correct number when the last numerical input is 12.

Ordinarily, when we encounter a person’s giving such a wild judgment as the one expressed in the “namely” clause of (c), we do not simply take the judgment expressed in

the first part of (c) as determining what the agent meant by “add 2”. Rather, we take into serious considerations the judgment expressed in the “namely” clause of (c) in the determination of what the agent means by “add 2”.

It is quite likely that Wright has the modified account in mind when he proposes his judgment-dependence account. Nevertheless, Wright’s modified account encounters problems that are no less serious than the original version at which Boghossian’s attack aims. First of all, Wright’s modified account is highly counter-intuitive. Our intuition about meaning something by a term is that it is not to be identified with either particular performances or judgments about particular performances. Meaning is something that dictates an infinite number of instances that covers a merely finite set of particular performances or judgments about particular performances. Therefore, meaning exists in a way that is independent and which reaches “ahead” of those particular performances or judgments about them. However, in Wright’s modified account, meaning is identified or determined by a best judgment about a particular performance. Or more accurately, it is a person’s past and current judgments about his own particular performances that collectively determine what he has been meaning by a term such as “add 2”. Wright’s modified account, therefore, offers a picture of meaning that conflicts with our commonsense intuitions about meaning.

We may illustrate the counter-intuitive feature of Wright’s modified account in the following way. Consider (d):

- (d) I judge that I mean(t) *add 2* by “add 2”, namely, that the correct number is 13 when the last numerical input is 12.

(d) differs from (b) and (c) in an interesting way. There are two distinct judgments in (d). The first judgment says what function was meant. The second judgment says that 13 comes after 12. However, we would naturally hesitate to think that the second judgment *specifies or identifies* what the content of the first judgment is. Rather, we would normally regard the second judgment as a seemingly *erroneous* judgment, which *fails to conform* to what I used to mean when I said “add 2”. The two judgments in (d) thus, in a sense, fall apart. In contrast, the two judgments in (b) or (c) do not appear to break apart: the second judgment in (b)/(c) seems to conform to what I used to mean, which is expressed by the first judgment. What does the contrast between (d) on the one hand and (b) and (c) on the other show? In my view, it shows that we actually implicitly understand the “namely” clauses which occur in (b), (c), and (d), to be *exemplars* of my past meaning, rather than *determinants* of my past meaning.⁶⁵ This implicit intuition is, unfortunately, undisclosed when we merely consider cases of (b) and (c).

The upshot is then that our intuitive idea is that for one to mean something by a word and for one to make a best judgment about his own particular performance are two distinct things which should not be conflated. The defect in Wright’s modified account is that it confuses the two in question. McDowell also makes the same criticism as follows:

...suppose I form the intention to type a period. If that is my intention, it is settled that only my typing a period will count as executing it. Of course, I am capable of forming that intention only because I am party to the practices that are constitutive of the relevant concepts. But if that is indeed the intention that--thus empowered--I form, nothing more than the intention itself is needed to determine what counts as

⁶⁵ Peter Sullivan appears to make a similar criticism. See his “Problems for a construction of meaning and intention”, *Mind*, April 1994.

conformity to it. Certainly it needs no help from my subsequent judgments...So there is something for my intention to type a period, conceived as determining what counts as conformity to it autonomously and independently of my judgments on the matter, to be: namely, precisely, my intention to type a period...This is common sense, not Platonism.⁶⁶

Although the above passage was only about intention, McDowell's point applies to meaning as well. McDowell argues that it is our commonsense intuition about meaning (and intention) that when an agent grasps the meaning of a term, the meaning of that term sets a standard of correctness and incorrectness for performances at the present and in the future. Such a standard of correctness and incorrectness, in our commonsense view, exists independently, and reaches ahead of one's performances and judgments. According to Wright's modified account, however, the content of one's meaning is not determined until his own judgment sets in. Wright's modified account thus seems too opposed to our ordinary notion of meaning and intention to be acceptable.

Wright's modified account faces another serious problem. Let's grant that it is one's best judgments about his own particular performances in the past and at present that collectively determine what one has been meaning by a term. The problem is, nevertheless, that a person's judgments about his particular performances are fallible. It is crucial, however, that the person's judgments are classified into correct and incorrect ones, so that what he means by a term can be determined. Does Wright offer us a standard of correctness and incorrectness to carry out the task of classification?

⁶⁶ McDowell (1992b), "Intentionality and Interiority", in his recent collection, p. 315.

The answer seems to be no. Wright grants best judgments a “positive presumptive” status, meaning that unless there is evidence to the contrary, best judgments are correct, and hence serve to determine what one means by a term. This treatment of correctness is tantamount to offering no standards of correctness at all. It amounts to merely suggesting that if a judgment seems right, it is (defeasibly) right. But as Wittgenstein famously expresses his worry: “that only means that we can’t talk about right”, the normativity problem to meaning and rule-following that is central to Kripke and Wittgenstein seems to be totally missed by Wright’s modified account.

Objection 3: Wright’s Judgment-Dependence Account Is Not the Only One Available

Thus, Wright’s judgment-dependence account of meaning and intention faces several difficulties. It seems circular, as Boghossian shows. Even if Wright’s account is clarified and modified to the extent that it can avoid Boghossian’s criticism, the account has been shown to be highly counter-intuitive, and incapable of solving the normativity problem. In the face of these difficulties, one may wonder: are we to adopt Wright’s judgment-dependence account in which the notion of meaning and intention is construed as an extension-determining concept? That is, are we to accept the view that one’s meanings and intentions depend on his own judgments? An alternative approach is to view meanings and intentions as dispositional entities, whose constitutions are independent of our judgments, and are thus appropriate objects of empirical theorizing.

Edwards⁶⁷ offers a convincing argument that shows that we are not forced to accept Wright’s account, and that it is open to us to adopt the view that meanings and intentions are dispositional properties up for empirical theorizing.

⁶⁷ Edwards, 1992.

Edwards' strategy centers on tackling the extremal condition, one of the four conditions a provisional equation, such as (I*), (M*), or (R*) has to meet in order for the concept under scrutiny, such as red or meaning or intention, to be an extension-determining concept. The extremal condition states that we can infer that the judgment-dependence account is the best account of the a priori covariance--which is non-trivially specified and meets the independence condition--between the extensions of the concept of, say, intention, and truths about the extensions of the concept of red, *as long as there is no other account of the covariance in question available*. Edwards's argument thus aims at showing that such an alternative account in question is available.

Edwards proves the existence of an alternative account in question as follows. It begins with brute anthropological facts about us which Wright himself acknowledges. We are inclined to ascribe intentional states to ourselves without any ground. These self-ascriptions tend to accord with the theory-generated or inference-based appraisals that others can make of what we do in order to predict and interpret our behavior. Others' intention-ascriptions to us cannot provide satisfying interpretations and predictions of our behavior, unless they accord with our own self-ascriptions. The above facts are common knowledge to the players of our language game. These anthropological facts are contingently true, and hence make our knowledge of a provisional equation like (I*) a posteriori, not a priori. For it is through possessing the empirical knowledge of those contingent anthropological facts that players of our language game understand (I*). But Edwards argues that there is a pragmatic sense in which these anthropological facts make our knowledge of (I*) a priori. Consider the anthropological fact that there is a certain limit to the height that we as humans can jump. In spite of our athletic limitations, we

can conceive in adequate detail of strategies, such as taking high-tech bio-chemical medicines of certain kinds, which help us transcend our jumping limitations. In contrast, according to Edwards, in the case of being a player of our language game, we know from the start that there is no way in which we can conceive in any detail of any strategy by which we may transcend our limitations as players in our language game. For example, we cannot conceive in any detail of a strategy of interpreting and predicting the behavior of other speakers which massively violates (I*). In this light, (I*) is a priori credible, in a pragmatic sense.

Such an alternative explanation of the a priori credibility of (I*) would seem to undercut Wright's account. For it shows that Wright's judgment-dependence account appeals to a metaphysical claim about intention (i.e., the claim that one's intending P is constituted by one's own judgment that he intends P) that is not necessary. Nevertheless, there is a worry that Edwards' alternative explanation may simply *imply* Wright's judgment-dependence account. To respond to Edwards' alternative explanation, Wright could insist that it is because of the metaphysical nature of intention and meaning as such that Edwards' alternative explanation in question works. Unless this worry is dispelled, Wright's judgment-dependence account is not undermined.

Edwards' reply is that the prospect for showing that his alternative account implies Wright's account is faint. The reason is as follows. Wright's metaphysical claim about intention and meaning implies that there is no possible world in which a speaker's best judgments about his own intentions are massively false. On the other hand, central to Edwards' alternative account is the thought that we are unable to *contemplate* in any detail a possible world in which (I*) or (M*) is massively violated when we try to

interpret and predict our fellow speakers. Contemplating a possible world is different, Edwards points out, from *constructing* a possible world. When constructing a possible world, we have the freedom of stipulating a world in which (I*) or (M*) be massively violated, despite the fact that any detail of constructing it is not open to our cognitive capacities of contemplation. As a result, there is a gap between the central thoughts of Edwards' alternative account and of Wright's judgment-dependence account when they explain the substantial and a priori truth of (I*). Insofar as this gap is not bridged, Wright cannot take his metaphysical claim about meaning and intention to be supported or implied by Edwards' alternative account.

Edwards' argument, therefore, shows that an alternative account of the a priori credibility of (I*) or (M*) is available, and thus the inference that Wright's judgment-dependence account is the best account of meaning and intention cannot be legitimately made. Moreover, Edwards's alternative account of (I*) and (M*) does not cast doubt on the possibility that intentions or meanings are entities suitable for empirical theorizing, contrary to Wright's judgment-dependence account. Consequently, it is left open for us to view intentions (and meanings) as theoretical entities, such as dispositions. In the next two chapters, we will closely examine various dispositional accounts of meaning and intention, to see if they can respond to KW's paradox satisfactorily.

Chapter V. The Dispositionalist Solution to The Skeptical Paradox: A More Likely Approach (I)

We have so far discussed the anti-realist and anti-reductionist approaches to KW's skeptical paradox. Although they have added some new insights into the nature of rule-following or linguistic meaning, they failed to provide satisfactory solutions to the skeptical paradox. As we have seen, in chapter III, once the sceptical conclusion is conceded, which is characteristic of anti-realist positions, there is no way to remedy the disastrous consequences of accepting the skeptical conclusion. On the other hand, the anti-reductionist approach, reviewed in chapter IV, is a form of a "straight solution" to the skeptical paradox. A "straight" solution is one which tries to invalidate the skeptical argument by pointing out defects inherent in it.

The major argument made by anti-reductionists, notably McDowell's and Wright's, is that KW's skeptical argument relies on the erroneous assumption that meaning must be reduced to some other fact. In the anti-reductionists' view, meaning facts are real and irreducible, and we come to know them with authority and immediacy. McDowell on the one hand, and Wright on the other, employ very different approaches to the skeptical paradox despite their sharing the common anti-reductionist ground. McDowell argues that a genuine Wittgensteinian approach to meaning and intention combined with a quietist meta-philosophy could dissolve KW's skeptical paradox. However, McDowell's attempt failed. The claim of McDowell's Wittgenstein that norms can exist only in the context of communal practice is not convincing. For our commonsense intuition strongly suggests that an individual competent speaker is sufficient to guarantee linguistic understanding. Moreover, there exists a genuine unresolved puzzle concerning

the exhibition of both first-person accessibility and third-person dispositional projectivity by meanings and intentions that McDowell's Wittgenstein leaves untouched. Wright offers a constructive account, i.e., a judgment-dependence account, that attempts to solve the genuine puzzle just mentioned and at the same time refutes KW's skeptical paradox. However, Wright's account was discounted both on account of internal theoretical difficulties, and for running against some of our fundamental intuitions about the notion of meaning, intention, and rule-following. As a result, the anti-reductionist approach does not seem to be a likely solution to KW's skeptical paradox.

The only option left for responding to the skeptical paradox is the dispositionalist approach. A dispositionalist, like an anti-reductionist, aims to find a straight solution to the skeptical paradox. Unlike an anti-reductionist, however, a dispositionalist does not call into question the reductionist assumption held in KW's skeptical argument. A (typical) dispositionalist confronts KW's skeptic directly, and tries to reduce meaning-constituting facts to dispositional facts⁶⁸. Although the dispositional approach is widely regarded as very unpromising (or even as the least promising) in terms of solving KW's skeptical paradox, still quite a few philosophers consider it as a plausible approach. In this chapter, I will examine several rigorous dispositionalist attempts that try to solve the skeptical paradox.

⁶⁸ This general characterization of the dispositional approach needs some modification, however. Later in this chapter we will see that Yalowitz (2000) explicitly adopted a non-reductive dispositional account, based on considerations of some crucial constraints on meaning and intention other than the normativity constraint imposed by KW. I will examine this non-reductive version of a dispositional account later when we get to Yalowitz's account. Also, Millikan's teleological approach rejects the reduction of meaning to a disposition, even though it is a naturalistic account. I will also discuss Millikan's position later in this chapter.

As I have mentioned in chapter I, the major problems that the dispositional approach faces in solving KW's skeptical paradox are the finitude and the error problems. Let me briefly rephrase these two problems below.

The major thesis of the dispositional account is that meaning something by a word (or intending something) is constituted by one's possessing a certain disposition to apply the word in a certain way. However, a problem raised by KW is that a disposition is finite and hence unable to encompass the infinitely many judgments that are sanctioned by meaning (or a rule or an intention). My disposition, being manifestable only finitely many times in the past due to my finite capacities, is subsumed under numerous rules which are distinct and mutually incompatible, and whose instances diverge right after the end point of my finite performances. How then can I (or we) be *justified* in claiming that one particular rule rather than another of these numerous ones has been followed during my life-time? There appears to be an *under-determination* problem: our finite performances/dispositions under-determine the rule we follow. This is known as the "finitude" problem.

It should be noted that the finitude problem poses a difficulty not only to the dispositional approach, but also to anti-reductionists. Recall that our having knowledge about (or our having justification in) meaning self-attributions is taken for granted by an anti-reductionist. Holding that meanings are real and irreducible, and granting that we do know what we mean by our words, the philosophical task left for an anti-reductionist to do is to show either one of the following two things. Either it has to be shown how we shall be contented with resting on the meaning--"bedrock" without needing to dig below the "bedrock" to seek justification in the "sub-bedrock" terms (this is McDowell's

Wittgensteinian account), or how it has to be shown how the special epistemic characteristics of knowing meaning self-attributions at the “bedrock”-level can be nicely explicated (this is Wright’s Judgment Dependence Account). However, a rule (or meaning) such as the plus rule, has an infinity of instances which are impossible to grasp by a human speaker. How then can an anti-reductionist claim that a competent speaker knows which rule he follows? Thus, the under-determination problem which underlies the finitude problem haunts the anti-reductionist approach as well.

A much more serious problem for the dispositional approach is the error problem. Given that a person could potentially make mistakes in his performances when he tries to follow a rule that he grasps, for example, saying that $2 + 2 = 5$, how do we determine which rule he follows? How do we pick out a set from the totality of the person’s performances in such a way that the rule being followed is not already presupposed (so as to avoid the circularity problem)? How are we *justified* in selecting one pattern of performances rather than another pattern in the determination of the rule that is followed? If no non-circular specification of the correct performances can be offered, we do not seem to have justification to claim, and hence do not know, that we are following one rule rather than another. The consequence is that, as Kripke’s Wittgenstein maintains, there exists no such thing as rule-following (or meaning something by a term or intending something). This is the “error” problem for dispositionalists.

Note that the problem we are dealing with concerns the ontology of *rule-following*, rather than the ontology of *rules*. KW does not doubt that there exist arithmetic rules, and that they dictate an infinity of instances necessarily. What KW calls into question concerns what arithmetic rule a subject uses a sign, such as “+”, to refer to, and KW’s

claim is that no fact of the matter can be found to settle this issue. In the rest of this dissertation, I will be dealing with the ontological problem of rule-following, rather than the ontological problem of rules. However, as I will point out in the last chapter, the two problems are intimately related, and offering a dispositional solution to the former problem will have a lot more to say on the latter problem.

Can dispositionalists solve the finitude and the error problems? This is what we are going to investigate in this and the following chapters. Before we start, however, there is one crucial problem that is peculiar to the dispositional approach. Even if the dispositionalist can solve the error problem and also accommodate the finitude objection, Kripke seems to think that there is still an important normative element in our notion of meaning that is missing in a dispositional account. Unfortunately, the matter regarding what this omitted element involves is presented very obscurely in Kripke's main text (1982), and is either inadequately explicated or entirely ignored in the subsequent literature⁶⁹. Because it is crucial for my subsequent investigation of whether or not a satisfactory dispositional account of meaning can be found, it is important to figure out what normative element a dispositional account must incorporate.

The Missing Element of Normativity in Dispositional Accounts of Meaning

One of the main passages in which Kripke discusses the crucial inadequacy of a dispositional account of meaning is as follows:

Suppose I do mean addition by '+'. What is the relation of this supposition to the question how I will respond to the problem '68 + 57'? The dispositionalist gives a *descriptive* account of this relation: If "+" means addition, then I will answer '125'.

⁶⁹ There are possibly some exceptions: Wright (1984), Goldfarb (1985), Zalabardo (1997), Yallowitz (2000).

But this is not the proper account of the relation, which is *normative*, not *descriptive*. The point is not that, if I meant addition by '+', I *will* answer '125', but that, if I intend to accord with my past meaning of '+', I *should* answer '125'. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be disposed to respond as I *should*, but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is normative, not descriptive.⁷⁰

This passage depicts the quite natural idea that my meaning plus by "+" implies normative truths about me, namely, truths about how it would be correct to use "+," or how I ought to use "+" in order for the usage to be correct. As a result, if I meant plus by "+" in the past and intend to accord with what I meant, I ought to answer "125" to "68 + 57" in order to be correct. The problem for a dispositional account is then the following. I may be disposed to make a mistake, and thus I may answer "124" rather than "125" to "68 + 57", despite the fact that I meant plus by "+" in the past and am currently intending to accord with what I meant in the past. This is, however, not what a crude dispositional account would say. For what I mean by a term is, in this account, to be determined by how I am disposed to apply the term. Hence, my disposition to answer "124" to "68 + 57" only indicates that I meant a non-standard function in the past. The conceptual difficulty for a (crude) dispositional account is thus that it cannot accommodate the possibility of errors. So presented, the normativity objection is the error problem that we discussed earlier in chapter II.

⁷⁰ Kripke (1982), p. 37.

However, Kripke also appears to think that the concept of dispositions is *descriptive* while the concept of meaning is *normative*, and thus that these two things cannot be of the same kind. Kripke seems to think that no matter how a disposition is to be specified, something normative about meaning is essentially lacking in a dispositional account which is descriptive in nature. But what is the missing normative element that is thought not to be captured by a dispositional account capable of solving the error problem? Unless more explication with textual support is given, we cannot help but feel sympathetic with Fodor's (1990) attitude toward KW's normativity requirement that there is no such missing element of normativity to be found that is over and above what is presented in the error problem. Fodor writes:

requiring normativity to be grounded suggests that there is more to demand of a naturalized semantics than that it provide a reduction of such notions as, say, extension. But what could that 'more' amount to? To apply a term to a thing in its extension *is* to apply the term correctly; once you've said what it is that makes the tables the extension of "table"s, there is surely no *further* question about why it's *correct* to apply "table" to a table. It thus seems that if you have a reductive theory of semantic relations, there is no job of grounding normativity left to do.⁷¹

Some philosophers have tried to explicate the missing element of normativity by notions such as commitment or faithfulness. For example, Blackburn writes as follows:

I share Kripke's view that whatever dispositions we succeed in identifying they could at most give us standards for selection of a function which we mean. *They couldn't provide us with an account of what it is to be faithful to a previous rule.* It is

⁷¹ Fodor (1990), p. 135.

just that, unlike Kripke, I do not think dispositions are inadequate to the task of providing standards. Indeed, I think they must be.⁷² (Emphasis mine)

Blackburn reads Kripke as uncovering a crucial element of meaning that is something other than standards of correctness and incorrectness in a disposition. And this crucial element is an agent's *being faithful to* a rule that he followed in the past. The problem, however, is this: what does it mean to say that one is *faithful* to a rule which he followed in the past? Blackburn and others who take the idea of faithfulness or commitment seriously, such as McDowell, offer no further clear explication, however.

A promising place to locate the missing element of normativity in rule-following behavior seems to lie in the notions of justification and guidance. The following quotation from Kripke is suggestive of this reading:

Precisely the fact that our answer to the question of which function I meant is *justificatory* of my present response is ignored in the dispositional account and leads to all its difficulties.⁷³

This quoted passage suggests that there is a close connection between normativity and justification. Any proposed meaning-fact has to offer a justificatory ground for my correctly answering “ $68 + 57 = 125$ ”. The reason that a dispositional account of rule-following fails to provide such a justificatory ground is that dispositional facts themselves are not justificatory of my responses. But why? What does it mean to say that one’s rule-following disposition is not justificatory of his response?

⁷² Blackburn (1984), p. 289-91. Similar expressions are given by Wright (1984), p. 771-2; and by McDowell (1984), p. 329.

⁷³ Kripke (1982), p. 37. See also p. 11.

An attempt to follow this line of reading which addresses the relation between normativity and justification is given by Gampel (1997). Gampel suggests that the missing element of normativity is the capacity, within or outside of a person, to justify or guide a person's use of an expression. For example, consider the meter bar in Paris, which was traditionally used as a standard of measurement. It certainly could be said that physical facts about the meter bar were used to justify various measurement claims. But this was not essential or sufficient. For these physical facts' capacity to justify such claims *depend on our taking* the bar as a standard of measurement. Similarly, the dispositional facts themselves do not have the capacity to justify my using a term in a certain way even under the ideal conditions, for it is not essential or sufficient for such dispositional facts of a physicalistic nature to *serve as a potential guide* to our linguistic behavior. Rather, whether or not these dispositional facts may serve as a potential guide to my/our linguistic behaviors depends on whether or not we can offer an explanation for them. Therefore, dispositions are analogous to physical facts about the meter bar in Paris: dispositions in themselves are not essential or sufficient to serve as a source of normativity. Dispositions are at most contingent facts which could possibly provide a standard of correctness. But they could not by themselves offer a reason or justification of why they could serve as a standard of correctness. Thus, Gampel argues that the capacity to justify is crucial to meaning or rule-following.

It is worth noting that Gampel is neutral about what constitutes the essential justification of a rule-following or meaning behavior: it could be Platonic entities, communal agreements, first-person properties of some sort, or what ever, depending on which theory one adopts. For any account to be recognized as a theory of meaning or

rule-following, according to Gampel, it must either show that the proposed facts about meaning have the requisite essential normative character or explain why the intuitive notion of the requisite normative element of meaning need not be respected by the theory. The dispositional account apparently fails to do either task and is hence not satisfactory.

Gampel's appeal to the notion of justification or guidance to capture the missing element of normativity seems to be on the right track. However, Gampel's depiction remains empty regarding what it is that is missing in a disposition such that a disposition fails to be justificatory of or a guide for my responses.

Zalabardo (1997) offers a very explicit and convincing explication of the missing element of normativity in terms of the notions of justification and guidance. Zalabardo claims that the missing element of normativity involves features of *first-person phenomenology*, which a disposition does not have. The following passage by Kripke is a clue that strongly suggests Zalabardo's construal:

So it does seem that a dispositional account misconceives the skeptic's problem--to find a past fact that justifies my present response. As a candidate for a 'fact' that determines what I mean, it fails to satisfy the basic condition on such a candidate...that it should *tell* me what I ought to do in each new instance.⁷⁴

This quoted passage appears to connect the *justification* requirement of meaning with the *guidance* requirement. That is, for something to constitute my meaning, say, *plus*, it has to justify my answering, say, " $68 + 57 = 125$ ". More crucially, for something to justify my response, it must be able to *guide* me when I respond " $68 + 57 = 125$ ". This guidance requirement on meaning fits our commonsense talk about what meaning's major function

⁷⁴ Kripke (1982), p. 24.

is when we apply a term: the meaning of a term such as “+” is usually cited by a speaker when he tries to *justify* his calculations, and we also ordinarily say that the function of the “+” sign which we grasp *guides* our arithmetic performances. Meaning something by a term is not constituted simply by a speaker’s being disposed to use the term in a correct way. Rather, it involves the speaker’s being *guided* by the meaning he grasps to give correct performances. In other words, logical covariation of correct use and the meaning-constituting fact is not at issue here. What KW is requesting is a speaker’s being in an epistemic state of knowing the fact about correct use. The meaning-constituting fact has to be, in a way, *transparent* to the speaker with regard to offering him information about the correctness of his use of a term⁷⁵.

This first-person guiding aspect of the normativity requirement of meaning appears to be what a disposition cannot possess, for a disposition is considered, by Kripke and many others, to be a deep-seated physiological fact which is apt only for being posited as a hypothesis and being known only inferentially and indirectly. This point was made by Kripke in the following passage:

How does any of this indicate that--now or in the past—‘125’ was an answer *justified* in terms of instructions I gave myself, rather than a mere jack-in-the-box unjustified and arbitrary response? Am I supposed to justify my present belief that I meant addition, not quaddition, and hence should answer ‘125’, in terms of a *hypothesis* about my past dispositions? (Do I record and investigate the past physiology of my brain?)⁷⁶

⁷⁵ See Goldfarb (1985) and Loar (1985) for the same observation about the first-person guiding aspect of the normativity requirement of meaning.

⁷⁶ Kripke (1982), p. 23.

The contrast between our intuitive notion of meaning on the one hand and dispositions on the other hand becomes clearer in the following passage:

I immediately and unhesitatingly calculate '68 + 57' as I do, and the meaning I assign to '+' is supposed to justify this procedure. I do not form tentative hypotheses, wondering what I should do if one hypothesis or another were true.⁷⁷

Kripke's idea in the above passage is that our intuitive notion of meaning is that meaning is *transparent* to us. We know how to go on applying a term non-inferentially and with a fair degree of certainty when we grasp the meaning of the term, so that in the process of applying it, no physiological investigations are involved. The information about what my past disposition was is not directly available to me, because it requires forming a hypothesis and conducting empirical research to test the hypothesis. Meanings and dispositions thus exhibit very different characteristics. As a result, they must be of distinctive kinds. Zalabardo (1997) turns these reflections on the differences between meanings and dispositions into a formulation of the "justification requirement":

(J) The procedures that I use for deciding whether to apply a predicate to an object would only be justified if they involved *conscious engagement* with the facts that determine which objects I should apply each predicate to.⁷⁸

A dispositional account of meaning fails to meet (J), because what I mean by a term is known directly and non-inferentially, while what my disposition is can only be known indirectly through "recording and investigating" my neuro-physiology. I will call the problem of meeting (J) "the first-person epistemic problem."

⁷⁷ Kripke (1982). p. 40.

⁷⁸ Zalabarbo (1997), p. 481.

However, we may question whether or not the first-person epistemic problem, which derives from the *guidance* or *conscious engagement* requirement, is a problem that is worth solving. It is one thing that KW has called our attention to the problem, but it is quite another thing whether or not the problem is a significant one. It might be suggested, for example, that the connection between justification and guidance be divided by adopting an externalist theory of justification, so that one may be able to justify one's disposition to perform in a certain way and yet not have to be guided by the meaning-constituting fact when he performs certain acts.⁷⁹

Nevertheless, to deny the connection between justification and guidance will miss the most fundamental intuitive idea of KW that started the whole discussion of the rule-following paradox. The idea is, meaning or a rule is something that I seem to grasp firmly, and it not only dictates what I ought to do here and now, but also guides my linguistic or arithmetic performance. It is this first-person dimension of meaning and rule-following that KW has eagerly emphasized. Any satisfactory account of meaning, especially the dispositional account, must accommodate this point. Wright's (1989a) reason to reject the dispositional approach was based on the ground that a disposition by its theoretical nature could never accommodate the first-person phenomenology of meaning and rule-following. In particular, Chomsky's theoretical project of explaining human linguistic behaviors in terms of universal generative grammars is construed by Wright as a form of dispositional account. Wright rejects it on the ground that those deep-seated dispositional entities posited in Chomsky's theory cannot constitute meaning or rule-following, for they are incapable of guiding an agent's linguistic behavior at the

⁷⁹ See Miller for this suggestion, in *The Philosophical Quarterly*, 2000, p. 172. See also Martin & Heil (1998), and Coates (1987), in *Mind*, for a similar suggestion.

conscious level. As Wright rightly pointed out, the real challenge presented by KW is how anything can possess the two seemingly opposing features, i.e., first-person accessibility and third-person dispositional theoreticity, which are simultaneously exhibited by meaning and rule-following.⁸⁰

It is by now clear that Kripke's famous but obscure normativity objection to the dispositional account consists of two problems that are often conflated: the error problem and the first-person epistemic problem. Having teased apart the two previously conflated problems helps us see why the dispositional account is often regarded as the least plausible account of meaning. These two problems set two seemingly conflicting constraints on any account of meaning. On the one hand, there is the constraint that conditions of correctness must be non-circularly specified so that the meaning of a term can be determined. On the other hand, there is the constraint that the meaning of a term must be known directly and non-inferentially. The former constraint is an objective matter, whereas the latter is subjective. These two constraints thus seem to pull the dispositional account in two opposite directions and hence are very unlikely to be satisfied simultaneously.

Any satisfactory dispositional account of meaning has to show how the finitude, error, and the first-person epistemic problems can be handled. In what follows, I will review some major dispositionalist attempts and argue that none of them is adequate in terms of solving these three problems.

⁸⁰ As we have seen, Wright takes up the challenge himself and provides a judgment-dependence account. But as I have pointed out, his account is not likely to succeed. Yalowitz (2000) disputes the claim that a disposition cannot accommodate features of self-knowledge. As we will see later in this chapter, he attempts to offer a dispositional account of meaning self-ascriptions.

Shogenji's Initial Dispositional Account

Shogenji (1993) distinguishes two forms of meaning skepticism inherent in KW's skeptical paradox: modest skepticism and radical skepticism (to be discussed shortly). He further proposed an "initial dispositional account" of meaning as an attempt to solve the modest skepticism. As Shogenji himself notices, however, his account cannot solve radical skepticism. In this section, I will examine both forms of skepticism and how Shogenji's dispositional account fares respectively in solving each of them. I will argue that KW's skepticism is surely the radical skepticism, and hence that Shogenji's dispositional account fails to solve KW's skeptical paradox. By seeing why Shogenji's dispositional account fails, I believe that we can come to have a better understanding of some of the persistent difficulties that underlie KW's skeptical challenge.

In Shogenji's characterization, modest skepticism poses a challenge to the defender of the reality of rule-following to formulate a theory of rules which may non-circularly specify which facts are relevant to the determination of the rule being followed and how those facts determine it. For example, the past usage theory is a theory of rules which states that facts about the past usage of a person are relevant to rule-determination, and in so far as a particular rule is consistent with those facts, the particular rule is the rule that is being followed. Therefore, in the past usage theory of rules, a person's past usage is relevant in rule-determination and consistency with it determines the rule. On the other hand, a radical skeptic may grant that such a theory of rules required by a modest skeptic is possible, but challenges how any proposal of such a theory of rules can be justified. Radical skepticism's claim is that we can offer no justifiable reason to accept one theory of rules rather than another.

To fully appreciate the difference between the demands of modest and radical skepticisms, we may start by considering Shogenji's initial dispositional account which was offered as a theory of rules, and then see how it fares in responding to the challenges of each of the two skepticisms.

Shogenji follows Kripke in using addition as the paradigmatic example for discussion. He first decomposes the plus rule into two major components: the basic routine rules, which include adding single-digit numbers, such as $1 + 1 = 2$, $4 + 7 = 11$, and so on, and carrying, and the recursive rule, which dictates applying the basic routine rules repeatedly. Shogenji then proposes treating an agent's disposition to mean plus by "+" as consisting of two corresponding parts: one consists of the initial disposition to follow the basic routine rules of addition; the other is the disposition to follow a recursive rule to repeatedly apply these basic routines to problems under consideration.

The problem is then to specify how the agent's disposition determines the basic routine rules and the recursive rule, given that facts about an agent's disposition are relevant to rule-determination in Shogenji's theory of rules. Shogenji's idea is that by dividing the disposition of an agent into two parts and in turn by specifying how each part determines each of the two corresponding component rules of the plus rule, the error and finitude problems presented in KW's skeptical paradox can be solved separately.

Shogenji begins with the error problem. Remember that the error problem asks how we can specify a disposition in such a way that it may non-circularly determine the rule being followed. Shogenji proposes a theory of rules which makes two claims: 1) a competent adder's initial dispositions to solve addition problems are facts relevant to rule-determination, and 2) the initial disposition of an agent is so specified that only the

dispositions to produce *stable* arithmetical performances determine the basic routine rules. Adders may sometimes give unstable responses because they are drunk, tired, distracted, etc.. In most cases under favorable conditions, however, adders will exhibit stable performances. Hence, in Shogenji's theory, dispositions that produce stable performances constitute one's following the basic routine rules. One interesting characteristic of this theory of rules is that it is open as to what those stable performances an agent produces turn out to be. An agent might produce stable and convergent performances such as answering " $4 + 3 = 1$ ", " $3 + 3 = 8$ ", and so on. According to Shogenji's dispositional account, as long as those performances are stable and convergent, they determine the basic routine rules.

Shogenji's appeal to dispositions which produce stable performances in the determination of the basic routine rules fits our commonsense intuition. A normal person after several years of training normally exhibits a stable pattern of arithmetic performances, such as answering " 4 " 998 times out of 1000 occasions in which he has encountered the problem " $2 + 2 = ?$ " thus far. Our ordinary practice typically would take the person as following the basic routine rule $2 + 2 = 4$, rather than as following another non-standard basic routine rule such as $2 + 2 = 5$ or $2 + 2 = 3$ (suppose the person answered " 5 " and " 3 " in the two remaining occasions out of the totality of 1000), or even a much more bizarre non-standard basic routine rule like $2 + 2 = 3$ at t_1 in context 1, $2 + 2 = 4$ at t_2 in context 2, and $2 + 2 = 5$ at t_3 in context 3 (there are numerous ways to construct a non-standard basic routine rule from performances that include unstable ones). We would ordinarily regard answering " 5 " or " 3 " as the person's making an error.

However, the skeptic would object to Shogenji that his appeal to stable performances in the determination of the basic routine rules is blatantly circular, because it presupposes that it is *stable*, rather than *unstable* performances that constitute basic routines. Why is it that stable, rather than unstable, performances are selected to fix the basic routine rules, the skeptic asks. Unless some justification can be given for the selection of the former, Shogenji's account would be circular. Shogenji answers this charge:

The task is similar to the construction of a formal proof theory in symbolic logic, which is designed to capture semantic relations, but which does not mention any semantic relations. In an analogous way, the concept of normal conditions is designed to deliver the correct routines, but the specification of the conditions need not involve any mention of the correct routines.⁸¹

Shogenji claims that once his theory of rules has specified which facts are relevant to rule-determination (i.e., initial dispositions to perform addition problems), he can go on to freely select a set of dispositions (i.e., only those dispositions that produce stable performances) that determine the basic routine rules. This procedure, Shogenji argues, is as legitimate as one in which we arbitrarily choose certain non-semantic facts to deliver semantic relations in the construction of a formal proof theory. The point is that insofar as no other rules are employed in the task of specifying the relevant facts in rule-determination, the theory of rules which Shogenji proposes, i.e., the initial dispositional account, is not circular.

⁸¹ Shogenji (1993), p. 496.

To make his point clearer, Shogenji claims that it is logically possible to propose an alternative theory of rules which specifies initial dispositions to solve addition problems as relevant to rule-determination, but chooses the set of dispositions that produce *unstable* performances in the determination of the basic routine rules. This alternative theory of rules, Shogenji argues, cannot be accused of circularity either, because it is similar to constructing a formal proof theory, where no semantic relation is mentioned before the theory is established.

The real issue here, however, is this: How can Shogenji be sure that the theory of rules he favored, i.e., the one in which the dispositions to produce stable performances are chosen in the determination of the basic routine rules, is the correct theory of rules rather than the alternative one that we mentioned in the last paragraph? What justifies Shogenji's favoring the standard theory where stable performances are selected over a non-standard one where unstable performances are chosen?

Shogenji promises to solve the error problem. Has he succeeded in doing so? Recall that the key to solving the error problem is to meet the non-circularity requirement. Since Shogenji does not presuppose any basic routine rules before he specifies facts relevant to rule-determination, we shall not charge him with violating the non-circularity condition. However, a second level of epistemological doubt arises. That is, how does Shogenji's account tell us which theory of rules is the right one, the standard one (fixed by stable performances) or non-standard ones (fixed by unstable performances)? Why do we favor the standard theory over non-standard ones? Shogenji's initial dispositional account as a theory of rules by itself does not offer any justification for us to do so.

As a matter of fact, Shogenji sets as his goal to meet the challenge of *modest skepticism*, which, as we have mentioned in the beginning of this section, puts forward a challenge to formulate a semantic theory which may non-circularly specify facts relevant to rule-determination in such a way that the rule being followed is delivered. Shogenji's dispositional account offers a theory of rules by appeal to stable performances in the determination of the basic routine rules, which does not appear to employ other rules. Shogenji's dispositional account, therefore, is adequate in responding to modest skepticism. However, as Shogenji himself admits, his account does not meet the challenge of *radical skepticism*, which questions why a particular theory of rules is chosen, rather than alternative ones. In other words, a radical skeptic may grant that a non-circular semantic formulation is possible, but deny that justification can be given for favoring a standard theory of rules over alternative ones. Shogenji's initial dispositional account by itself does not offer any such justification. Therefore, if Shogenji's account succeeds at all, it works against modest skepticism, but is impotent against radical skepticism. And clearly, Kripke's sceptic is the radical skeptic.

Shogenji's solution to the finitude problem has a similar result in terms of responding to Kripke's radical skepticism. Recall that Shogenji separates the plus rule into two parts: the basic routine rules and the recursive rule. Hence, the challenge to solve the finitude problem, in Shogenji's formulation, is to specify facts relevant to rule-determination without circularity on the basis of a speaker's dispositions to produce adding performances such that the recursive rule can be selected. However, the consideration of something analogous to normal conditions under which correct basic routine rules can be determined by picking out stable performances does not work this

time. Shogenji recognizes that no facts of the matter can be specified without circularity with respect to fixing the recursive rule. To explain why, consider the following case:

If my brain had been stuffed with sufficient extra matter to grasp large enough numbers, and if it were given enough capacity to perform such a large addition, and if my life (in a healthy state) were prolonged enough, then given an addition problem involving two large numbers, m and n . I would have responded with their sum, and not with the result according to some quus-like rule. But how can we have any confidence of this? How in the world can I tell what would happen if my brain were stuffed with extra matter, or if my life were prolonged by some magic elixir?⁸²

Kripke questions how we can determine that, under the condition of a subject's brain being supplemented with extra stuff, the subject will be disposed to apply the basic routines *recursively*, rather than *requisively*. Shogenji argues that any attempt to answer the question Kripke just raised is bound to beg the question. Consider a powerful argument offered by functionalists against the type identity theory of mind and body. According to the functionalist, a psychological state is not type identical with a physical state since the same psychological state can be realized by different physical systems. Further, even in the same physical species, different individuals may realize the same psychological state by different physical states. Thus, psychological states are functional states whose physical realizers may differ among different systems and individuals. Similarly, in the case of the brain supplement proposal, what a 15-year-old needs for giving a correct sum of million-digit numbers may be different from the brain supplement that an 80-year-old needs for giving the same answer. As a result, the supplement has to

⁸² Kripke (1982), p. 27.

be specified functionally, rather than physically, as follows: "The speaker's brain must be supplemented with whatever appropriate materials so that he will give the standard answers to indefinite complex problems."⁸³ To specify the supplement functionally is obviously circular, for it requires a prior determination of the standard answers. This leads Shogenji to conclude that there is no non-circular way of specifying exactly the kind of brain supplement that delivers the correct application of basic routines.

Shogenji in fact gives up the idea of searching for facts that may determine the right application of the recursive rule. He instead proposes a trivial solution: "the correct way of applying the basic routines is their recursive application."⁸⁴ The proposal is, however, obviously circular, for it simply maintains that the recursion rule is the correct rule.

Shogenji argues that in the context of responding to modest skepticism, his account does not beg the question. For as explained earlier, a rule is divided into two parts: the basic routines and the recursion. The basic routines determine which language (English, Chinese, or any other) is used, and the recursion is universal in the sense that it is contained in every language, whether it is English, Chinese, or any other. In the context of discussing modest skepticism, according to Shogenji, the task is to formulate, rather than justify the acceptance of, a theory of rules. We might formulate an alternative theory which has the same set of basic routines but incorporate the *requusive* rule. To decide which one should be favored, the alternative or the standard theory, is a task that belongs to responding to radical skepticism. As far as formulating a theory of rules is concerned, the burden of proof is on the skeptic to show that the theory of rules formulated in the initial dispositional account fails.

⁸³ Shogenji (1993), p. 497-8.

⁸⁴ Shogenji (1993), p. 498.

Shogenji thus formulates a theory of rules that has the basic routine rules and the recursive rule as the two major constituents of the plus rule that is followed. Shogenji claims that the introduction of the recursive rule overcomes the difficulty of the past usage theory in which the logical possibility of the quus rule's being followed cannot be excluded. The finitude problem, according to Shogenji, is then solved because the two component rules uniquely constitute a rule. The problem with Shogenji's account, however, is that it still owes the radical skeptic a justification concerning whether people are disposed to apply the basic routines *recursively* or *requusively*. And Kripke's skeptic would clearly be the radical skeptic.

To sum up, Shogenji's account meets the requirement of modest skepticism by providing a way to formulate a theory of rules that delivers the standard rule without circularity and without having the finitude problem that the past usage theory faces. Nevertheless, Shogenji's dispositional account offered as a theory of rules still leaves room for radical skepticism. The radical skeptic's challenge is that the logical possibility of a person's following some non-standard rule, such as one consisting of the requusive rule and the basic-routine rules that are determined by unstable performances, cannot be ruled out. Justification for favoring one theory of rules over an alternative one is needed to solve radical skepticism. Besides, Shogenji's account entirely ignores the first-person epistemic problem of meaning and rule-following.

Martin and Heil's Realist Dispositional Account

In the discussion of Shogenji's dispositional account, we saw that underlying the error and finitude problems is persistent epistemological doubt about why a particular set of dispositions is selected rather than another (or why a particular theory of rules is

favorable over another). It is because no suitable reason or justification is available to answer those epistemological doubts that KW's skeptical reasoning leads to a metaphysical claim that meaning and rule-following are nonexistent.

Martin & Heil (1998) take a dispositional approach that is very different from Shogenji's. Martin & Heil do not directly confront epistemological questions that underlie the error and finitude problems as Shogenji did. Rather, they emphasize the metaphysical dimension of the issue about meaning and rule-following. More specifically, Martin & Heil offer a realist dispositional approach⁸⁵ in which the idea of a disposition is very different from that held by KW in his discussion about the dispositional account (What these two different views are and how they differ will be discussed shortly). They argue that it is because of the highly inadequate conception of a disposition held by KW that all the major problems of dispositional accounts such as the error and finitude problems arise. Martin & Heil claim that once misconceived ideas about dispositions are removed and replaced with correct ones, it can be shown how rule-following or meaning can be suitably constituted by dispositions. In this section, I will argue that the error and finitude problems remain major obstacles for Martin & Heil's realist dispositional account. Nevertheless, their account provides important insights about the metaphysical characteristics of a disposition. Those insights will prove to be highly valuable for a continued search for a satisfactory dispositional account.

To better introduce Martin & Heil's realist dispositional account, it will be useful to go into some detail about the traditional view⁸⁶ of a disposition held by KW. The

⁸⁵ Martin & Heil, "Rules and Powers", 1998.

⁸⁶ E.g., Ryle (1948), *The Concept of Mind*, ch5; Goodman (1954), *Fact, Fiction, and Forecase*, 2nd edn, p. 39; Quine (1960), *Word and Object*, p. 222-3.

traditional view is one in which disposition-ascribing statements are equated with counterfactuals, which have the following form: “if p were to have happened, q would have happened”, where p is presupposed to be false. Kripke explicates the dispositional account of meaning as follows:

To mean addition by ‘+’ is to be disposed, when asked for any sum ‘x + y’, to give the sum of x and y as the answer (in particular, to say ‘125’ when queried about ‘68 + 57’); to mean quus is to be disposed when queried about any arguments, to respond with their quum. True, my actual thoughts and responses in the past do not differentiate between the plus rule and the quus hypotheses; but, even in the past, there were dispositional facts about me that did make such a differentiation. To say that in fact I meant plus in the past is to say—as surely as was the case!—that had I been queried about ‘68 + 57’, I would have answered ‘125’. By hypothesis I was not in fact asked, but the disposition was present none the less.⁸⁷ (Emphasis mine)

Here, Kripke apparently adopts a counterfactual analysis of statements ascribing semantic dispositions with respect to “+” which is claimed by a dispositionalist to constitute one’s meaning plus by “+”. The strategy of the counterfactual analysis is to replace a meaning-ascribing statement like (m)

(m) S meant plus by “+”

by a counterfactual conditional statement like (c)

(c) If S had been queried about “68 + 57”, he would have answered “125,”

⁸⁷ Kripke (1982), p. 22-3

where (c) uses only two simple indicative statements. As a result, the seemingly problematic or mysterious term “meant’ in (m) is eliminated and explicated by a relationship that is claimed to hold between two plain indicative statements.

For a counterfactual conditional such as (c) to reductively explain a statement about a disposition like (m), the main task is to find out the necessary and sufficient conditions under which a counterfactual conditional such as (c) is true. The task in question is not so straightforward, for a counterfactual conditional like (c) is not truth-functional. A truth-functional compound sentence is one whose truth-value is entirely determined by the truth-values of its components in every possible case. For example, the falsity of “George Bush is the president of China” or “George Bush has an office at Beijing” logically entails the falsity of “George Bush is the president of China and has an office in Beijing”. A conjunction is false when any of its conjuncts are false. However, the counterfactual conditional “If George Bush were the president of China, he would have an office at Beijing” is true, even with the falsity of both components. Other counterfactual conditions with false components are nonetheless false, such as “If George Bush were the president of China, he would ride a donkey traveling to Tibet”. The truth-value of a counterfactual conditional seems to depend, at least partly, on the contents of its components.

It turns out to be very troublesome to identify the truth conditions of a counterfactual conditional. Consider (m) and (c) again. The indicative statement (m) reporting a semantic disposition with respect to “+” is sometimes true while the counterfactual conditional (c) is false, and *vice versa*. For example, a subject may have meant *plus* by “+”, but could have failed to answer “125” if he had been queried “68 +

57", for he might have been drunk or too tired at the time of the query. Conversely, a subject may have meant something else by "+" such as *quus*, but somehow had answered "125" when being queried "68 + 57", for he could have been very drunk or tired then when queried. The former case shows that the counterfactual (c) is not necessary for the indicative statement (m), for (m) can be true while (c) is false. The latter case shows that the counterfactual (c) is not sufficient for the indicative statement (m), for (m) may be false when (c) is true. If a set of counterfactuals that includes (c) is intended to be a reductive explanation of the disposition statement (m), then the attempt apparently fails, because the set of counterfactuals is neither necessary nor sufficient for (m).

A typical defense of the counterfactual analysis of disposition against the above problem is to add auxiliary conditions in the antecedent of a counterfactual conditional such that the consequent can be derived. For example, one might modify (c) by (c') as follows:

(c') If S had neither been tired nor drunk, and had been queried about "68 + 57", he would have answered "125"

The problem, however, is that the auxiliary conditions in (c') are not complete. For example, when one was distracted, or had brain damage, one would not be able to answer "125" when queried "68 + 57". We soon find out that a competent specification of the list of the auxiliary conditions is extremely hard to give. This difficulty may tempt us to try the following formulation:

(c'') If nothing had interfered, and if S had been queried about "68 + 57", he would have answered "125"

The formulation (c''), apparently, simply trivializes the counterfactual conditional. For the insertion of "had nothing interfered" in (c'') simply amounts to ensuring the correctness of the original counterfactual (c) and hence the correctness of (m). In other words, (c'') is tantamount to (c'''):

(c''') If S had been queried about "68 + 57", he would have answered
 "125", unless he did not.

As a result, the prospect of explicating a disposition-ascribing sentence in terms of counterfactuals seems dim. If we understand a disposition-ascribing sentence in terms of a counterfactual conditional as Kripke does, then we appear to have very good reason to follow Kripke in rejecting a dispositional account of meaning.

Martin & Heil (1998) are fully aware of the difficulties that surround a counterfactual construal of disposition, and they agree that counterfactual analysis is highly inadequate. They propose that we replace the problematic counterfactual analysis of dispositions with a realist position. The major reason to take such a position is that by so doing, we can make better sense of the world—including objects in general and mental phenomena in particular. In a realist view, a disposition is a robust property inherent in any object (organic or inorganic, intelligent or mechanical). For example, salt dissolves in water because it has a disposition to dissolve in water, a glass breaks because it has a disposition to break, humans follow rules because they have dispositions to follow rules, and so on. The great explanatory power that a realist dispositional account of a wide range of natural phenomena has, according to Martin & Heil, is due to the fact that "dispositionality is a fundamental, irreducible, and ineliminable component of nature."⁸⁸

⁸⁸ Martin & Heil (1998), p. 295.

Three important features of a disposition construed as robust in Martin & Heil's (1998) realist theory will be reviewed below, because they are relevant to Kripke's arguments against the dispositional account of rule-following.

The first important feature of a disposition is that a disposition occupies "a region of a holistic web" with other reciprocal disposition partners for possible mutual manifestations. For example, just as salt tends to dissolve in water, water tends to break up salt molecules. There is a potentially infinite range in which a disposition may come to have reciprocal relations with other disposition partners. For instance, salt can also melt when it is heated by fire in a spoon. A disposition itself may also be a disposition partner of alternative dispositions for alternative mutual manifestations. For example, a human body has a disposition to react to salty material in a certain way. Consequently, a glass of water containing dissolved salt can cause certain physiological effect on a human body when drunk. The physiological effects can have further effects on either the body or the world when other potential dispositional partners enter into causal contact with the body.

What this "holistic-web" picture shows is that a disposition possesses a feature of "projectivity". That is, a disposition is "projective" for its alternative reciprocal partners for alternative manifestations. However, the way in which a disposition is projected is not unconstrained. A disposition is selective or directed to its disposition partners in terms of how it manifests itself and how it allows other dispositions to manifest themselves. For instance, if salt is put into a glass of sand, it will normally not dissolve. That is, the disposition of solubility that salt possesses does not manifest itself when its

reciprocal partner is sand. These features of “directedness” and “selectivity” are responsible for a disposition’s having a “content” and “specificity”.

The second important feature of a disposition is that a disposition has an infinite characteristic. There are at least two senses of infinity that can be found in a disposition⁸⁹. Both of them are derived from the ways by which a disposition projects itself. Consider quarks of some sort, say, type-A quarks. Type-A quarks possess a disposition to be annihilated by quarks of a complementary sort, say, type-B quarks. By virtue of having this disposition, type-A quarks may come to be annihilated by an arbitrarily large number of type-B quarks that might happen to be in the vicinity of type-A quarks. As a result, type-A quarks have a disposition for an infinite number of possible manifestations because each type-A quark could potentially get into contact with any one of those infinite number of B-type quarks. However, this does not mean that a type-A quark’s dispositions will manifest themselves on an infinite number of occasions. As a matter of fact, a type-A quark’s disposition to manifest itself when it encounters a type-B quark on a particular occasion will preclude its manifesting itself on any other occasions. for the type-A quark may turn into other kinds of sub-quarks upon the encounter with a type-B quark.

There is another sense in which type-A quarks have a disposition of an infinite nature. A type-A quark may come into contact with a type-B quark at an infinite number of times. Thus, a type-A quark’s disposition to react may manifest itself at t_1 , t_2 , t_3 , ... and so on, depending on when it encounters a type-B quark. And again, a type-A quark’s actual manifestation at a particular instant when it runs into a type-B quark precludes its

⁸⁹ See Martin & Heil, 1998.

manifestation at other times. But this does not prevent us from conceiving of a type-A quark as having an infinite disposition. To say that a disposition is finite, Martin & Heil (1998) argue, is only to confuse a disposition with its manifestation. A disposition of some sort possessed by an object, e.g., the solubility of salt, can manifest itself only finitely many times, due to the finite capacities of the object. However, the disposition is “there-and-ready”, and its projectivity extends indefinitely.

The third important feature of a disposition is that certain dispositions of higher-level organisms are capable of “cue manifestations”, which appear to account for the economy of their higher cognitive functioning. This third feature differs from the above two features in that the first two features are general features of a disposition of any objects while the third feature is possessed only by organisms of a higher sort. To understand what a “cue-manifestation” of a disposition is, consider our experiences of crossing a busy intersection. Suppose that we are capable of crossing a busy intersection. It should be unobjectionable that we have a disposition of some sort that enables us to do so. When this disposition manifests itself, we manage to cross a busy intersection. An interesting feature of our experiences of crossing a busy intersection that is not yet in this picture is this: when we are poised to cross a busy intersection, we may “feel” that we can or cannot cross the intersection. According to Martin & Heil, this feeling is a “cue” manifestation of an underlying disposition to cross the street, in contrast to a “typifying” manifestation of the same underlying disposition which constitutes our crossing a busy intersection. Cue manifestations in the form of “having a certain feeling of something” serve to inform an agent of the readiness of an underlying disposition’s typifying manifestation.

With these abovementioned features of a disposition at hand, we are in a position to see how Martin & Heil's view of meaning or rule-following is, in a metaphysical sense, constituted by dispositions and how KW's major objections to the dispositional account of meaning might be answered.

In Martin & Heil's account of meaning, for one to mean something by a term is for one to possess a disposition of a particular kind. This disposition, here and now, is projective for endless manifestations with alternative disposition partners. The projectivity of a disposition can be thought of as constituting a complex web. Dispositions differ when their complex webs differ. Thus, for an agent to mean plus by "+" is for him to possess a disposition with a complex web p, whereas for an agent to mean quus by "+" is for him to possess a disposition with another complex web q. The major difference between the structure of p and q is that when presented with the question " $68 + 57$ ", a possessor of p will answer "125" while a possessor of q will answer "5" when no intervening factors are present.

Solution to the error problem The error problem is solved as follows by Martin & Heil. An addition error made by a subject is viewed and explained as the failure of a disposition of a certain structured web to manifest itself because of some blocking factors, such as the subject's drinking too much alcohol, caused by some inhibitory disposition partners (there is presumably a rich neurophysiological story to be told about how alcohol affects human brains and behaviors). A person who follows the plus rule may give the same answer "5" to " $68 + 57$ " as another person does who follows the quus rule. However, it is not the performances that distinguish these two adders in terms of which rules they follow. Rather, it is the fine-grained details of their distinct

dispositional make-ups that do the trick. (But how can we distinguish two dispositions with distinct web-structures? This is a thorny epistemological problem that I will raise for Martin and Heil shortly). In Martin & Heil's view, both making a careful distinction between a disposition and its manifestations and having a correct metaphysical understanding of the complex web-structure of a disposition enable us to see how and why the error problem is misguided. Anyone who takes the error problem seriously shows, in Martin & Heil's view, a lack of deep appreciation for the metaphysical nature of a "real" disposition.

Solution to the finitude problem. The finitude problem, in Martin & Heil's view, also arises from a misunderstanding of the metaphysical nature of a disposition and involves a conflation of a disposition with its manifestation. The correct understanding of a disposition is that it has a bounded (because of its "directedness" and "selectivity") infinity (because it projects to anywhere and any time). The way in which a disposition is bounded explains and specifies its content. On the other hand, the inherent infinite nature of a disposition explains the "and so on" procedure of repeatedly applying basic routines. Martin & Heil suggest that recursion is built into human rule-following dispositions in a similar way to which recursion is built into a substance's disposition at the most basic level of nature. Just like a disposition possessed by a type-A quark to causally interact with a type-B quark can project itself to any time in the future and to come into contact with any of an unimaginably large number of type-B quarks for a manifestation, a person possessing a disposition to apply certain basic routines can in principle manifest the disposition at any time in the future when presented with any addition problem. The incapacity of a finite agent to perform an addition problem

involving very large numbers that he might never be able to finish calculating during his lifespan is not regarded as a limitation in the agent's disposition to follow the addition rule. Rather, it is regarded as a limitation in the manifestations of the disposition to follow the plus rule due to the agent's limited organic capacities (for example, the agent may become tired, distracted, or may even die before the super-long calculation is done) or limited environmental supplies (the pen may dry out or paper may run out, etc.). Thus, once the distinction is made between a disposition and its manifestations which are limited and finite, it becomes clear that a disposition is infinite in nature. The infinite nature of a disposition can then account for the recursive function which is an important constituent of the plus rule.

Solution to the first-person epistemic problem. To solve the first-person epistemic problem, the notion of "cue" manifestations is of great importance. Our experiences of following a rule are usually accompanied by a feeling of certainty that we are following the rule correctly or incorrectly. This feeling of certainty is a cue manifestation of the underlying disposition that constitutes rule-following or meaning. According to Martin & Heil, a cue manifestation provides an agent with a reliable, though not infallible, indication of the presence of a particular dispositional condition. This explains why while we may not know of the exact fine-grained structure of the complex web of our underlying disposition which constitutes our following a rule, we may know that we are in a particular dispositional state on the basis of some cue manifestations. Cue manifestations enable an agent to monitor, though fallibly, his own actions as he solves a problem, and result in his having a sense that his performance is

correct (correct in the sense that it is a performance manifested by a particular disposition without blocking or intervening factors).

The realist dispositional account of meaning provided by Martin & Heil appears to be very promising. The strength of their account lies in its realist view of dispositions which allows us to regard rule-following as constituted by an inherent property or capacity of the agent that is not identical to any particular rule-following behavior or mental occurrence. Indeed, because of its projective power, a disposition to follow a rule can exist “ahead” of any current semantic behavior and mental. Moreover, cue manifestations of a disposition provide useful conceptual resources for depicting the phenomenological dimension of rule-following and hence for solving the “first-person epistemic” problem.

The problem with Martin & Heil’s account, however, is that it seems unable to resolve the error and finitude problems completely. Martin & Heil solve the error problem by arguing that an error occurs because some blocking factors prevent a particular disposition from manifesting itself. However, how are we to determine whether an act is erroneous or not? To do so, we have to know whether the act is performed under a condition where no intervening factor occurs. How then are we to identify such a condition without having to presuppose what to count as intervening factors, and hence presuppose which rule is followed? Let me elaborate this point a little further.

Suppose an agent, say, Tom, has a real disposition to follow the plus rule. We can imagine, however, that there is an alien performing some trick on Tom’s brain through some sort of advanced remote control technology such that Tom’s disposition to follow

the plus rule fails to manifest itself. Though Tom has a real disposition to follow the plus rule, he gives the answer "5" to " $68 + 57$ " because of the alien's interfering. We can further imagine that the alien's remote manipulations make Tom feel certain that he is following the quus rule, even though he embodies the real disposition to follow the plus rule. According to Martin & Heil's account, Tom is following the plus rule, for his underlying disposition is to follow the plus rule. It's just that the alien intervened and caused the original disposition to manifest itself along another line. Given all the mental and behavioral evidences, we have to conclude that Tom follows the quus rule, unless the alien's conduct is detected by us. Without knowing whether an alien intervenes or not (and it seems plausible that we may not know it at all), we cannot decide without circularity which rule is followed by Tom. In short, from a sceptic's point of view, there is no observable difference between quus-Tom and plus-Tom-with-alien. As a result, Martin & Heil's account cannot avoid assuming that an agent is following a particular rule and that the agent possesses a particular dispositions to follow the rule. They thus fail to offer a non-circular way of specifying which rule is followed.

A skeptical worry arises as well in Martin & Heil's treatment of the finitude problem. In KW's original presentation of the problem in question, a rule, having infinite instances, outstrips humans' limited dispositional capacities. Dispositions and rule-following, therefore, cannot be of the same kind. More strongly, dispositions cannot possibly constitute rule-following. In Martin & Heil's account, the finitude problem arises from a misconception. A disposition is infinite, and has to be carefully separated from its manifestations, which are only finite. This view of a disposition is, in my view, insightful and correct. However, from a skeptic's viewpoint, this account is not

satisfactory. This can be seen more clearly using Shogenji's framework of rule-following in which the plus rule consists of basic-routine rules and the recursive rule. Suppose that an agent's disposition to follow the basic-routine rules are unproblematically specified (e.g., by an appeal to stable performances), and that we also agree that a disposition is infinite. A skeptic may still question, will the basic-routine rules be applied *recursively* or *requusively*? There could be many different ways in which a disposition of a certain structure repeatedly manifests itself infinitely: it could be repeatedly manifested in a recursive manner, a requusive manner, as well as many other bizarre manners. Martin & Heil seem to simply presume that an agent will be disposed to apply the basic routines recursively, rather than requusively. Justification remains to be offered that an infinite disposition of an agent embodies a recursive structure, rather than a requusive one.

One of the most important points made in Martin & Heil's account is concerned with the role played by the first-person characteristics in meaning or rule-following. In their view, first-person features involved in rule-following behavior cannot unambiguously determine which rule is being followed. However, they play an important role: these first-person features serve as a reliable, though fallible, indicator that informs the agent whether or not his underlying dispositions constituting the rule are manifested successfully (without being blocked or inhibited). The problem with Martin & Heil's account, however, is that it does not offer an adequate picture of the content and structure of this phenomenological aspect of rule-following. In particular, the following question is crucial for illustrating the phenomenology: how are the conscious thoughts and/or feelings of an agent capable of tracking or monitoring their underlying dispositions; in what way does an agent come to know (or be conscious of) the meaning-

constituting fact, i.e., an underlying disposition? Moreover, when an underlying disposition to follow a rule manifests itself, a salient occurrence in our introspective consciousness is the thought of a verbalized rule such as $f(x) = x + 2$, or a judgment such as $f(x) = 6$ when 4 is applied to x . Our commonsense intuition seems to be that either one of these mental occurrences (especially the verbalized rule) is sufficient to fix the rule that is being followed. But as Wittgenstein has famously and convincingly shown us, none of them is able to determine the rule being followed because of their leading to an interpretation regress. This thus appears to be paradoxical. We would hence like an explication of what roles a verbalized rule and a judgment about a particular instance of the rule play if they do not serve to determine the rule being followed. We would also like to know their relations to the underlying disposition that is claimed to constitute the rule being followed. Failing to provide an adequate answer to these questions, Martin & Heil's account does not satisfactorily solve the "first-person epistemic" problem.

Yalowitz's Expressivist Dispositional Account

Yalowitz (2000) points out that the most critical difficulty that any dispositional attempt faces lies in meeting two joint constraints, namely, how it can accommodate both the constitutive feature of meaning and content (i.e., how is the correctness condition of a term non-circularly specified) and the epistemic feature of direct self-ascription (i.e., how is the meaning-constituting fact, a disposition, known directly and non-inferentially?). These two features appear to pull in different directions, namely, objective and subjective, and thus render a dispositional solution to Kripke's skeptical paradox hopeless. Yalowitz attempts to offer a dispositional account that is directed at showing how these two seemingly opposing features can be accommodated in a disposition.

Yalowitz's primary task is hence tantamount to solving the error problem and then the first-person epistemic problem.⁹⁰ In addition, he also offers a solution to the finitude problem.

Yalowitz's solution to the error problem is that this problem arises from Kripke's mistakenly questioning the presumption of first-person authority. To assume first-person authority is to grant that we have knowledge of our intentions or meaning self-ascriptions. Yalowitz treats the status of our meaning and intention self-ascriptions as Wright does, namely, self-ascriptions have "positive-presumptive"⁹¹ status. They are deemed correct, unless there is evidence to the contrary. If an agent's first-person authority is granted, as Yalowitz thinks it should be, then an objective condition of correct use can be given, namely, by the agent's own verdicts about his performances⁹². As to the popularly held view that dispositional accounts cannot explain the fact that we seem to know our minds directly, Yalowitz points out that this is an unfortunate misperception largely due to Ryle's traditional view of dispositions in which an inferential epistemology is employed. In Ryle's view, a statement ascribing a disposition is equivalent to a set of infinitely many counterfactuals. As a result, to know what

⁹⁰ Yalowitz's account differs from Wright's judgment-dependence account in that the former tries to offer a plausible account of self-knowledge within a dispositional framework.

⁹¹ Wright (1989a), p. 251.

⁹² In Yalowitz (2001), "Individualism, normativity, and the epistemology of understanding", *Philosophical studies* 102, he modifies his claim that it is one's verdicts or judgments about his own performances that enjoy positive presumptive status. Yalowitz holds instead that it is particular inclinations to linguistic use that "take on the most basic positive-presumptive status needed in order to account for authority." (p. 59). The reason for the modification is that "the notion of 'judgment' runs the risk of overintellectualizing the performances of rule-following, and thus suggesting a regress-inviting model." (p. 59) I find Yalowitz's modification here appropriate and correct. It fits our ordinary practice better, too.

disposition an agent possesses, an inferential process is required in which every counterfactual in the set has to be tested and examined. Yalowitz argues that such an inferential epistemology employed in traditional theories of dispositions can be avoided. He offers a dispositional account which instead adopts an “expressivist” epistemology that allows a disposition to be known directly by its possessor. I will show that Yalowitz’s account is on the right track and a much improved one in terms of accounting for the first-person dimension of meaning which Martin & Heil (1998) helped illuminate but only gave a marginal role. However, Yalowitz’s account does not satisfactorily respond to the error problem.

The first thing to note about Yalowitz’s dispositional account is that while it attempts to avoid circularity in specifying the correctness condition, it rejects the reductionism widely thought to be intrinsic to a dispositional account. It is commonly held, by Fodor (1987, p. 98), Boghossian (1989, p. 509-16, 527-40), and others, that a reductive stricture is imposed on dispositional accounts of meaning by Kripke’s skeptic. However, in Kripke’s, as well as Ryle’s, own discussions, dispositions are characterized intentionally in terms of dispositions to act, to apply or use thought-tokens or linguistic expressions, and so on, under certain conditions. Yalowitz points out that in Kripke’s skeptical challenge, one is allowed the knowledge that one is attempting or intentionally following some rule or other, and that the issue Kripke was concerned with is how to determine the content of those intentional states. Yalowitz thus regards Kripke’s skeptical problem not as requiring a thorough behavioristic or other naturalistic reduction of intentional phenomena, but as raising doubt about the possibility that our dispositions to apply a term intentionally can indeed fix the content of the term. It would be

worthwhile to decide whether or not content can be fixed by dispositions which are characterized intentionally, because this is exactly what Kripke denied.

Yalowitz has an independent reason for not respecting a reductionist stricture on the dispositional account. This is, that other crucial characteristics of the mind cannot be ignored or sacrificed when providing an account of meaning or content. One such crucial component of mind that Yalowitz considers is its rational character. If the rationality of mind cannot be reduced to any naturalistic properties, reductive dispositional accounts would simply fail. This means that even if a reductive dispositional account could solve the problem of content presented by KW, the reductive dispositional account would not be a satisfactory general account of mind and meaning. As a result, Yalowitz proposes a non-reductive dispositional account to solve KW's skeptical paradox. His reason for adopting a non-reductive dispositional account seems to me to be fair and right.

Yalowitz thus formulates KW's skeptical questioning as follows⁹³:

Given the presumptions that 1) we authoritatively know our own (mental as well as linguistic) contents, that 2) we know when we intentionally apply an expression, and 3) we are, more generally, rational animals, by virtue of what does a linguistic expression or a thought have the content that it does?

In the above formulation of KW's skeptical challenge, the presumption 2) was derived from KW's allowing a disposition to be characterized intentionally, and the presumption 3) was derived from the Davidsonian principle of charity. It is not clear, though, where the presumption 1) came from. KW did not presume 1) in his formulation of the skeptical problem. On the contrary, Kripke rejected it. Remember that the

⁹³ Yalowitz (2000), p. 254.

skeptical conclusion Kripke derived states that we do not know the meanings or our linguistic terms and the contents of our thoughts. How then could Yalowitz reformulate KW's problem as accepting presumption 1)?

As a matter of fact, 1) is an assumption that Yalowitz himself accepts in his dispositional account. Holding the assumption that acknowledges first-person authority about what we mean and what we think is a key move in Yalowitz's response to the error problem (or the problem of providing correctness conditions). Yalowitz follows Wright in holding that self-ascriptions are "positive-presumptive": without having evidence to the contrary, self-ascriptions are to be treated as veridical. As a result, in Yalowitz's version of the dispositional account, the presumption of correctness is attached to our semantic inclinations and dispositions.

Yalowitz formulates his dispositional account as follows:

(y1) How one should use a term is how one's semantic disposition
normally lead one to use that expression.⁹⁴

He further formulates the normality condition, with the presumption of authority attached, as follows:

(y2) An expression 'E' is correctly applied to condition c by speaker S on a particular occasion if and only if S is inclined to apply 'E' intentionally to c on that occasion, and is sincere in doing so, and would reflectively endorse that inclination, and has suffered no recent or persistent physical or psychological trauma...and so on.⁹⁵

⁹⁴ Yalowitz (2000), p. 257.

⁹⁵ Yalowitz (2000), p. 259.

Yalowitz's dispositional account formulated in (y1) and (y2) is distinctive in two aspects. One is that it accepts the commonsense presumption of first-person authority of one's semantic self-ascriptions. This would enable one to account for the subjective pull of the tension noted earlier, for in every particular use of a term, what one is disposed, *ceteris paribus*, to think of as right is right (a more detailed "expressivist" epistemology will be offered later to describe the subjective dimension of meaning). The other aspect is that it emphasizes the *ceteris paribus* nature of a disposition. That is, the presumption of correctness attaching to semantic inclinations can be overridden whenever conditions under which the disposition manifests itself are not normal. This enables Yalowitz to account for the objective pull of the tension, i.e., the correctness conditions, because there is no sheer equation of a particular use with a correct use.

According to Yalowitz's account, the presumption that a particular linguistic use is correct would be overridden when there is a contrary evidence. For example, a person may apply the term "cow" to a horse very far away from him. However, were the person to take a closer look at the horse, he would apply the term "horse" rather than "cow" to the animal. The presumption of correctness attached to the original use of the term "cow" would thus be overridden under this situation. Nevertheless, Yalowitz points out that overriding the correctness of the original use is not equivalent to deeming the original use as incorrect. It might be the case that the animal being encountered was after all a cow wearing a horse's outfit in such a perfect manner that it looked exactly like a horse from outside. So, the person's original use of the term "cow" turned out to be correct. In light of the openendedness of possible overriding situations, Yalowitz treats

every particular use of a term, be it an initial use, or a later use correcting the initial one, as enjoying a positive presumption.

This treatment mentioned above in Yalowitz's account leads us to realize how heavily Yalowitz's solution to the error problem (or the problem of specifying the correctness condition) relies on the presumption of first-person authority. In particular, no presence or absence of any overriding factor on a particular occasion is a decisive indicator of whether a particular use is correct or not. Every particular use is presumptively correct, and the first-person authority is simply granted.

My objection to Yalowitz's account is that it simply dismisses the error problem instead of solving it. Recall that what the error problem presents us is with the problem of how we can classify the performances of an agent as correct and incorrect. As was shown in the cases Yalowitz considered above, no particular use, whether it is an initial or a later correcting use, seems to be decisively correct. The problem is exactly how we may offer a standard by which we can determine the correctness and incorrectness of a particular use. The challenge for a dispositionalist is then to provide such a standard. Instead of confronting the skeptic's challenge directly, Yalowitz simply grants the correctness of particular use, based on our first-person verdicts. The question that arises consequently is, why must we accept the presumption of first-person correctness? Kripke's sceptic would simply pounce on the presumption of first-person correctness and asks why it holds.

Yalowitz offers a reason for accepting the assumption in question. He argues that KW's skeptical argument would severely limit the interest of the normativity objection (i.e., the problem of correctness conditions) if it questioned the presumption of first-

person correctness, because “it is as fundamental to our commonsense picture as normativity.”⁹⁶ The problem with this reasoning, however, is that it underestimates the seriousness and the nature of the error problem. It is precisely because of the difficulty of the error problem raised by Kripke’s skeptic that our first-person authority is undermined. It is precisely because we cannot ever be sure which use of mine is correct and hence which meaning the use expresses that Kripke’s skeptic suggested that first-person authority be given up. Yalowitz’s treatment of the error problem thus seems to me unsatisfactory.

Let’s now move to Yalowitz’s treatment of the “first-person epistemic” problem. The problem in question concerns the epistemological dimension of the rule-following problem: how can the meaning-constituting fact, a disposition in our current case, be known in a direct and non-inferential manner? Meaning is transparent to us; we come to know what we mean by our words in an immediate way. A disposition, in contrast, is a theoretical entity⁹⁷, which can only be known on the basis of inference. How then can a disposition constitute meaning?

To solve this problem, Yalowitz directs our attention to the question of *what kind of entity* possesses a disposition. Paradigmatic cases of the *possessor* of dispositions usually center on physical objects in general. For example, salt possesses a disposition to dissolve in water; a glass has a disposition to break, etc. In these paradigmatic cases, the individuation conditions of a disposition can indeed only be known from a third-person, theorist’s perspective. However, when *persons* are the possessors of dispositions, especially semantic dispositions, it is a quite different matter. Being the subject of a

⁹⁶ Yalowitz (2000), p. 256.

⁹⁷ Kripke (1982), p. 40; Wright (1987), p. 394.

semantic disposition, one comes to know his own disposition directly by virtue of the *expression* of his semantic disposition, rather than by observation or inference. Yalowitz argues that those who object to dispositional accounts because of the first-person epistemic problem ignore the possibility that "...the theorist and the self-ascriber have different modes of access to the same underlying dispositional state..."⁹⁸ In this version of the dispositional account, our privileged access to semantic-constituting dispositions is explained by the fact that we are the *possessors* of these dispositions and we *express* them in the acts of self-ascription. Meanwhile, the individuation conditions of the same semantic disposition that is being expressed can be provided by suitable empirical investigations conducted by a theorist. In short, there is no major conceptual obstacle in our construing a semantic disposition as capable of exhibiting both third-person dispositional theoreticity and first-person accessibility.

The "expressivist" epistemology in Yalowitz's dispositional account is intriguing, and appears to answer some of Wittgenstein's puzzles. For example, some of Wittgenstein's remarks in *Philosophical Investigation* (e.g., PI 208-211) seem to raise doubts about our commonsense idea that there is an apparent first-person/third-person asymmetry in semantic self-knowledge. That is, we seem to be in a better position than other people to know what I mean by my words. However, when one tries to tell people which function he means by "+", what he can appeal to is nothing more than a limited number of explicit performances, and it is from these performances other people make inductive inferences about which rule the person is following. The paradoxical point is that the person does not know more than he can explicit perform (this is manifested in his

⁹⁸ Yalowitz (2000), p. 270.

saying “and so on”). Thus, there seems to be symmetry. Yalowitz’s “expressivist” epistemology helps preserve the asymmetry by showing that an agent does stand in a better position than other people with respect to knowing his own meaning. He does so because the agent is the *possessor* of his own semantic disposition; all he has to do to know what he means is to *express* his semantic disposition, without having to guess or interpret what he himself means, as other people have to do.

Yalowitz’s account also helps to answer the question raised in last section that is left unanswered in Martin & Heil’s version of a dispositional account. Martin & Heil’s notion of cue manifestation of an underlying disposition claims that an agent with higher-level cognitive capacities comes to know his disposition by virtue of having a feeling or thought which monitors the disposition and informs the agent that he is in such a particular dispositional state. However, Martin & Heil fail to explain how a feeling or thought, being a cue manifestation of a disposition, manages to track or monitor the disposition. Yalowitz’s answer to this puzzle is that a (semantic) disposition is simply directly *expressed* in the form of a thought or a speech act, such as thinking or saying “ $12 + 3 = 15$ ”. This is how we come to know our disposition directly.

Yalowitz notices, however, that there remains an under-determination problem to his account of “expressivist” epistemology. That is, granting the presumption of first-person authority and “expressivist” epistemology, one knows, presumptively, that his particular usage of a term, expressed in the form of a speech act or a thought, is correct. Nonetheless, he may still not know which meaning his usage of a term expresses. For one’s semantic inclinations to apply a term so far have only been finite. However, there are an infinite number of meanings that those semantic inclinations could equally

correctly express. So presented, the under-determination facing Yalowitz's expressivist dispositional account is the finitude problem. Hence, even if semantic inclinations may be known to be correct because of their positive presumptive status, it does not follow that one knows what he means by those usages. Using addition to illustrate this point, we may say that a person may know that a particular usage of "+", such as " $3 + 12 = 15$ ", is correct, but still not know whether he uses "+" to mean plus or quus. In short, the presumption of first-person authority may be necessary for semantic knowledge, but it may not be sufficient.

Yalowtiz argues that to solve the under-determination problem, it is crucial to ask the following question:

What is supposed to count as an adequate representation or articulation of semantic knowledge? In particular, is semantic knowledge best construed propositionally, in terms of knowing that one's words mean such-and-such; or rather, should it be construed in terms of knowing how to use one's words correctly?⁹⁹

Yalowitz argues that it is *knowing-how* knowledge that constitutes one's semantic knowledge, and *knowing-that* knowledge of what an expression means is only parasitic on knowing how to use it. As Wittgenstein has shown us, fixing a rule by appeal to another more basic rule, which is propositional in form, leads to an infinite regress because the more basic rule itself stands in need of further interpretation in order to have its own content fixed. Thus, knowing-that knowledge which is essentially propositional cannot constitute semantic knowledge. This leaves knowing-how to apply a term the only plausible candidate for adequate semantic knowledge, and Yalowitz claims that

⁹⁹ Yalowitz (2000), p. 265.

knowing how to use a term is sufficient for semantic knowledge. He writes: “Ultimately, all we can do is *show* what our words mean in these cases, by using them correctly.”¹⁰⁰

In my view, Yalowitz’s claim that knowing-that knowledge is parasitic on knowing-how knowledge is intriguing, for the former is doomed to generate an infinite regress because of its propositional form. Hence, knowing-that knowledge does not seem able to constitute semantic knowledge. However, it does not follow from this that knowing-how knowledge is *sufficient* for semantic knowledge. Even granting that all the finitely many performances or judgments of an agent are correct, they still underdetermine an infinity of mutually conflicting rules. It is still possible that some non-standard rule which begins to deviate wildly from the end point of my finite performances or judgments is being followed. Therefore, though making the distinction between knowing-how and knowing-that is illuminating, it does not solve the finitude problem.

Overall, Yalowitz’s “expressivist” dispositional account of meaning is, in my view, on the right track and is fruitful. It overlaps to a considerable extent with the introspective model of meaning and rule-following that I will develop in chapter VIII. I will show that the model I offer differs from Yalowitz’s “expressivist” account. In particular, mine has a larger element of knowing-that semantic knowledge.

This review of dispositionalist accounts has shown us that the error problem persists as a thorny difficulty for any attempt to naturalize content. A convincing metaphysical view of a semantic disposition can be offered, such as Martin & Heil’s, and a plausible delineation of first-person dimension of meaning within a dispositional framework can be

¹⁰⁰ Yalowitz (2000), p. 266.

proposed, as Yalowitz's account does. But even if so, skeptical questions concerning what one means by a word still arise, if conditions of correct and incorrect uses cannot be non-circularly specified. Such skeptical questions would naturally and legitimately lead to both a metaphysical doubt about whether or not a disposition is an appropriate fact constitutive of meaning and an epistemic doubt about any significant role attributed to first-person verdicts or phenomenology in the constitution of meaning. Therefore, the problem of content determination, which is sharply brought out by the error problem, remains a core problem for the naturalistic dispositional approach.

There are, however, several important and extremely influential naturalistic theories of content determination, such as Fodor or Drestke's causal/informational approach and Millikan's teleological approach, which all try very hard to solve the error problem. To find out whether or not the error problem has any prospect of being solved, it will hence be crucial that we examine the plausibility of those theories. I will, therefore, continue the survey of prominent dispositional accounts in the next chapter.

Chapter VI. The Dispositionalist Solution to The Skeptical Paradox: A More Likely Approach (II)

In the philosophy of mind, several vigorous approaches have been proposed to naturalize mind (as well as meaning). Representative of them are the causal/informational approach on the one hand and the teleological approach on the other. The causal/informational approach, the major advocates of which are Fodor¹⁰¹ and Dretske¹⁰², is characterizable as a dispositional approach, because this approach to mental content can be expressed in a dispositional form at its root. For example, John's mental or linguistic symbol "cow" expresses the property John is disposed to apply it to. So characterized, the causal/informational theories would clearly be subject to the error problem and obligated to offer a solution to it.

On the other hand, Millikan's teleological approach is, strictly speaking, not a causal/dispositional account of mental content, for its main theme is that a causal/dispositional property leaves out something crucial that is constitutive of mental content, i.e., biological function. In Millikan's view, therefore, mental content or meaning is not constituted by a disposition. Nevertheless, Millikan regards the challenge of offering a standard of correctness and incorrectness presented by KW as "a welcome one"¹⁰³, and attempts to offer an account of normativity in her teleological approach.

It is therefore important that we examine whether or not these leading naturalistic theories offer satisfactory solutions to the error problem. I will examine three of the most

¹⁰¹ See Fodor, 1987, 1990.

¹⁰² See Dretske, 1981.

¹⁰³ Millikan, "Truth rules, hoverflies, and the Kripke-Wittgenstein paradox", in *White Queen Psychology*, 1993, p. 211.

important naturalistic theories, developed by Fodor, Drestke, and Millikan. I will not go into the details of their theories of content, but will rather concentrate on how they solve the error problem. I will begin with Millikan's theory.

Millikan's Teleological Approach

Millikan takes the error problem to be the main problem. Millikan views the error problem as involving the normativity of meaning and mental content, which in turn involves the possibility of offering a standard of correctness and incorrectness within a naturalistic framework. She writes: "The problem is to account for the *normative* element that is involved when one means to follow a rule, to account for there being a *standard* from which the facts, or one's dispositions, can diverge."¹⁰⁴ Millikan attempts to solve the error problem by bringing in biological considerations to offer such a standard for the evaluation of human (and non-human) rule-following behaviors. In this section, I will examine whether or not Millikan's solution to the error problem in the dispositional account is satisfactory.

Millikan adopts a teleological approach¹⁰⁵ to language understanding and mental representation, according to which understanding a linguistic expression, possessing a mental state, or following a rule, is constituted by *biological competence*. The most crucial notion in understanding biological competency is *purposiveness*. According to Millikan, purposes in the biological sense are "functions fulfilled in accordance with evolutionary design."¹⁰⁶ Take an organ such as a heart for example. Hearts have the

¹⁰⁴ Millikan, 1993, p. 216. Note that though Millikan in this passage views a disposition as a "fact", she does not regard it as a meaning constituting fact.

¹⁰⁵ See Millikan, 1993, 1984.

¹⁰⁶ Millikan, 1993, p. 217.

function of pumping blood which causes blood to circulate in an organism. The function of hearts has been historically performed so that continued proliferation of hearts is made possible. The fulfillment of the biological function of hearts, i.e., pumping blood, helps explain both the reproductive success of certain species that possess hearts and the continued presence of hearts in those species.

Similarly, to follow a rule, to mean something by a term, or to possess a mental state, Millikan claims, all have to be accounted for in terms of their biological purposes. According to Millikan, there are certain cognitive mechanisms, such as memory, perception, and so on, typically neuro-physiologically based, that are responsible for producing propositional attitudes, linguistic understanding, rule-following behaviors, and so on. The cognitive mechanisms in question have historically performed so as to fulfill their biological functions in a way that is in accord with their evolutionary design, and in a way that enhanced the fitness of the ancestors of the species which possess those mechanisms.

How then does Millikan's evolutionary story account for the normative element that is much needed in human rule-following and meaning? To put the question another way: when and how is a rule-following behavior classified as correct or incorrect in Millikan's account? Her answer is that biological purposes provide the primary and ultimate source of normativity. More specifically, a disposition to apply a term in a certain way is correct insofar as it accords with the biological purpose of whatever linguistic and mental mechanisms are responsible for tokening a linguistic term, and incorrect insofar as it does not accord.

To illustrate how biological purpose provides a standard of correctness, Millikan uses a very simple case, which involves a hoverfly's displaying a competence in conforming to a rule. Male hoverflies exhibit behaviors of darting after female hoverflies. Most interestingly, their intercepting behaviors exhibit a certain uniform pattern that can be accurately described by a mathematical rule. Given the standard speed of a female hoverfly, a roughly uniform distance of detection, and a certain constant acceleration on the part of a male hoverfly, the geometry of motion dictates that the male hoverfly has to make a turn of a certain angle in order to intercept the female hoverfly. It turns out that a male hoverfly's intercepting movement matches the descriptions specified by the rules of the geometry of motion. In other words, a male hoverfly's behavior conforms to a rule. Millikan calls the rule "the proximal hoverfly rule"¹⁰⁷, because it is a rule a male hoverfly conforms to in reaction to a proximal stimulus.

Millikan does not distinguish between the notions of rule-conforming and rule-following in her discussion of the hoverfly case. Pettit¹⁰⁸ (1993) makes a useful distinction between these two notions. An object's (or subject's) behavior is rule-conforming if the behavior can be described by some rule from a third-person objective point of view. On the other hand, for a behavior to belong to the rule-following category, not only does it have to be describable by some rule, but also the subject who produces the behavior subscribes himself to a rule that may or may not coincide with the ascribed rule. So, in Pettit's terminology, a male hoverfly cannot be said to follow a rule, but can only be said to conform to a rule. For a hoverfly is presumably such a low level species

¹⁰⁷ Millikan. 1993, p. 219.

¹⁰⁸ P. Pettit, *The Common Mind*, 1993.

that it is incapable of subscribing to a rule, which requires conscious and reflective mental capacities that a hoverfly apparently lacks. In Millikan's discussion, such a distinction between rule-conforming and rule-following does not matter, for, as we will see shortly, it is unconscious biological purposes that matter in the constitution of rule-following in her account. In the following discussion, I will adopt Millikan's conflation of the two notions in question. But it's important and useful to bear Pettit's distinction in mind.

There is one important fact to know about the hoverfly case. A male hoverfly not only darts after a female hoverfly, but it also darts after pebbles, distant birds, midges, and so on, as far as they appear in the image on the male hoverfly's retina with a similar speed and size as a female hoverfly. In both cases, one in which the object of interception is appropriate (i.e., female hoverflies) and another in which the object of interception is inappropriate (i.e., pebbles or others), a male hoverfly's intercepting behaviors conform to the proximal hoverfly rule.

Millikan points out that when a male hoverfly's behavior conforms to the proximal hoverfly rule, it does not calculate over the inner representation of the rule in order to follow it. The rule being followed is both *unexpressed* and *unconscious*. Due to its endowment of a general genetically determined mechanism of some kind, a male hoverfly is able to conform to (or to follow) the unexpressed rule. Millikan argues that the intercepting mechanism of a male hoverfly has a biological purpose, which is similarly unexpressed, that enables the hoverfly to conform to the unexpressed rule. The unexpressed biological purpose of a male hoverfly is achieved through normal functioning of the mechanism the male hoverfly possesses. The mechanism in question

has historically acted so as to fulfill the function of getting male and female hoverflies together and of enhancing the reproductive success of the hoverfly genes. Millikan thus claims that the male hoverfly shows a *biological competence* or *ability* in following the proximal hoverfly rule.

Millikan emphasizes that a biological competence to follow a rule is not equivalent to a disposition to follow a rule. For a disposition is neither sufficient nor necessary for a biological competency. A male hoverfly has a disposition to be blown by a wind, but this disposition does not correspond to a biological competence or purpose. Hence, a disposition is not sufficient for a biological competence. A male hoverfly might be blind, and consequently may never exhibit a disposition to behave in a way that conforms to the hoverfly rule. Nevertheless, it possesses a biological ability to do so. Therefore, a disposition is not necessary to a biological competence. Because the rule-following of a male hoverfly is constituted by the biological competence or purpose it possesses and a disposition is neither necessary nor sufficient for a biological competence, Millikan thus claims that the rule-following of a male hoverfly is not constituted by a disposition.

It is, however, not clear that a blind hoverfly does have the ability to conform to the rule. Millikan's claim seems to be based on the following idea. A hoverfly inherits certain intercepting traits in its genes. These inherited traits remain intact even when their possessor happens to be blind. It is those inherited traits in the genes that constitute a biological competence. Therefore, even though a blind hoverfly may lack a disposition to intercept because of its blindness, it still has intercepting traits in its genes and hence still has a biological competence and hence an ability to conform to the hoverfly rule. The idea, in short, is that because hoverflies in general have the intercepting disposition,

the defective ones do, too, in light of its evolutionary history in ancestry. The problem with Millikan's idea is that Millikan conflates the distinction between having a gene whose function it is to create some behavior and the organism with the genes actually showing that behavior. Millikan apparently uses the former property to characterize having an ability, but having an ability can also be characterized by the latter property. Something, such as blindness, may keep the gene from performing its function. According to the latter characterization of having an ability, a particular male hoverfly has no ability to conform to the hoverfly rule. For even though it has the genetic basis of the ability, the basis in this case is not activated. Contrary to Millikan's claim, I am inclined to maintain that having a disposition to perform what a genetic trait enables an organism to do is necessary to having the ability, and hence to conforming or following a rule. In what follows, however, my criticism of Millikan's teleological position will not hinge on this point. It suffices that we register doubt on Millikan's claim that having a disposition is not necessary to conform or following a rule.

Human rule-following behavior, according to Millikan, is fundamentally the same as a male hoverfly's rule-following behavior. It is the biological purpose or competence that constitutes human as well as non-human rule-following. Millikan thus claims that human dispositions to follow a rule are neither necessary nor sufficient for rule-following. Moreover, a biological purpose in a human is the same as one in a hoverfly in that it is unexpressed and unconscious. A biological purpose does not occur in the consciousness of a human being when he follows a rule.

We notice, however, that there is an apparent dissimilarity between the case of a hoverfly and a human being. That is, a hoverfly is *unaware of* its following, whereas it is

typical of a human being that he is consciously reflective of the fact that he is following a rule. It is often the case that the rule a person follows is *expressed*. For example, when questioned about what he is doing, a person typically says: "I am doing *addition*", or "I am following the *plus* rule." Millikan argues that the dissimilarity in question is irrelevant when we consider the question of what *constitutes* (human or non-human) rule-following. It is biological purposes that matter in the constitution of rule-following because only by evaluating a behavior with regard to whether or not it is produced to fulfill a biological purpose that we can classify the behavior as either correct or incorrect. Millikan claims that the biological purpose a human being's rule-following behavior fulfills is, due to its being deeply rooted in human's evolutionary history, *unexpressed* and *unconscious*, as in the case of the hoverfly. An unexpressed biological purpose in a human being has to be presupposed as basic.

Millikan does not regard the first-person epistemic problem as serious, nor does she attempt to offer a solution to it, even though she recognizes that this problem seems to have bothered Kripke's Wittgenstein¹⁰⁹. In my view, her disregard of the significance of the first-person epistemic problem is unfortunate but unavoidable. It is unfortunate because having awareness of being guided by a rule is distinctive of human rule-following behavior, while species with lower cognitive capacities such as hoverflies lack such first-person phenomenology. It indeed appears to be mysterious and needs at least some explication what it is that a subject seems to be guided by a rule that he seems to have a conscious grasp. However, to acknowledge the role of first-person epistemic features in rule-following, she will encounter genuine difficulty accommodating first-

¹⁰⁹ See Millikan, 1993, p. 215, n4.

person epistemic characteristics of meaning in a naturalist framework in which meaning is reduced to physical/dispositional/causal properties. The paradox the dispositional account of rule-following must resolve is that on the one hand, meaning seems to be known directly, whereas on the other hand, a disposition, claimed to constitute meaning, can only be known inferentially, or so many philosophers appear to believe. How, then, can meaning be possibly constituted by a disposition? This difficulty is bound to occur as well in Millikan's naturalistic account in which meaning is constituted by biological properties, or more accurately, by biological purposes. A biological purpose is deeply rooted in the evolutionary history of an organic system and hence the organic system that possesses it is destined to be unconscious of it. How then can a biological purpose possibly constitute meaning, which appears to be known directly and consciously? Millikan's teleological solution is apparently unable to offer an answer.

An even more serious problem for Millikan's teleological account of rule-following concerns her solution to the error problem, which is Millikan's major concern. (So I will focus on her solution to the error problem.) Millikan's solution to the error problem is that a rule-following behavior is correct insofar as it is in accord with its biological purpose. In this light, a male hoverfly's behavior of darting after a female hoverfly conforms to the proximal hoverfly rule and is deemed correct, because it fulfills the biological purpose of the hoverfly's mechanism (we may call it a quasi-cognitive mechanism) that produces its rule-conforming behavior. Does this really solve the error problem? It is highly dubious for the following reason.

The mechanism of the male hoverfly, as we have mentioned earlier, also produces behaviors of darting after other objects flying by, such as a pebble, that occur in the

image on its retina. This behavior of darting after a pebble is also correct according to Millikan's account, because it not only conforms to the proximal hoverfly rule, but is also a product of the normal functioning of the mechanism fulfilling its biological purpose. The problem, however, is then that the male hoverfly can be described as having a disposition to follow the rule of darting after a female hoverfly and as having another disposition to follow the rule of darting after a female hoverfly *or* a pebble. Millikan's teleological account is not able to differentiate these two cases because both behaviors (darting after a female hoverfly and darting after a pebble) are correct on her account. Hence, the above two disposition ascriptions fit the male hoverfly's rule-following behaviors equally well. There is thus an indeterminacy lurking in Millikan's account of a male hoverfly's rule-following behavior. Thus, the error problem has not been solved by Millikan's account.¹¹⁰

However, there seems to be sufficient conceptual resources in Millikan's teleological account to accommodate the above difficulty. Millikan offers the concept of a "distal" hoverfly rule, in contrast to a "proximal" hoverfly rule. The distal hoverfly rule says that "if you see a female, catch it."¹¹¹ Compared to the distal hoverfly rule, the proximal hoverfly rule is an "overkill" rule, because darting after female as well as non-female hoverfly or darting after non-hoverfly objects conforms to it. Even though to follow the proximal hoverfly rule is also to follow an "overkill" rule, Millikan regards a male hoverfly's following the proximal hoverfly rule as fulfilling its biological purpose. It is just that, Millikan emphasizes, following the proximal hoverfly rule is only a *means*

¹¹⁰ See Fodor (1990), especially chapter 3, for taking this line of criticism of Millikan's teleological solution to the error problem.

¹¹¹ Millikan, 1993, p. 222.

for a male hoverfly to follow the distal hoverfly rule. It is the conformity to the distal hoverfly rule, rather than the conformity to the proximal hoverfly rule, that explains the *successes* of hoverfly ancestors' procreation, which in turn provides the ultimate source of a normative standard.

The appeal to the concept of a distal rule thus offers Millikan a way out of the difficulty of the error problem. It could now be claimed that a male hoverfly has a disposition to follow the rule of darting after a female hoverfly rather than another disposition to follow the rule of darting after a female hoverfly *or* a pebble. For it is the former disposition that does the job of explaining the reproductive successes of a hoverfly species. The key to identifying an erroneous rule-following behavior is not by looking at what stimulus *causes* it, but by looking at what *results* the stimulus brings about. Even though it is an instance of the proximal hoverfly rule, the rule-following behavior of darting after a pebble would be erroneous, because it cannot explain (only the behavior of darting after a female hoverfly can) why the hoverfly species has actually procreated.

Is this move successful in solving the error problem? It seems not. The difficulty concerns how do we determine a human distal rule the following of which explains human reproductive successes. There may be more distal human rules than we are able to identify without controversy. In the case of human rule-following and meaning, a huge portion of meaning something by a term and following a rule is not linked in any obvious and direct way to human's reproductive successes. Consider the following case. A mathematician, John, gives an explicit definition of a non-standard rule q^* for the mathematical sign "+", and decides to operate the "+" sign according to the rule q^* . John

then does many calculations in a regular and seemingly competent manner, but those calculations look quite unfamiliar and odd to us. When questioned about his answers, he would justify them by citing the explicit rule q^* that he intended to follow. He might, quite naturally, occasionally make some mistakes in his calculations when he followed the rule q^* . Now, is there any way Millikan's teleological account of the distal rule could identify John's erroneous performances, and hence determine the distal rule he follows? Millikan's account would suggest that we have to look at how John's dispositions have been affected by the evolutionary history in his ancestry in order to determine whether those performances are erroneous or not. The problem, however, is that John's performances do not seem to have anything to do with the procreation of his species. In the case of the hoverfly, it might be plausible for a biologist, or lay-people, to identify a male hoverfly's darting after female hoverflies, not pebbles or birds, as contributive to the reproductive successes of the hoverfly species. But, in the case of a person's performing a calculation, according especially to an artificial rule such as q^* , it is not at all clear how to determine whether the calculation is correct or incorrect by its evolutionary history. Our intuition, however, is that there is a matter of correctness and incorrectness as to the person's calculation. As a result, Millikan's account does not solve the error problem that KW raises.

Fodor's Asymmetric Dependence Account

Fodor proposes an asymmetric dependence account¹¹² of mental representation to solve the error problem, or the problem of misrepresentation. As a naturalist who holds a version of the causal theory of content, Fodor finds that the major conceptual obstacle for

¹¹² Fodor (1990), *A Theory of Content and Other Essays*.

the naturalization project of meaning and mind is how to pick out semantically or mentally relevant causal factors from all the other kinds of causal relations that tokens of a linguistic or mental symbol such as “+” or “cow” can enter into. For example, we use the word “cow” to mean cows. A causal theorist tells us that this meaning fact can be reductively explained as consisting in a causal relation between cow-caused “cow” tokens and cows. Roughly, the word “cow” means *cows* as far as the tokening of the word “cow” is caused by the presence of a cow. The problem is, however, that a horse as well as a cow might cause the tokening of the word “cow”, especially in a bad light. In that case, “cow” would have to mean *horse or cow*, because cows and horses could both be causally responsible for the tokenings of “cow”. This is undesirable, for we know perfectly well that “cow” means *cow* only, rather than *cow or horse*. Our commonsense intuition is that tokenings of “cow” in the presence of horses are errors, which should be excluded from the correct extensions of the term “cow”. Fodor calls this problem as “the disjunction problem”.

The disjunction problem is fundamentally the same as the error problem. According to the dispositional account, what a speaker means by his term, say, “cow”, has to be directly read off from the instances of a particular property to which the speaker is disposed to apply the term. The problem is then that the speaker may make a mistake by applying the term to a horse when it is dark or the viewing distance is too far. The term “cow” then comes to mean *cow or horse*, instead of *cow*. Thus, for a naturalistic causal theorist like Fodor as well as a dispositionalist, the main task of ascribing meaning to a term is to solve the “error” problem, which requires specifying causal factors in a non-circular way such that certain causal factors, i.e., those that lead to errors, are excluded

from the construction of the extension of a term, so that the meaning of the term can be determined.

Fodor's solution to the error or disjunction problem is as follows. "Cow" means *cow*, rather than *cow or horse*, because the existence of cow-caused "cow" tokens is a pre-requisite for the existence of horse-caused "cow" tokens. It cannot be the case that there are horse-caused "cow" tokens without the prior existence of cow-caused "cow" tokens. Put in Fodor's terms, "cow" means *cow*, rather than *cow or others*, because non-cow caused "cow" tokens are *asymmetrically dependent* upon cow-caused "cow" tokens.

Fodor's idea of asymmetrical dependence can be expressed in a general way by the following principle (P)¹¹³:

supposing that there are two laws that respectively express two causal relations: $A \rightarrow B$ and $C \rightarrow D$,

(P) The law $C \rightarrow D$ is asymmetrically dependent on the law $A \rightarrow B$ if and only if the A/B connection cannot break without breaking the C/D connection; and the C/D connection can break without breaking the A/B connection.

Using principle (P), we can determine dependency relations between two causal relations, and in turn determine the meaning of a term. Take the term "cow" for example. If we find out that i) the cow/"cow"-tokens connection cannot break without breaking the horse/"cow"-tokens connection, and ii) even if the horse/"cow"-tokens connection were to break, the cow/"cow"-tokens connection would remain intact, we may conclude that the law that horses cause "cow" tokens is asymmetrically dependent on the law that cows cause "cow" tokens, and consequently conclude that the term "cow" means *cow*, rather

¹¹³ The formulation is based on L.R. Baker. "Has Content Being Naturalized", in Loewer & Rey (1991), p. 18.

than *cow or horse*. Notice that this account does not require either that only cows cause “cow”- tokens or that all cows would cause “cow” tokens, in order for “cow” to mean cow. All that this account requires is that there are some cow-caused “cow” tokens. If all non-cow caused “cow” tokens are asymmetrically dependent on those cow-caused “cow” tokens, it can be concluded that “cow” means cow.

How does Fodor’s asymmetric dependence account of meaning ascription fare? I do not think that Kripke’s skeptic would regard Fodor’s account as satisfactory. For Fodor’s idea simply amounts to saying that errors are parasitic on the true meaning. When faced with two causal relations, Fodor’s account tells us that one of them determines the meaning of a term while the other is responsible for the production of errors, so far as the latter causal relation is asymmetrically dependent on the former. The problem, however, is how we are to determine which of the two causal relations to choose so that one is asymmetrically dependent on the other? Only after the dependency relation between two causal relations is determined can the true meaning of a term be fixed. However, when one, such as Fodor, comes to favor a causal relation on which the other comes to asymmetrically depend, he has to presuppose what a term such as “cow” means. Consequently, the charge of circularity is bound to apply to Fodor’s account.

Fodor would have a reply to my objection. He could argue that the determination of causal relations is a matter for empirical investigation. Behavioral dispositions can be extinguished by punishment. Suppose that extinguishing my tendency to call cows “cows” in the presence of cows also led to the disappearance of my disposition to call far away horses “cows”, but extinction of my tendency to call far away horses left the

cow/"cow" correlation intact. That would be evidence that the horse/"cow" disposition depends on the cow/"cow" one, and hence that "cow" means *cow*.

The problem with this method of investigation is that extinction of one disposition would necessarily lead to the extinction of the other. Presumably, administering an electric shock every time I said "cow" in the presence of a cow would extinguish my cow/"cow" correlation. But certainly the proximal stimuli for my saying "cow" in the presence of a cow is a visual appearance of a certain sort. So what I am really learning is not to say "cow" when presented with certain visual appearances. But distant horses also present those appearances. So I have also learned not to say "cow" when I see a horse in the distance. This is how extinction of the first disposition would, arguably, lead to extinction of the second. The crucial point, however, is to notice that similar treatment or punishment would extinguish the first disposition (the cow/"cow" correlation) by extinguishing the second one (the horse/"cow" correlation), since both cows and far-away horses present to an agent the same or very similar appearances. As a result, it is not clear how empirical investigations could help Fodor determine the dependency relation between two causal correlations.

Is this reply enough to refute Fodor's account? Couldn't there be indirect evidence that helps determine what the dependency relation is between two causal relations? It is not clear, however, what indirect evidence might be and how it helps determine the dependency relation in question. Unless it can be made more clearly, Fodor's asymmetric dependence account is not tenable.

Dretske's Distinctive Learning Stages Theory

Dretske¹¹⁴ accounts for errors or misrepresentations in meaning and mental representation by appeal to distinctive learning stages, i.e., the learning stage and after-learning stage. In the learning stage, a person is trained to establish a certain correlation between a term and the world, e.g., “cow” and cows. It is presumably the job of a teacher in the learning stage to ensure that the correlation is established in such sufficient strength and intensity that only, say, cows cause “cow” tokening behaviors. Once the nomic relation, i.e., “cow”-cows correlation, is established, it may be triggered by objects other than cows. These “wild” tokenings of a word are construed as “errors” or “misrepresentations” in Dretske’s account, because they fall outside the pattern of the appropriate manifestations of the dispositional correlation established during the learning stage. In short, in Dretske’s account, errors cannot be correctly understood unless we pay attention to the distinction between learning stages. Conceptually, errors cannot be appropriately attributed to a person during the learning stage; only after the learning stage can errors be legitimately ascribed.

In my view, Dretske’s learning-stage account captures a very important intuition regarding meaning/content fixation. When we are children, we come to know the meaning of a word by going through a learning stage in which we are exposed to a repeated and regular pattern of feedback from the environment as to how to apply a word to objects. Thus, a very natural and appropriate way to find out the meaning of one’s word is to ask the question: what is learned during the learning stage when a child is taught how to use it? It would be quite senseless to ask a question like “what does John mean by the word ‘dog’?” during the stage when John is still learning how to use the

¹¹⁴ See Dretske (1981), *Knowledge and the Flow of Information*.

word “dog”, or before John enters into a learning stage. Correspondingly, only after the learning stage does it make sense to ask if John has made a mistake in applying the word “dog”. In short, our intuition is that there is a word/world nomic relation established within the child during the learning stage, and that this nomic relation essentially contributes to the determination of the meaning of his word or the content of his mental representation. Errors, in Dretske’s account, have to be construed relative to learning stages. As Peng nicely puts the point: “Before the content of a mental representation is fixed, it makes no sense to say that some occurrences of it are mistaken. The content-fixation period is both *conceptually and temporally* prior to the post-fixation period.”¹¹⁵

There are, however, two major objections made by Fodor¹¹⁶ to Dretske’s learning-stage account. The first objection is that there is no way of telling when the learning period has ended and when the period of competently using a sign has begun. As Fodor asks rhetorically, does someone blow a whistle to signal the end of the learning period of a person and the beginning of his application of a word? There seems no sharp boundary between the learning stage and the after-learning stage. If so, then no attribution of error is possible, and hence meaning/content is impossible.

As Dretske¹¹⁷ points out, however, Fodor’s first objection is based on a sort of the Sortes argument: If you can’t say when a poor man ceases to be poor as you continue to give him pennies, then there is really no principled distinction between rich people and poor people. The Sortes argument implies that, due to the lack of a sharp boundary

¹¹⁵ See Eric Peng, “Disjunction, Ambiguity, and the Problem of Content-fixation”, in *Taiwaness Journal for Philosophy and History of Science*, 1996, p. 33.

¹¹⁶ See Fodor (1990), esp. chapters II and III.

¹¹⁷ See Dretske (1991), *Dretske and his Critics*, esp. p. 208.

between being poor and being rich, we cannot legitimately attribute the property of being rich (or being poor) to anyone. Similarly, based on the observation that there is no sharp boundary between the learning and after-learning stages, Fodor claims that we cannot distinguish between a competent user of a term and an incompetent, still-learning person, and consequently, we cannot legitimately attribute the status of meaning something by a term to someone. But this is obviously absurd, Dretske argues. In many cases, we do have a clear idea about whether a person is poor or not. The existence of borderline cases does not constitute any fundamental difficulty for our attribution of someone's being poor (or being rich). Similarly, even though no clear line can be drawn between the learning and the after-learning periods, it is absurd to go so far as to declare, as Fodor does, that no one can be unambiguously identified as being in the after-stage learning stage. I agree with Dretske that Fodor's first objection appears to be based on some version of the Sortes argument, hence is unconvincing.

The second major difficulty for Dretske's learning-stage account, according to Fodor, is more serious: Dretske's account ignores relevant counterfactuals. We may grant that a trainee received an error-free training during his learning stage in which his trainee's utterances of, say, "cow" tokens were only caused by cows. Fodor argues that this does not guarantee that the training established the cow/"cow" correlation rather than cow-or-horse/"cow" correlations, and hence does not ensure that the trainee uses "cow" to mean *cows* rather than *cows or horses*. For it *could have been* the case that during the learning stage the trainee encountered a horse, perhaps in a dark night, which caused him to utter "cow". The trainee thus could have learned to mean *cow or horse* by "cow". The disjunction problem, Fodor argues, returns.

A natural and seemingly convincing response to Fodor's second objection that Dretske could make is that Fodor has failed to consider further counterfactuals, such as the following one: had the trainee encountered a horse and uttered a "cow" token in the learning stage, he would have been corrected by the teacher. Fodor argues, however, that this response is hopeless, too, for it is the teacher's pedagogical *intentions*, and thus *meaning*, to reward certain behaviors, rather than actual or counterfactuals correlations that hold between a word and the world, that come to determine which nomic property is established, and hence which meaning of a word is learned. The skeptical constraint that no semantic or intentional terms be used in the specification of a meaning fact would be violated in such a response.¹¹⁸ Hence, the circularity problem remains.

I do not think that a teacher's intentions are necessarily involved in the characterization of his correcting behaviors. What is important is that, had the trainee issued certain utterances, he would have received certain negative feedback from the environment (a teacher in this case). The feedback in question can be described in an objective and neutral way, such as in terms of a distribution of actual and counterfactual punishments and rewards so that it is not contaminated by anything intentional. For example, a counterfactual may go like this: had a trainee responded "horse" to a cow, he would have received an electrical shock from the teacher; whereas had he responded "cow" to a cow, he would have received a cookie from the teacher. As a matter of fact, Fodor seems to think that this is a plausible response, and thus goes on in a latter chapter of the same book to argue that the appeal to an objective distribution of feedbacks still cannot do the trick of determining what is learned. The reason Fodor offers is that "it

¹¹⁸ See Fodor (1990), Ch II, p. 42.

would be impossible for one to learn a disjunctive concept when a nondisjunctive concept is being taught.”¹¹⁹ For example, it may somehow occur to me that you are trying to teach me *cow or horse*, rather than *cow*. However, according to Dretske’s theory, I could never acquire the concept *horse*, but can only acquire the concept *cow*, no matter what I would do and how I would respond in the learning situation, because it is the objective pattern of punishments and rewards that fix what I learn. Fodor’s point is that there is more involved in the determination of what one learns than mere externally specified pattern of punishments and rewards. According to Fodor, a learner is actively participating in a learning environment, rather than passively receiving external stimuli and feedback which are objectively specifiable. As Fodor writes: “...what constitutes my concepts is not *the objective reinforcement contingencies*, but rather *the reinforcement contingencies that I take to obtain*.”¹²⁰

I think Fodor has made a strong case against Dretske’s learning-stage account, because it has been shown that Dretske would have a hard time determine without circularity which pattern of reinforcement contingency shapes the concept acquired by an agent during the learning stage. Dretske’s account is thus unable to solve the error problem presented in KW’s skeptical argument.

It is worth noting, however, Fodor’s effective objection not only shows the inadequacy of Dretske’s learning-stage account of concept or meaning acquisition, but also reveals that Fodor’s own asymmetric dependence account is on a par with Dretske’s learning-stage account with regard to answering to KW’s skeptical challenge. As we have seen, Fodor’s asymmetric dependence account of meaning ascription only tells us

¹¹⁹ Fodor (1990), Ch III, p. 63.

¹²⁰ Fodor (1990), p. 63.

that what a person means by a term is determined by certain dependency relations between two distinct causal connections. The account does not by itself tell us which causal connection is asymmetrically dependent upon which. Fodor's account simply has to presuppose that a particular nomic relation is in place while the other is not in order for the meaning of a term to be fixed. Kripke's skeptic would hence charge Fodor's account with circularity. Here in raising his second objection against Dretske, Fodor has exactly the same tone as Kripke's skeptic: why is it that this rather than that pattern of reinforcement contingencies is favored in determining the concept acquired in the learning stage? Dretske himself does not offer a satisfactory response to this challenge. However, it is equally clear that Fodor faces the same problem, and does not offer a satisfactory response either.

The upshot then is that to the extent that Fodor's second objection is effective against Dretske's learning-stage account of meaning/concept determination in responding to KW's skeptical challenge, Fodor's own asymmetric dependence account is subject to the same difficulty. No determinate meaning can be fixed either by appeal to a pattern of counterfactual reinforcement contingencies given in a learning stage or by a nomic property on which other nomic properties come to causally depend. For circularity is bound to arise in either account.

The review of these causal/dispositional/teleological accounts of meaning and mental representation shows that the error problem remains unresolved. Therefore, KW's skeptical challenge of finding a fact that can determine the meaning of one's linguistic expression or the content of one's thought has not been met. Nevertheless, the accounts reviewed in this and last chapters, I believe, have offered useful conceptual

resources and important insights in our search for a satisfactory solution. In the next chapter, I will try to show how those resources and insight can help solve the error problem (and also the finitude problem) within a dispositional framework. In the chapter after that, I will focus on solving the first-person epistemic problem within the same framework.

Chapter VII. Proposing A Satisfactory Dispositional Account (I):

Solving the Error and the Finitude Problems

Old Plagues and New Lights in the Dispositional Account of Meaning and Rule-Following

The review of various dispositional and functional theories of rule-following and meaning-ascription in the previous two chapters shows that the skeptical paradox raised by KW has not been satisfactorily solved. In particular, the error problem continues to plague dispositionalists (as well as functionalists). To solve the problem, a dispositionalist must find a way to distinguish between the disposition that produces correct usage of a word and the disposition that produces incorrect usage of the same word. Many philosophers indeed have attempted to do so. However, no one has succeeded so far.

For example, Shogenji (1993) appeals to the concept of stable rule-following behaviors to serve as the criterion of correct behaviors in rule-following. The criterion does not appear to be circular at the first glance, because what those stable behaviors a person is disposed to produce are left entirely open. For example, if a subject stably responds "1" to the question " $4 + 3 = ?$ ", then the basic routine rule he is following is $4 + 3 = 1$ according to Shogenji's account. But upon reflection, epistemic doubt arises concerning why a particular set of performances (i.e., stable ones) rather than another set of performances (e.g., unstable ones that include actual or potential "deviant" performances) is chosen to fix the rule. A justification must be provided to explain why a particular pattern of behaviors is designated as correct. Otherwise, Shogenji's account is bound to presuppose which rule is followed. Moreover, even if a reason is provided to

justify why a particular method of selecting correct behaviors is adopted, further question may still arise as to why the provided reason stands. An infinite regress for justification would thus occur.

Although the error problem seems insurmountable to a dispositionalist, some new insights have been shed on the nature of a human disposition and the possible faults in KW's discussion about human dispositions. For example, due to the difficulties involved in providing a justifiable criterion for distinguishing the disposition that produces correct behaviors from the one that produces incorrect behaviors, perhaps we should simply abandon the attempt. We may instead claim that each particular behavior of an agent that is judged by the agent himself as correct is correct, presumptively. That means, one's verdicts about one's own behaviors serve as a criterion of correct (or incorrect) behaviors, unless there is contrary evidence available to override those verdicts. Yalowitz (2000), following Wright (1989), adopts this approach. He argues that first-person authority, which is displayed in one's verdicts, is fundamental to human rationality, and is as fundamental to our commonsense picture of the mind as normativity (i.e., correctness conditions) is. In Yalowitz's view, Kripke is wrong in bracketing the presumption of a human's first-person authority. The problem with Yalowitz's approach, however, is that one's verdicts, either in the form of a judgment or a feeling of certainty, can sometimes guide one wrong when following a rule. It cannot be determined decisively which verdict is right or when a verdict is right. Consequently, a standard for distinguishing the correct behaviors from the incorrect ones remains needed to fix the rule being followed. It is precisely because we cannot be sure which use of a term is correct and hence which meaning the use of the term expresses that Kripke's skeptic suggests that we give up first-

person authority as a means to fix the meaning of a word. Yalowits' emphasis on the first-person authority of one's disposition to make judgments about one's own rule-following behaviors simply leaves the error problem intact.

Martin & Heill (1998) offer another fresh perspective on how to solve the error problem which also helps develop a dispositional account of rule-following. They direct our attention to the ontological character and status of a disposition claimed to be constitutive of rule-following or meaning. According to Martin & Heil, a disposition is an actual property. A real disposition can persist or undergo changes without ever manifesting itself or its changes. Thus, a disposition has to be suitably distinguished from its manifestations under various conditions. Kripke, Martin & Heil claim, conflates the distinction between a disposition and its manifestations. The error problem arises as a result of this conflation. For the fact that constitutes rule-following or meaning lies in a real and actual disposition, rather than its manifested behaviors which may or may not actually occur. It is thus wrongheaded to search for a criterion or standard by which we can distinguish correct manifested behaviors from the incorrect ones. The right way to go, according to Martin & Heil's realist account, is to identify what the real rule-disposition is that is possessed by its subject and responsible for the production of rule-following behaviors. An erroneous behavior of an agent in their account is then viewed as occurring when an interfering factor enters into causal contact with the particular real disposition constitutive of the agent's following a particular rule.

This way of looking at how an infallible disposition may produce erroneous behaviors is similar to Chomsky's (1965, 1968) criticism of behaviorism, based on the important distinction between performance and competence in psychology and

linguistics. Chomsky argues that behaviorists have an unfruitful research program that tries to provide a theory of language by reference to overt behaviors. This is similar to forming a theory of weather by the actual observation of weather, rather than constructing theories of thermodynamics, hydrodynamics, etc. that can serve to explain weather that we actually get¹²¹. Kripke's error problem for the dispositional account of rule-following, in Martin & Heil's view, makes a mistake similar to the behaviorists'. Kripke treats a disposition to follow a rule entirely in terms of rule-following behaviors produced under certain conditions, while he ignores the underlying real disposition that produces them.

The problem is, however, that the error problem cannot be gotten rid of so easily. Even if the distinction between a real disposition and its manifested behaviors is suitably made, and even if it is granted that it is a real disposition that constitutes rule-following, epistemological difficulties remain as to how we can non-circularly identify a real disposition. A real disposition is, unlike an observable manifested behavior, invisible. The only plausible way to know what an agent's underlying disposition is is to infer it by observing its manifestations under various conditions and by forming a hypothesis or theory about what the underlying disposition is. However, there surely exist potential counterexamples, i.e., erroneous behaviors, to any hypothesis that ascribes a certain disposition to a subject. How then can we be justified in treating those counterexamples as mere erroneous behaviors rather than as evidence that the hypothesis is false? Forming a hypothesis about a subject's possessing a particular disposition unavoidably needs to make certain presuppositions about the rule being followed and thus appears to be circular.

¹²¹ See Pietrosky and Rey (1995) on this point.

Despite these difficulties, Martin & Heil's realist dispositional account of rule-following is worth pursuing. Later in this chapter, I will take up the error problem that is left unresolved in their account. More specifically, I will argue that hypothesizing what a real disposition is is a perfectly legitimate scientific practice, and hence that charge of circularity can be met.

There is another way to approach the error problem within the same realist framework that may also shed some light on our understanding of a human disposition to follow a rule. If it is the real disposition, rather than its manifestations, that is to be identified as the fact that determines the rule being followed, then we need to say something about how such a disposition comes into existence. In my view, a disposition of an agent to follow a rule is presumably acquired through learning. Therefore, to determine which rule an agent follows (or equivalently, which real disposition the agent possesses in a realist dispositional account), it is natural and reasonable to ask which rule is learned (or which real disposition is established) during the learning stage. If we can determine which real disposition gets established during the learning stage, we can determine which rule is being followed in the after-learning stage. This intriguing way of solving the error problem is offered by Dretske (1981).

Dretske's approach is similar to Martin & Heil's in the following way. From the metaphysical point of view that Martin & Heil are keen to emphasize, the error problem is misleading because it is the result of Kripke's adopting an ontologically misconceived view of a disposition. Similarly, from the conceptual point of view inherent in the Dretskean dispositional account, the error problem is no less misconceived, because erroneous behaviors are possible only in the after-learning stage. That is to say,

erroneous behaviors are conceptually possible only in the after-learning stage when a particular real disposition is already established. Before or during the learning stage, it is not appropriate to attribute errors to an agent, simply because the agent is still learning. Thus, for Kripke, to ask for a criterion that can non-circularly specify correct rule-following behaviors, he must have presumed that rule-following is existent (or that a real disposition is in place).

The problem that faces the Dretskean approach is, again, an epistemological one: how we determine which rule is learned (or which real disposition gets established) during the learning stage. There are potentially an infinity of rules that are mutually inconsistent with each other but are compatible with the finite set of examples and feedback given to a pupil during his learning stage. How are we to determine which rule gets picked up (or which real disposition a pupil acquires)? How are we justified in claiming that one rather than another rule is learned during the learning stage? This issue is neither trivial nor simple. Later, in this chapter, I will show that the epistemological difficulties that are left unresolved in Dretske's account are solvable. By doing so, a clearer characterization of a real disposition to rule-following can, hopefully, be offered.

The rest of this chapter has four sections. In section II, I will show that Martin & Heil's realist view of a disposition must be supplemented with a counterfactual analysis to express what a real disposition is. The error problem then re-emerges. In section III, I will show how the error problem can be solved by arguing for the legitimacy of adding a normality condition to a counterfactual conditional of a disposition. This amounts to arguing for the legitimacy of *ceteris paribus* laws. In section IV, I will first try to show that the source of difficulty in Dretske's distinctive learning-stage account comes from

his failing to take into consideration an agent's inner cognitive characteristics involved in learning. I will then argue that once the inner constraints imposed on human learning are taken into consideration, we will have good epistemic ground to claim that one rather than another rule is learned and hence followed afterwards. In section V, I will outline a solution to the finitude problem within a realist dispositional framework. In particular, I will argue that the finitude problem is based on a sort of inductive skepticism and the solution to it requires a reliabilist epistemology.

A Counterfactual Analysis of Dispositions, and the Error Problem in Martin & Heil's Realist View of Dispositions

In KW's discussion, a semantic disposition is construed counterfactually. For example, an agent's meaning plus by "+" is analyzed as "If an agent were queried so and so, he would answer this and that." In this section, we will first see how the error problem arises as a result of adopting such a counterfactual analysis of a disposition and consequently why it constitutes a major obstacle for a dispositionalist's attempt to account for meaning and rule-following. We will then review Martin & Heil's proposal to see how it might offer a more compelling view of dispositions. Nevertheless, I will argue that the error problem persists. I will then argue that a counterfactual conditional is still needed to describe what particular real disposition an object or subject may possess. Consequently, there is no escape from the fact that a dispositionalist has to solve peculiar difficulties, such as vacuity and circularity, resulting from the employment of counterfactuals.

The Counterfactual Idea of A Disposition Held by KW

In KW's discussion of a dispositional account of rule-following, the traditional view of a disposition is employed. The traditional approach, taken by Ryle and others, uses a counterfactual conditional to analyze a disposition, intending to give a reductive explanation of something mysterious by something that is less mysterious. For example, the statement (f) ascribes a disposition, e.g., fragility, to a piece of glass:

(f) The piece of glass is fragile.

The traditional view suggests that we replace (f) by the counterfactual conditional (g):

(g) If the piece of glass on the table were to fall on the floor, it would break.

The counterfactual conditional (g) comprises two indicative statements: the antecedent describes a certain testing situation, while the consequent describes a certain reaction of the piece of glass when put in the testing situation. If (g) can be viewed as a necessary and sufficient condition for (f), then (g) offers a reductive explanation of (f). In that case, the mysterious unobservable disposition, i.e., the fragility, is reductively explained as nothing over and above the object's reaction in a certain way under certain conditions.

Similarly, in KW's argument against a dispositional account of meaning or rule-following, meaning is first equated with a semantic disposition, and then statements about the semantic disposition is further analyzed in terms of a set of counterfactual conditionals. Take an agent's meaning plus by "+" for example. In a counterfactual analysis, a meaning-ascribing statement (m):

(m) S meant plus by "+"

is replaced by a disposition-ascribing statement (d):

(d) S had a certain disposition with respect to the operation of the sign "+"

(d) is in turn analyzed in terms of a set of infinitely many counterfactual conditionals of the form (c):

(c) If S had been queried about " $m + n$ ", he would have responded with the sum r ,
where the triple of (m, n, r) is an instance dictated by the plus rule.

An example of the schema (c) would be (ce)

(ce) If S had been queried about " $72 + 25$ ", he would have responded with 97.

The conditional schema (c) describes a counterfactual relation holding between two observable states: a certain testing situation and S's reaction in that situation. The idea of the counterfactual analysis of a disposition is thus to reduce mysterious meaning to a less mysterious disposition, and then in turn explain a disposition in terms of an observable relation between situations and behavior.

In order for the counterfactual (c) to analyze the disposition ascription (d) and hence reductively explain the meaning ascription (m), however, the infinite set of instances of (c) has to be necessary and sufficient for (m). This turns out to be very difficult to accomplish. For the indicative statement (m) is sometimes true while the counterfactual (c) is false, and *vice versa*. For example, an agent might mean plus by "+" but fail to answer "125" to the question " $68 + 57 = ?$ " when queried. At the moment of being queried, the agent might have been in some unfavorable condition such as being drunk or tired. In this case, the counterfactual (c) is not necessary for (m). On the other hand, an agent might not mean plus by "+" but answer "125" to " $68 + 57$ " when queried. Again, the agent might have been in some unfavorable condition such as being drunk or tired. In such case, the counterfactual (c) is not sufficient for (m). Since (c) is neither necessary

nor sufficient for (m), (c) does not reductively explain a semantic disposition successfully.

As explained above, the difficulty of providing the necessary and sufficient conditions for a counterfactual conditional like (c) derives from the fact that people more often than not make mistakes under unfavorable conditions. An agent might answer “123” to “68 + 57” but still means plus by “+”. This is because his answering “123” is a careless mistake. Similarly, an agent might answer “125” to “68 + 57”, but actually means quus by “+”, because his answering “125” is also a careless mistake relative to the meaning of his word. The challenge of the error problem can hence be viewed as the challenge of offering a necessary and sufficient condition for a set of counterfactual conditionals such as (c) that is indicative of a rule-following disposition. A typical way to meet this challenge is to supplement the antecedent of a counterfactual with a normality condition from which (together with the antecedent) the consequent of the counterfactual can be derived. For example, a normality condition can be specified as one in which an agent is neither drunk nor tired. In such a condition, how one is disposed to answer determines what he means by a sign. Underlying the specification of a normality condition is the idea that the performances which an agent carries out under such a condition will be correct. Hence, if an agent follows the plus rule, a normality condition ensures that he is able to answer “125” when queried with the question “68 + 57” under this normality condition.

The problem, however, with adding a normality condition to a counterfactual conditional is that the resulting conditional is either vacuous or circular. An agent might make a mistake not only when he is drunk or tired but also when he receives a heavy

blow on his head. He might thus answer “5” when queried “ $68 + 57$ ”, even though he actually follows the plus rule. Such an unfavorable situation surely has to be excluded from the normality condition. We soon find that it’s extremely difficult to specify a competent normality condition that covers all conceivable situations under which an error might occur. For not only could receiving a heavy blow on the head lead to erroneous performances, but numerous other factors could. It thus seems that we have to add a normality condition to (c) as follows:

(c1) If S had not been drunk, not tired, not received a heavy blow on the head, and *not in any other unfavorable conditions*, S would have responded with the sum “r” when queried “ $m + n$ ”.

The conditional (c1) is obviously vacuous, for what “any other unfavorable situations” are is not clearly specified. The difficulty of giving a non-vacuous specification of a normal condition might in turn tempt us to give the following formulation of a counterfactual:

(c2) If nothing had interfered, S would have answered the sum “r” when he had been queried about “ $m + n$ ”.

Specifying the normality condition this way as in (c2) renders (c2) trivially true, for the insertion of “had nothing interfered” simply amounts to presupposing that the original counterfactual (c) is true, and hence presuppose the truth of the meaning-ascribing sentence (m). Adding a normality condition to a counterfactual condition is thus either vacuous or circular. As a result, the prospect for a dispositional account of rule-following, where a disposition is given a counterfactual analysis, is dim.¹²²

¹²² The difficulty of providing a necessary and sufficient condition for the truth of a statement ascribing a

The error problem for a dispositional account of rule-following is thus fundamentally a task of saving a counterfactual conditional from vacuity or circularity, given the counterfactual analysis of a disposition held by KW's skeptic.

Martin & Heil's Realist Idea of A Disposition

Martin & Heil's dispositional account of rule-following does not confront the error problem directly by offering a plausible set of counterfactuals. Instead, it solves the error problem by discarding KW's counterfactual analysis of dispositions. In KW's counterfactual construal of dispositions, rule-following is ontologically nothing but a set of infinite potential behaviors exhibited under an infinity of possible testing conditions. In this view, a disposition is a mere *possibilia*, while its manifestations are actual and real. In other words, a disposition is ontologically reducible to a set of testing conditions and observable behaviors. Martin & Heil reverse the order of explanation by arguing that a disposition is actual, while its manifestations are mere *possibilia*. In their realist view, an object possessing a disposition has duration, and hence may undergo changes. So, for example, a piece of glass can be fragile for an hour and cease to be fragile for another hour, due to some natural factors such as changes of temperature or (imaginary) influences such as the interference of a Martian's remote control or a divine agent's order. Hence, a disposition and the change of a disposition may not (or even never) manifest itself. Therefore, manifestations are mere *possibilia*, while the disposition that causes them is actual.

disposition to follow the plus rule presented here parallels Martin's (1994) criticism of a conditional analysis of a disposition. There Martin raises a bizarre but admissible electro-fink case and argues that a conditional analysis of a disposition is neither necessary nor sufficient for a disposition-ascribing statement. I will discuss this case shortly in the next section. See Martin, "Dispositions and Conditionals", in *The Philosophical Quarterly*, 1994.

Martin (1994) offers an important argument against the conditional analysis of a disposition. He asks us to consider first the following disposition ascription A:

A: The wire is live.

And then, consider the conditional B:

B: If the wire is touched by a conductor then electric current flows from the wire to the conductor.

On the reductionist view, B is meant to be A's *analysans*. It turns out, however, that B is neither necessary nor sufficient for A. Martin shows this by considering an electro-fink machine that is connected to the wire. The function of the electro-fink machine is to detect when the wire is about to be touched by a conductor and can instantaneously make a live wire dead or a dead wire live. An electro-fink thus produces the following two cases. In the first case, imagine that the wire is dead and not touched by a conductor. In that case, A is false but B is true. B is true because the wire's connecting to an electro-fink machine ensures that if the wire were to be touched by a conductor then electric current would flow to the conductor. As a result, B is not sufficient for A. In the second case, imagine that the wire is live and not touched by a conductor. A is thus true while B is false. B is false because, once again, the wire is connected to an electro-fink machine, which renders it the case that if the wire were to be touched by a conductor, then current would not flow from the wire to the conductor. Hence, B is not necessary for A. The electro-fink case thus shows that no conditional is necessary or sufficient for a disposition-ascribing statement.

The result of the two cases of the wire example can be generalized as follows. A disposition ascription is not equivalent to a conditional, because the disposition ascription

could be true while the conditional false and the disposition ascription could be false while the conditional true. In Martin's view, a conditional analysis of a disposition is "clumsy" and "inexact." Interfering factors such as an electro-fink machine might enter into causal contact with a disposition such as the electric power of a wire and render the disposition unable to manifest itself. Martin's primary concern is hence not only to point out the inadequacy of a conditional analysis of a disposition, but also to advance our realist intuition about a disposition.

Martin & Heil's emphatic distinction between a disposition and its manifestations has an important implication for their solution to the error problem. To follow a rule, as noted, is to possess a disposition of a certain sort, which is in turn characterized as one that occupies a particular position in a holistic web of reciprocal and inhibitory dispositional partners. For example, when put under a testing condition, an agent may not manifest a particular semantic disposition because of the absence of some reciprocal dispositional partners (e.g., the agent is deprived of oxygen for, say, two minutes) or because of the presence of some inhibitory dispositional partners (e.g., the agent has earlier taken some sleeping pills). Nevertheless, there still exists a rule-constituting fact, a real disposition occupying a certain position in the holistic structure of dispositions.

In Martin & Heil's metaphysical picture, therefore, errors occur because a particular disposition fails to manifest itself due to blocking factors, so are not at all critical or even relevant for the determination of the rule that is being followed. What is critical for rule-determination is instead the determination of the underlying disposition that is embodied by an agent. The disposition should not be equated with its successful or unsuccessful manifestations, or omissions. The fundamental mistake, according to Martin & Heil, of a

counterfactual analysis is that it confuses disposition with manifestations. As a result, the theorist wrongly looks for an epistemic criterion for classifying manifested behaviors into correct and incorrect, and thus faces the thorny error problem. The right way to determine the rule being followed, according to Martin & Heil, is to make a direct appeal to the web structure of a particular real disposition under scrutiny.

Martin & Heil's rejection of a counterfactual construal of a disposition exploits our metaphysical insight about what a disposition should be. Their metaphysical considerations are, in my view, insightful and plausible. They capture our important intuition that even if an agent is not put under a testing condition and hence does not manifest a behavior, the rule-constituting disposition is inherent in the agent and is there-and-ready for a suitable manifestation. This is similar to our intuition about, say, a bridge. Even if no car is passing over it, the bridge's disposition to support cars is inherent in the bridge and is there and-ready for a suitable manifestation. If what a counterfactual analysis does is merely to reduce a disposition to a pair (or a set of pairs) of testing conditions and manifestations, then the analysis is very likely to be wrong. We can easily imagine that a bridge still possesses its enduring capacities even though no cars ever drive over it.

Philosophers favoring a counterfactual analysis of the disposition, however, may argue that they do not deny that a bridge has an enduring capacity even though no cars ever drive over it. Rather, their claim is only that the disposition ascription "the bridge has the capacity of supporting cars at *t*, when no cars were driving over it" has to be analyzed in terms of the counterfactual conditional "If any car had been driving over the bridge at *t*, it would have held". To me, this argument is similar to Berkeley's treatment

of the existence of an unseen object. Berkeley does not deny that objects could exist unseen; he merely contends that the existence statement "An object O exists unseen at t" must be treated as being equivalent to the counterfactual conditional "If someone had been looking at t, he would have had O-ish experiences." The consequence of his treatment is to reduce the existence of an object to a counterfactual conditional, which amounts to denying that an object can exist independently of a potential perceiver who is to perceive the object under certain conditions. As Berkeley himself writes: "...the absolute existence of unthinking things without any relation to their being perceived, that seems perfectly unintelligible. Their *esse* is *percipi*, nor is it possible they should have any existence out of the minds or thinking things which perceive them."¹²³

Berkeley's counterfactual analysis of the existence of objects is, however, counter-intuitive. Although we may agree that the existence statement P: "There is a table in my room" has some logical relation to the counterfactual statement P': "If someone entered my room, he would or could have the experience of seeing a table," our commonsense intuition is that the second statement P' is the outcome of the first statement P. That is, we intuitively think that the existence of the table does not require its being perceived by an agent. Rather, the existence of the table is directly responsible for an agent's being able to perceive it under certain conditions, and the existence of the table also explains the truth of the counterfactual statement P'. Berkeley's semantic analysis of existence statements may be onto something, but the ontological consequences he draws from his analysis are misguided. Similarly, our realist intuition about a disposition should be suitably appreciated and kept salient. Contrary to counterfactual reductionists' claim, my

¹²³ Berkeley, G. *Principles of Human Knowledge and Three Dialogues*, 1948-57.

claim is, therefore, that ascriptions of dispositions are categorical, that they refer to real properties of objects which possess them, rather than counterfactual.

Rejecting a counterfactual construal of a disposition based on realist insights still leaves us with a puzzle: In what sense is a disposition real and actual, as Martin & Heil claim? If a disposition is not to be equated with a counterfactual conditional, then what can it be identified with? Is a disposition to be identified with the causal basis of an object or subject that possesses the disposition? Or does a disposition have a causal basis at all? If a disposition has a causal basis but is distinct from its causal basis, then what are their relations, and how do they interact? If a real disposition is not identical with its causal basis, then the disposition will not be causally efficacious. In that case, in what sense can it be real?

These are difficult and controversial questions.¹²⁴ I don't think they have to be answered in the context of responding to KW's skeptical paradox. At this stage I will stick to a minimalistic realist view about a disposition. Namely, a disposition is real in the sense that a disposition ascription is not to be reduced to or identified as a mere counterfactual conditional that describes manifestations under various conditions. A disposition is something that makes a counterfactual conditional true.

The real difficulty that Martin & Heil's account faces, however, is that the error problem does not go away. Martin & Heil solve the error problem by arguing that an error occurs because some intervening factors prevent a particular disposition from manifesting itself. So, in order to identify an erroneous act, we have to know whether the

¹²⁴ Major positions and discussions of the metaphysics of dispositions can be found in the following two excellent books: *Dispositions: A debate*, by Tim Crane (ed), 1996, and *Dispositions*, by Stephen Mumford, 1998.

act is performed absent blocking factors. How then are we to identify such a condition without having to presuppose what to count as blocking factors, and hence presuppose which real rule-constituting disposition is embodied?

Let me elaborate the difficulty a little more. The difficulty that Martin & Heil's view encounters is an epistemological one. Martin & Heil emphasize the distinction between an underlying disposition and its manifestations under certain conditions. However, they say relatively little about how we know what disposition an agent possesses, if he has one. We cannot point to a disposition and see it directly, because it is theoretical and not directly visible (on their account). So, if there is any way we can know what a disposition is like, there has to be some intimate connection between a disposition and its manifestations so that we may know unambiguously what disposition it is via the mediation of its observed manifestations under various conditions. But then the error problem emerges again. For only "suitably" (or correctly) manifested behaviors can tell us which disposition is embodied by an agent, where "suitably" means that the real disposition is not being blocked by the presence of any inhibitory dispositional partner or by the absence of reciprocal dispositional partner. The skeptic can reasonably doubt that we can decide without circularity whether a blocking factor is present. Because of the interlocking and interdependent relations between a particular real disposition and its dispositional partners, it does not seem plausible that one can identify a blocking factor without presupposing what a particular disposition is in place. As a result, the rule that is being followed cannot be determined without circularity.

A Solution to the Error Problem in Martin & Heil's Realist Dispositional Account of Rule Following

As we have seen in the last section, the error problem re-surfaces in Martin & Heil's realist account. We cannot know which disposition an object or subject possesses, unless we know whether certain behaviors are its suitable manifestations or not. To know the latter, we have to be able to tell whether certain inhibitory dispositional partners are present or certain reciprocal dispositional partners are absent. This proves hard to do, for we seem to be bound to presuppose that a particular disposition is in place when we claim that a blocking factor is or is not present.

My claim is that the error problem that is left unresolved in Martin & Heil's account can be solved by adding the following two supplementary points. The first point is that ascribing a real disposition to an object or subject is equivalent to an act of hypothesizing or theorizing. The second point is that such hypothesizing or theorizing is abundant in scientific practice, and it is perfectly legitimate to infer the underlying property of an object. In this section, I will argue for these two points respectively.

Hypothesizing about Particular Dispositions and the Resulting Problem of Vacuity of Provisoes and Ceteris Paribus Clauses

We have granted that the realist view of a disposition proposed by Martin & Heil is better than a view which reduces a disposition to a counterfactual conditional. We thus agree that a real disposition must be suitably distinguished from its manifestations. However, the underlying real disposition that is possessed by an object or subject is invisible and hence unknown. How then do we know what it is? Hempel's¹²⁵

¹²⁵ Hempel, "Dispositional Explanation", in *Dispositions*, by R. Tuomela (ed.), 1977.

characterization of a disposition by means of a deductive-nomological model of explanation may give us a clue.

Hempel (1977) gives a lucid explication of why ascribing a particular disposition to an object or subject can be viewed as a matter of hypothesizing and can figure in a scientific inference. According to Hempel, a theorist employs theoretical terms that stand for various dispositional properties, such as fragility, magnetism, and so on, depending on what the objects under investigation are. Theoretical statements, which contain theoretical terms, about what particular disposition an object or subject possesses are then made. Theoretical principles, normally expressed in the form of counterfactual conditionals, to which those ascribed dispositional properties are assumed to conform may also be constructed by a theorist. An argument is then formulated in which the premises are the theoretical statements, relevant theoretical principles, and initial conditions, and the conclusion is a statement about the object's predicted behaviors derived from these premises. For example, Hempel (1977) characterizes such an argument schematically as the follows:

(DN)

(C1) *i* was in a situation of kind S.

(C2) *i* has the (dispositional) property M.

(L) Any *x* with the property M, if it were in a situation of kind S, would behave in manner R.

(E) *i* behaved in manner R.

In (DN), (C1) describes an initial condition, and (C2) contains a theoretical term M

that ascribes a dispositional property to *i*. (L) states a theoretical principle, which contains M, in the form of a conditional $(x)(Mx \rightarrow (Sx \rightarrow Cx))$. And finally, (E) is a conclusion stating the behavior of *i*. The derivation of (E) has a deductive-nomological (D-N for short) form, which allows the object *i*'s actual or potential behavior to be explained or predicted. Such an argument in a DN form helps us see how ascribing a disposition does its explanatory work.

For example, let's use the magnetic theory to illustrate how a DN dispositional explanation of an empirical phenomenon works. Suppose we hypothesize that an object, say, an iron bar *b*, is magnetic and hence ascribe to it the magnetic property M. Suppose T is a simple theory of magnetism that contains theoretical terms such as "magnet", "north pole", "south pole", and so on. T also lists a few theoretical principles, such as (L1) "for any object *x* with the magnetic property M, iron filings would cling to *x*, if they were placed close to *x*". (L2) "for any object *x* with the magnetic property M, if *x* is broken into two shorter bars and these are suspended by long thin threads close to each other at the same distance from the ground, they will orient themselves so as to fall into a straight line", and so on. Supposing now that we place the iron bar *b* close to some iron filings, *b*'s actual or potential behavior of attracting those iron filings will be explained or predicted by the following argument (DN'):

(DN')

(C1) Iron filings were placed close to the iron bar *b*.

(C2) *b* has the magnetic property M.

(L1) For any object *x* with the magnetic property M, iron filings would cling to *x*,
if they were placed close to *x*.

(E) Those iron filings clung to the iron bar b.

It is worth noting that in adopting Hempel's DN model of dispositional explanation, there is a question whether stating the theoretical principle (L) as a conditional involves reducing the disposition ascription (C2). Hempel is not fully explicit on this. However, even if he holds a reductivist view, we can still, in my view, construe a disposition ascribed by (C2) realistically, and regard (L1) as a consequence of an object possessing M. What is crucial and relevant in the context of solving KW's error problem is this: Hempel's DN model of dispositional explanation helps us see that ascribing a disposition to an object is equivalent to forming a hypothesis like (C2) and that a conditional like (L1) follows from an object's having the particular real disposition that the hypothesis (C2) is about.

Hempel's characterization of a dispositional explanation of an empirical phenomenon can be similarly applied to a realist dispositional account of rule-following. Recall that the skeptic raises an epistemological doubt as to how we know which real disposition a subject possesses. Adopting Hempel's DN model of dispositional explanation shows that one can regard the answer to the skeptical question a hypothetical business. A theoretical statement is first made that ascribes a dispositional property to a subject. For example, a subject *s* is ascribed a disposition to use the sign "+" in accordance to the plus rule. Theoretical principles, expressed by counterfactual conditionals, such as "A subject *s* would answer the sum *r*, if he were queried with the question '*m + n*'", can then be constructed to explicate what the ascribed disposition is. The rule-following behaviors of the subject can then be explained or predicted by a

deductive-nomological argument whose premises include the theoretical statement that ascribes a rule-following disposition, a theoretical principle that describes what the particular disposition is, and initial conditions about the subject. This can be represented as follows by (DN*):

(DN*)

(C1*) A subject s was presented with the question " $m + n = ?$ ".

(C2*) s has the disposition P with regard to operating the sign "+".

(L*) Any subject x with the disposition P , if he were presented with the question " $m + n = ?$ ", would answer the sum r .

(E*) s answered the sum r .

Before I continue to discuss the problem involved in the Hempelian DN model of dispositional explanation, it is worth noting that Hempel's DN model is employed here only because it provides a convenient and lucid format to illustrate how disposition ascriptions are explanatory and hypothetical. If there are other models of explanation that serve equally well or better than Hempel's model, I could as well use them, as far as they also help illuminate the discussion of a dispositional account of rule-following. Hence, no tight relation is intended between a disposition and a DN model of explanation.

The logic of my argument is hence to view a disposition as a theoretical entity, which can be known by inductive inference based on the observation of behaviors. Hempel's DN model of explanation is useful for illustrating it. KW, however, might raise an objection at this point, that viewing rule-following as consisting in a theoretical disposition is very unnatural. For rule-following is not a theorizing enterprise which

requires inductive reasoning. Rather, it is an activity that does not seem to be mediated by any inductive inferences. My response to this objection is that contrary to what KW might think, construing rule-following as a hypothetical disposition is very natural. We do ascribe dispositions to people to explain and predict their rule-following behaviors. The same thing can be said in the first-person case as well: it is by virtue of the fact that I possess a certain disposition that the production of my rule-following behaviors is accounted for. Therefore, a hypothetical disposition does seem to be a perfectly legitimate and natural candidate to be considered as the fact that constitutes rule-following.

The element of unnaturalness KW has in mind is actually the first-person epistemic problem, which concerns how a subject can ever know in a non-inferential manner his own rule-following disposition, which is being construed as a theoretical entity and hence can be known only through some form of indirect inductive inference. I grant that this is an important question. I will deal with it in the next chapter.

Hempel (1988) points out, however, that a DN model of dispositional explanation of an empirical phenomenon faces the problem of provisoes, which, as I will show shortly, is the underlying source of the error problem for the realist dispositional account of rule-following. The problem of provisoes has to do with the fact that there are cases in which all the premises of a DN argument are true while the conclusion false, and thus the DN argument is invalid. Let's again take the theory of magnetism. We may grant the truths of the disposition ascription (C2) and theoretical principles regarding magnetic properties, and we may also suppose that the initial condition (C1) is also met. There are, however, cases in which the iron filings do not cling to an iron bar, thus conflicting with

the predicted result (E). This conflict between an actual observed result and a theoretically derived statement may be due to the fact that there is some other stronger magnetic force present in the vicinity of the iron filings which deflect them. It may also be due to the fact that the iron bar suddenly undergoes a physical change such that it loses its magnetic power and is hence unable to attract those iron filings. There are many other potential factors that might prevent the iron filings from clinging to the iron bar when the two are put close to each other. However, the theory of magnetism itself does not ensure that those potential disturbing factors are absent. Thus, the inference DN' model has to presuppose as a premise that no such disturbing factors are present; or, to state the premise in question more positively, namely, the behaviors of an iron bar and of those iron filings put next to the iron bar are subject *only* to the magnetic forces they exert upon each other. Hempel calls such a premise a "proviso". A proviso is not explicitly stated, but is essentially employed in a DN dispositional explanation of an empirical phenomenon.

One of the major problems, as noted by Hempel (1988: 153-4), that results from the need for provisos is that a theory is not falsifiable by observation. For whenever an object's behavior is observed to differ from the predicted result, one can always claim that the implicit provisos are violated (i.e., that some disturbing factors are present). It is clear that the problem of unfalsifiability is a result of the fact that provisos are only vaguely expressed. Unless a criterion can be given as to when a proviso is satisfied and when it is violated, a dispositional explanation will never be falsifiable because it inherently contains an unstated escape clause. The error problem that arises in Martin & Heil can thus be characterized as the problem of provisos in Hempel's DN model of

dispositional explanation.

So far, I have shown that if following a rule is regarded as constituted by a subject's having a real disposition, on Hempel's DN model of dispositional explanation identifying a subject's disposition is equivalent to making a theoretical statement. (DN*) includes a theoretical statement about a subject's having a particular rule-following disposition P, a theoretical principle describing the disposition P, and an initial condition about the subject. Based on these three premises, a conclusion regarding the manifestation of the disposition can be drawn. Thus, prediction or explanation of a subject's rule-following behaviors seems to become possible. Nevertheless, the observed rule-following behaviors of a subject may "deviate" from the DN* prediction. A proviso thus has to be implicitly inserted in the premises of the argument to guarantee that the "deviations" of the subject's actual behaviors from the predicted behaviors are the products of blocking factors. The vacuity of provisoes and a lack of criteria for the satisfaction and violation of provisos may render a hypothesis that ascribes a disposition to a subject impossible to confirm or disconfirm. As a result, which rule a subject follows remains undetermined.

The rule-underdetermination problem can be put in an alternative way. Given that the disposition ascription (C2*) does explain observed behavior, it seems that we have inductive evidence for the theory that the subject *s* has the rule-following disposition P. However, because of the proviso problem, we cannot say definitely what C2* does and does not explain. As a result, skepticism about the reality of rule-following stays alive.

It is clear by now that the error problem for a realist dispositional account is caused by the problem of vacuity inherent in provisoes. If we can solve the problem of vacuity, the error problem may be resolved. In the next section, we will see how to do so. But

before we do that, one thing is worth noting. The function of provisoes in theoretical inferences, such as those characterized by a DN model, is similar to that of *ceteris paribus* clauses that are widely employed to protect laws of special sciences from counterexamples. With a *ceteris paribus* clause attached to a law, counterexamples to the law can be claimed to occur because *other things* are not *equal*. Adding a *ceteris paribus* clause to a law is tantamount to making an assurance that when disturbing factors are absent, a nomological pattern of behaviors will be observed as predicted. Provisoes are also similar to normality conditions typically added to a traditional counterfactual conditional that is to analyze a disposition ascription. Normality conditions are those under which no disturbing or intervening factors enter into causal relation with the particular disposition under scrutiny. Therefore, whenever an object's or subject's disposition manifests itself in a way which conflicts with what the counterfactual conditional describes, it can simply be claimed that conditions are not normal, and the manifested behaviors cannot be attributed to the normal functioning of the disposition. It is clear then that the problem of unfalsifiability that confronts Hempel's DN model of dispositional explanation also confronts a *ceteris paribus* law, and a counterfactual conditional supplemented with normality conditions. The difficulty in each of these cases arises from the fact that there is serious vacuity inherent in a proviso, a *ceteris paribus* clause, or a normal condition. Thus, a solution to the problem of vacuity in any one of the three cases applies to the other two as well.

A Solution to the Problem of Vacuity of Provisoes, Ceteris Paribus Clauses, and Normality Conditions

The key to solving the problem of vacuity lies first in seeing the important role

played by provisoes or *ceteris paribus* clauses in scientific theorizing. There are certain properties or relations in the world that are nomologically significant and are of interest to us or to scientists. However, the world is so complex that without suitable abstraction or idealization no such nomologically interesting properties can be uncovered. Take for example the theory of magnetism that we discussed earlier. The theory tells us that an iron bar has a magnetic property, and that it would attract iron filings if they were placed close to it. In reality, however, there are many cases where the predicted magnetic phenomenon would not be observed. The iron bar might lose its magnetic power, or those iron filings might be glued to the plate. Numerous factors can enter into causal contact with the iron bar and the nearby iron filings so that the magnetic force they exert upon each other cannot be successfully manifested. But failing to observe the magnetic effect does not mean that the magnetic force does not exist. It is the presence of some interfering factors that prevents the magnetic property of the iron bar from manifesting itself. Due to the potential interference of multifarious causal factors, *ceteris paribus* clauses (or provisoes) thus have to be introduced into statements about laws to the effect that if those interfering factors are absent, certain nomologically significant properties or relations exist. Conversely, if a theorist has to specify all the potential intervening factors in order to secure his theory about the nomological properties under scrutiny, then the theory would be almost impossible to come by. For the world is too complex for the theorist to name all those potential intervening factors.

Pietroski & Rey¹²⁶ argue for the importance of *ceteris paribus* clauses in science. In their view, *ceteris paribus* clauses are indispensable tools for the abstraction and

¹²⁶ Pietroski & Rey (1995), "When other things aren't equal: saving *ceteris paribus* laws from vacuity". *British Journal of Philosophy of Science*.

idealization much needed in scientific theorizing. Take another law of physics; this one does not involve dispositions. Boyle's law of gases tells us that there is a nomologically important relation between the pressure, volume, temperature, and a fixed number of moles of the gas ($PV = nRT$). The law says that for a fixed quantity of gas, if the temperature of the gas remains constant, then the gas' pressure varies inversely with its volume. However, Boyle's law of gases holds only given certain conditions, e.g., the electrical attraction between molecules is absent, molecules do not take up space, containers of the gases are inelastic, and so on. *Ceteris paribus* clauses thus have to be implicitly attached to Boyle's law to ensure that when and only when those conditions are met, will certain nomological truths about gases hold. In other words, *ceteris paribus* clauses are used as a tool for idealizing an object under investigation and its environment in such a way that certain nomologically significant properties can be abstracted from our complex world. Accepting Boyle's *ceteris paribus* law of gases, accordingly, entitles one to explain the "deviant" behaviors of a gas from the law's predictions by citing those factors ignored when one only considers the pressure, volume, and temperature of the gas.

Conversely, the vindication of Boyle's theory would be impossible if it requires that all those numerous unstated conditions be clearly specified. The reason is that the counterfactual conditionals which Boyle's theory supports will become enormously complicated, and the corresponding initial conditions will become too deviant from conditions in the real world to be actually fulfilled. Thus, the employment of *ceteris paribus* clauses is unavoidable when a theorist generates statements about laws. The consequence, however, is that the problem of vacuity in *ceteris paribus* (cp) laws is

equally inescapable.

Pietroski & Rey stress the indispensable role played by cp-clauses in scientific theorizing, but they also recognize that vacuity of cp-clauses is ineliminable. They thus offer a sufficient condition for the non-vacuity of cp-laws. If their attempt is successful, then the problem of unfalsifiability inherent in scientific theories can be solved, and so can the error problem in a dispositional account of rule-following. I will explicate their proposal below.

The core of Pietroski & Rey's proposal is that, if all apparent counterexamples to a cp-law can be explained by *independent* factors that are actively interfering with the nomologically causal relation described by the law, then the cp-law is non-vacuous. Independent interfering factors are independent in the sense that, roughly, they can be explained by theories that invoke laws other than the one under investigation (we will go into this in more detail shortly). Pietroski & Rey express their proposal metaphorically. "cp-clauses are cheques written on the bank of independent theories."¹²⁷ Just as writing a cheque carries a commitment that the bank where the cheque is deposited will cash it, adding a *ceteris paribus* clause carries with it an explanatory commitment that other theories than the one being offered will explain counterexamples to the law. In Pietroski & Rey's "cheque-cashing" proposal, therefore, the substance and warrant of a *ceteris paribus* clause are derived from the substance and warrant of those other theories. If the *ceteris paribus* clause of a law can be "cashed" by other theories, then the cp-law is a good one. If not, then the cp-law is genuinely vacuous and better jettisoned.

In Pietroski & Rey's sufficient condition, it is crucial that counterexamples or

¹²⁷ Pietroski & Rey (1995), p.89.

exceptions to a cp-law must be explainable by citing *independent* interfering factors. But what exactly does the notion of independence mean here? We may take clairvoyance for example to illustrate it. Some people may believe that the power of clairvoyance exists, and accept the following cp-law as true: cp (if medium x predicts p, then p)¹²⁸. Suppose that experiments are conducted to test whether the cp-law in question is true, and it turns out that people claimed to be clairvoyant fail to exhibit clairvoyant power in most cases. Suppose further that no obvious interfering factors, such as subjects' being distracted, drunk, tired, etc., are observed in those experiments that can explain the subjects' failures, and as a result ectoplasmic interferences are appealed to. The appeal to ectoplasmic factors, however, obviously has the following problem. We do not have a respected scientific theory about ectoplasmism in which other phenomena than the failures in clairvoyance experiments are explained. Ectoplasmic factors thus seem to be a mere concealed way of tautologizing the law about clairvoyance. The requirement in Pietroski & Rey's proposal that interfering factors be "independent" thus means that those factors not only explain the occurrence of counterexamples to the law under scrutiny, but they can also do some explanatory work in some independently vindicated theories. Hence, if counterexamples to a cp-law can be explained by independent interfering factors in the above sense, the cp-law is not vacuous.

The clairvoyance case shows that if factors cited to explain counterexamples to a cp-law are subsumed under ill-warranted theories or under no theories, then the cp-law has to be regarded as vacuous. On the contrary, if counterexamples to a cp-law can be explained by factors described by well-warranted and substantive theories, then the cp-

¹²⁸ See Pietroski & Rey (1995), p. 90.

law is non-vacuous. An interesting case in the history of science can help us see how Pietroski & Rey's proposal works. Newton's gravitational theory had been successful in explaining and predicting objects' motions until it encountered certain perturbations in the orbit of Uranus; there the theory's predictions persistently failed. It was conjectured that a proviso in Newton's theory was violated, which means that Uranus was subject to some additional force that was not taken into consideration in the original calculation of Uranus' movement. A hypothesis was thus proposed that there existed an unknown planet that was responsible for exerting the additional force upon Uranus. The hypothesis about the existence of an unknown planet was confirmed when the planet Neptune was subsequently discovered. Moreover, the existence of Neptune also explained observed optical phenomena independently of the deviations of Uranus' orbit. As a result, Newton's gravitational theory is not vacuous, according to Pietroski & Rey's proposal, because interfering factors that were cited (i.e., the existence of Neptune) not only explained counterexamples to the theory (i.e., deviations of Uranus' orbit), but also explained other phenomena. These interfering factors are subsumed under a theory that is substantive and independently warranted. On the other hand, if all the efforts to confirm the "hypothesis of Neptune" had repeatedly failed, then the deviations of Uranus' orbit would have constituted genuine counterexamples to Newton's theory and consequently the cp-clause it contained would have been genuinely vacuous. We would then have a reasonable ground to reject Newton's gravitational theory as useless.

Note that Pietroski & Rey's proposal allows that the interfering factor in a law of a theory T can be conjoined with the very law itself to account for the counterexamples to T. As was shown in the case of Newton's theory, the existence of Neptune was cited as

an interfering factor. In conjunction with the gravitational law in Newton's theory, it explained why the counterexample to Newton's theory occurred, i.e., why Uranus' orbit "deviated". In most cases, however, interfering factors that are cited to explain counterexamples to a theory are subsumed under theories other than the one under investigation. For example, suppose one found an object that fell freely with an acceleration that deviated from the predicted value derived from Newton's gravitational theory and second law of force. Careful studies may be carried out and show that there existed a strong magnetic force in the surroundings of the free-falling object to which it was subject. In this case, the explanation of the counterexample to Newton's theory, i.e., the deviant acceleration of the object, has to be offered by a theory different from Newton's theory, namely, the theory of magnetism. It is crucial that the theory of magnetism on which the explanation of the counterexample is based be able to explain other phenomena and to be itself well-warranted. Only so can the proviso or cp-clause inserted in Newton's gravitational theory be cleared of the charge of vacuity.

Pietroski & Rey, in my view, has offered a plausible sufficient condition for the non-vacuity of cp-laws. The main idea is that cp-clauses or provisos carry with them promises or commitments that counterexamples to a theory can be explained by (mostly) other theories that are themselves substantial and well warranted. If it turns out that no such theories are available after our repeated efforts to find them, then the theory with a proviso or cp-clause attached would be construed as genuinely vacuous. We would then disregard the theory.

Pietroski & Rey's proposal has an important bearing on solving KW's error

problem¹²⁹. As we have seen earlier, the error problem for a realist dispositional account of rule-following concerns how a particular real disposition claimed to constitute a subject's following a particular rule can be identified without circularity. More specifically, the difficulty has to do with the fact that a real disposition is invisible. Thus, the way to discover which real disposition a subject possesses is via inferences from the subject's manifested behaviors in various conditions. This inference task requires that we distinguish behaviors that are "suitable" manifestation of a disposition from those that are not, where "suitable" means that no interfering factors are present. But how are we to decide whether or not interfering factors are present? According to KW's skeptic, any attempt to claim that an interfering factor has occurred and hence an erroneous behavior has been produced would necessarily presuppose that a particular real disposition is in place and thus renders the account circular.

My proposal is that the error problem in a realist dispositional account of rule-following can be solved as follows. Hempel's model of dispositional explanations of rule-following phenomena shows us that ascribing to a subject a particular disposition to follow a particular rule is a matter of hypothesizing in a way that is similar to scientific theorizing. That is to say, a hypothesis is formed regarding which particular disposition a subject possesses based on the observation of the subject's relevant rule-following behaviors. When such a hypothesis faces counterexamples because the subject's actual or potential behaviors "deviate" from the predictions of the hypothesis, an implicit proviso or a *ceteris paribus* clause is employed to account for these deviations.

Furthermore, the problem of vacuity for a proviso or *cp*-clause can be satisfactorily

¹²⁹ P & R also apply their proposal to KW's error problem. See especially section 3.3 of P & R (1995). My aim here is to pursue and elaborate on their idea.

solved from Pietroski & Rey's view. Insofar as interfering factors cited to explain a subject's "deviant" rule-following behaviors are subsumed under a substantive and warranted theory, the hypothesis ascribing a real disposition to the subject should not be charged with vacuity or circularity. For example, a subject may answer "5" or "122" or another number when queried " $68 + 57 = ?$ ". The fact that the subject has been hit on his head can be reasonably cited to explain why those performances are "deviant" or "erroneous" relative to a real disposition to follow the plus rule. Meanwhile, substantive and warranted theories of neurophysiology are available to explain how those deviant performances of calculation may occur when one's head was hit, in addition to explaining many other phenomena. All other seeming counterexamples to a disposition-to-follow-a-rule hypothesis (i.e., seemingly "erroneous" behaviors) can be treated similarly. Namely, interfering factors will be cited to account for their apparent deviation from the underlying disposition and they must be subsumed under well warranted theories in order for a disposition-ascribing hypothesis to be non-vacuous. On the other hand, if repeated efforts to find an interfering factor, which is subsumed under a substantive and warranted theory, have failed, the "deviant" behavior would have to be regarded as a manifestation of a disposition that differs from the one hypothesized. The original hypothesis then should be given up in favor of another.

The account Pietroski & Rey offers fits quite well with Martin & Heil's realist account of a disposition. In Martin & Heil's account, a real disposition occupies a particular position in a holistic web in which many other dispositions are related to the disposition under investigation in a complicated way. In such a web, the disposition under scrutiny has inhibitory and reciprocal dispositional partners. Each of those

inhibitory and reciprocal dispositional partners is in turn itself a disposition that has its own specific web-structure characterizable by its various inter-related dispositional partners. Moreover, each of the inhibitory and reciprocal dispositional partners of the one under examination can belong to very different kinds and can enter into causal contact with the disposition under scrutiny. For example, a subject's arithmetic performance can be affected by factors such as the amount of oxygen supply to the subject, whether or not he takes sleeping pills, whether or not he receives a heavy blow on his head, the amount of time he has spent on performing a long calculation continuously without having a break, and numerous other factors. All these interfering factors are proper objects of vigorous studies in various scientific domains, such as neuroscience, physiology, cognitive psychology, and so on.

Pietroski & Rey's "cheque-cashing" proposal then makes perfect sense within this realist web picture of dispositions. A proviso or cp-clause is inserted in a disposition-ascribing hypothesis to deal with counterexamples (i.e., "erroneous" behaviors that deviate from the hypothesis' predictions). Adding a cp-clause or a proviso, according to Pietroski & Rey's proposal, amounts to making an explanatory commitment that there exist interfering factors causing those deviations, and that there are or will be scientific disciplines to study how those interfering factors affect a subject's rule-following behaviors. Studying interfering factors in various scientific domains is equivalent to uncovering the web-structure of the subject's rule-following disposition. It is clear that Martin & Heil offer a plausible metaphysical account of dispositions, and Pietroski & Rey provide a sound methodology with which the empirical reality of dispositions can be investigated in a scientifically respectable way. The result is that a subject's real

disposition, claimed to constitute his following a particular rule, can be determined non-vacuously and non-circularly. Consequently, the error problem that emerges in Martin & Heil's realist dispositional account of rule-following is satisfactorily solved.

How might Pietroski & Rey's account fare if KW's skeptic insists that cp-clauses and provisoes are hopelessly vacuous, or, "hedged", that no sufficient condition for non-vacuity can be offered, including Pietroski & Rey's proposal? Indeed, KW rejects any theory or hypothesis that employs a cp-clause or proviso to ascribe a disposition. However, the problem with rejection of a cp-clause or proviso is that KW would have to face the undesirable consequence that he has to call into doubt almost all respectable sciences, including physics. For not only do special sciences, such as psychology, linguistics, economics, and so on, rely on cp-clauses or provisoes extensively to establish causal laws, but very basic physical theories, such as Newton's theory of gravitation, theory of magnetism, Boyle's theory of gases, etc., also implicitly embody provisoes or cp-clauses, as we have seen in earlier examples. There is serious doubt whether there is any current fundamental science that is free of cp-clauses or provisoes¹³⁰. Even if there were, they would be extremely rare. As a result, if KW's skeptic rejects the legitimacy of cp-laws in the practice of science, he will inevitably rule out almost all of our knowledge about the world. This consequence renders KW's reply seriously deficient. Also, it trivializes KW's position as a form of inductive skepticism.

KW would insist on the distinction between basic sciences and special sciences. One way he can do this is to make a distinction between a theory and its application. The theory of gravity *per se* has no cp-clause, for it only claims that any two masses attract

¹³⁰ Cartwright is noted for her claim that "the laws of physics lie". See N. Cartwright (1983), *How the Laws of Physics Lie*.

each other in a certain way, period. It is only when the theory is used to predict the motion of specific bodies that a cp-clause becomes necessary. The distinction between a theory and its applications, KW would claim, does not seem to apply to theories in special sciences, especially in the branch of psychology that ascribes rule-following dispositions to people. If the distinction in question can be sustained, then the problem of vacuity of cp-laws only applies to special sciences but not to basic sciences. KW's dismay at employing cp-clauses in psychological laws of rule-following would then revive.

We may well recognize that there exists the distinction in question between basic sciences and special sciences. Nonetheless, I would make two points in reply. First, even if we grant that only theories in special sciences employ cp-clauses, Pietroski & Rey have offered us a plausible way to formulate the sufficient condition for the non-vacuity of cp-theories. It would thus be methodologically legitimate to use those cp-theories that meet the non-vacuity criterion. Second, theories in special sciences, such as biology, linguistics, psychology, economics, and so on, do provide, at least sometimes, as good explanations and predictions for the human world as basic sciences for the physical world. It is not clear how KW might explain the explanatory and predictive power of theories in special sciences, unless he is able to recognize that theories in special sciences may indeed pick out nomologically significant properties in the world. Therefore, the laws or dispositions referred to in the psychology of rule-following may well be properties as real as those picked out by theories in basic sciences.

There is also a possible regress problem that KW's skeptic would raise against Pietroski & Rey's proposal. What if the independent theories under which the cited

interfering factors are subsumed are themselves cp-theories? Will they depend on yet further theories, which might contain cp-clauses? Must we eventually get to theories that are wholly strict, not involving any cp-clauses? If there are indeed such theories, why don't we just use them and jettison cp-theories?

Even if we grant KW's worry that independent theories themselves contain cp-clauses and that they might in turn rely on further theories that also employ cp-clauses, my claim is that cp-theories remain methodologically sound. The reason is that cp-theories in a science network are mutually corroborative and coherent. A cp-theory cannot contradict the evidential results produced by other cp-theories. Otherwise, the credibility of the cp-theory would be undermined. Conversely, if a cp-theory receives and continues to receive evidential support from other cp-theories, then the cp-theory is very likely to be true. The same criterion applies to all other cp-theories as well. They themselves have to be coherent with other cp-theories in order to be supported and to gain credibility. This process of seeking corroboration in a knowledge network ensures that cp-theories are subject to critical and sophisticated examination. Consequently, the regress problem about cp-theories that arises in special sciences in general and rule-following in particular can be adequately handled.

But what if all theories contain cp-clauses? Would there be any independent confirmation? I tend to favor the view that at least some theories in basic sciences do not employ cp-clauses. As a result, cp-theories in special sciences can be ultimately grounded in them without running in an epistemic circle. As to the question of why we don't simply jettison cp-theories and use theories in basic sciences, my answer is that there may be nomologically significant properties in the world that can be picked out

only by cp-theories in special sciences, rather than by theories in basic sciences. As Fodor¹³¹ has forcefully argued, psychological properties are similar to, say, the monetary property in that they both cannot be described by vocabularies at the physical level. The monetary property can be realized by different physical material, such as papers, gold, diamond, etc. But then for any laws at the physical level to capture what all those possible realizations share in common, the laws will become too wildly disjunctive to be genuine laws. The idea to reduce all real properties to physical properties and to accordingly abandon cp-theories in special sciences in favor of theories in basic physical sciences may thus be simply a prejudice of positivist philosophy of science.

To emphasize, I do not have to offer an explicit specification of a cp-clause in ascribing a rule-following disposition. All I have to do is to argue that a dispositional ascription and explanation can be non-vacuous. Since a suitable sufficient condition of non-vacuity can be offered, the epistemological difficulty in the error problem is resolved.

Realism for a Disposition to Follow a Rule

Thus far I have shown that the fact of the matter that constitutes rule-following can be explained by Martin & Heil's realist dispositional account. The error problem in such a realist account becomes the problem of how an underlying real disposition can be identified, and I have also argued that identifying which underlying disposition a subject or object possesses is an act of hypothesizing of the sort that is essential in scientific practice. A major worry about such an enterprise of hypothesizing or theorizing is its vacuity due to *ceteris paribus* clauses or provisoes it employs. However, the problem of

¹³¹ See Fodor, "Special Sciences, or The Disunity of Science as a Working Hypothesis," in Block (1980).

vacuity of cp-laws is solved by Pietroski & Rey's sufficient condition: that an interfering factor cited to explain the violation of a cp-clause can itself be independently explained by a substantial and warranted theory. As a result, when this condition is met, the rule that is being followed by a subject can be determined in a way that is epistemically grounded and methodologically legitimate. Hence, no skeptical worry about the reality of a rule-following disposition can be reasonably sustained.

However, if it is a real disposition, rather than its manifested behaviors, that is to be identified in the determination of the rule being followed, then a natural question to ask is how such a real disposition ever comes into existence in the subject. Presumably, the question of the ontogeny of the real disposition to rule-following may be characterized by a person's innate cognitive constraints and his learning history. If we can determine which real disposition gets established during the learning stage, then we can settle on which real disposition a subject possesses in the after-learning stage, and hence determine which rule he follows. In this section, I will further explore the real nature of the disposition to rule-following from a learning perspective. Specifically, my idea is that by attending to the dimension of learning, we may come to see what factors are crucial in the determination of the acquisition of a real disposition. As a result, we may obtain a clearer and more complete picture of how we identify rule-following dispositions.

In chapter VI, I have introduced Dretske's learning stages account to resolve the error problem. Although I have argued that Dretske's account fails to respond adequately to some of Fodor's criticisms, it retains genuine insight worth further pursuing. I will first rehearse the insight inherent in Dretske's account and the problems that remain unresolved in it. I will then show how those problems can be solved, and that the rule-

following phenomenon can be located and depicted in a broader context than the one in which KW's error problem is formulated.

Insight in Dretske's Distinctive Learning-Stage Account: A Disposition Is Acquired Through Learning

On Dretske's dispositional account of meaning and rule-following, meaning something by a term is viewed as consisting in a nomic dispositional term/world correlation. Such a nomic dispositional correlation is established through learning. Once a particular word/world nomic disposition, say, a disposition D to apply "cow" to cows, is established in an agent during the learning stage, D can be mistakenly triggered by something similar to but different from cows, such as horses at night, in the after-learning stage. An error thus occurs. The insight inherent in Dretske's account is that errors can occur and be attributable to an agent only in the after-learning stage, because only in the after-learning stage can the established nomic correlational disposition be triggered by something that lacks the appropriate piece of "information." Hence, he denies the skeptic's claim that it is impossible to vindicate the reality of rule-following because it is impossible to identify an erroneous rule-following behavior non-circularly. In Dretske's account, the crucial issue in the determination of the rule being followed is not the identification of erroneous behaviors. Rather, what is critical is the determination of the rule being learned during the learning stage.

The point that errors can exist only in the after-learning stage is well-taken. However, as we have pointed out in the last chapter, Dretske's account faces two major problems. The first problem, raised by Fodor (1990), concerns the lack of a distinctive

point at which a person finishes learning. Dretske disposes of this problem by calling it an instance of the Sortes fallacy.

The second problem that Dretske's account faces concerns how we determine which real, nomic disposition is established during the learning stage. In Dretske's account, this is determined by the pattern of input stimuli and reinforcement contingencies a trainee receives from the learning environment. For example, a teacher would give instruction and feedback in such a way that a trainee utters "cow" only in the presence of cows and not other things. In that case, the trainee would acquire a disposition to apply "cow" to cows only and hence mean *cow* by "cow". But as we mentioned in the last chapter, Fodor points out that Dretske's account ignores relevant counterfactual situations. For example, a trainee could have encountered a horse at night, and could have uttered "cow" in that situation. In such case, we can as well say that the trainee means *horse or cow* by the word "cow." The disjunction problem or error problem recurs.

Dretske could have argued that in the counterfactual situation just described, if the trainee had encountered a horse at night and had uttered "cow", he would have been corrected by the teacher. However, Fodor then raises a further difficulty. Which rule or meaning a trainee learns (or which nomic disposition a trainee acquires) does not depend merely upon those objective reinforcement contingencies provided by the trainee's environment (primarily the teacher). Rather, it depends on the teacher's pedagogical *intentions*, and thus *meaning*, when he rewards or punishes learning behaviors of the pupil. As a result, contentful states, such as the teacher's intentions or meanings, would have to be employed in the specification of the rule being learned. Dretske's account would thus be circular.

I replied that that the difficulty is only superficial. For the teacher's feedback to a trainee does not necessarily have to be characterized in intentional terms. It can be described in a purely objective manner, such as in terms of a distribution of actual and counterfactual punishments and rewards. This way of extensionalizing teaching, however, faces a problem. It may be asked: Doesn't the teacher have to *recognize* that the thing is a cow, when he gives suitable feedback to the trainee? If so, the teacher's intentional states have not been eliminated.

My response is that the real problem in Dretske's learning-stage account does not come from the worry of whether or not external stimuli, including a teacher's intentions, in a learning environment can be expressed in terms of purely objective counterfactuals. Rather, the real problem comes from how those external stimuli in a pedagogical context are *taken* by a trainee. It is assumed under the current dialectic that the teacher is a competent speaker and that he is trying to convey the meaning of a word or the intentional content of his thought to the trainee. The difficulty is how and whether or not this is achieved. Whatever is said or done by the teacher, it could be taken by the trainee in a wide variety of different ways. How then is it determined which concept or rule the trainee learns in a pedagogical context? Therefore, the core issue involved is not to extensionalize the teacher's intention. The real challenge is to explicate how and whether or not the trainee picks up a determinate concept or rule that the teacher tries to convey to him. As we have seen earlier, Fodor notices this point, and raises the same problem.

To save Dretske's account, therefore, it is necessary to posit that what an agent learns in the learning stage does not wholly depend on characteristics of purely external stimuli presented to the agent by the environment. Inner characteristics of an agent's

cognitive mechanisms that are responsible for learning are crucial in the determination of what the agent eventually learns. If we can bring into consideration the inner constraints on an agent's cognitive mechanisms and conduct suitable investigations on how they influence concept learning, there may be a chance to resolve the underdetermination problem.

My plan in what follows is to offer a solution to the underdetermination problem. My purpose is that, by offering a solution to it, we may see more clearly in what way a rule-following disposition is real, and what important characteristics it possesses. The pursuit of this study will, hopefully, enable us to understand a disposition not merely in terms of hypothetical explanation and various testability conditions, but also in terms of the kind of realism a rule-following disposition embodies. The result of this work will hence be a unified treatment of a rule-following disposition.

Inner Cognitive Constraints on Human Learning: Solving the Underdetermination Problem in Rule/Concept Learning.

Let me restate the problem of learning that we are facing. Given a finite set of closed feedback loops and instructions presented to a subject by his teacher and other resources such as textbooks or parents, there is a potentially infinite number of meanings/concepts/rules that are mutually incompatible but fit equally well with the stimuli. All these concepts or rules can be consistently said to be conveyed to a pupil by a teacher's feedback. For example, when the teacher says to a pupil that 2 plus 2 equals 4, the teacher might be teaching the plus rule or the quus rule as well. So, naturally, we would ask, which rule is learned by the pupil?

The underdetermination problem of learning is explicitly considered by Kripke and regarded by him as the major problem to those who take the Dretskean learning approach. Kripke first notices¹³² that the predicate “grue”, which was invented by Goodman (1954), plays a role similar to “quus”. The predicate “grue” contains a peculiar time factor as follows: x is grue = either (x is green and observed before time t , say, midnight on 31 December 2002) or (x is blue and observed after time t , i.e., midnight on 31 December 2002). Both “grue” and “quus” are “bent”¹³³ predicates whose meanings are wildly different from that of our normal predicates like “green” and “plus”. Some have contended that “bent” predicates are not suitable for projection because “bent” predicates are more complicated than normal predicates. This is due to “bent” predicates’ being “temporal” or “positional” while the latter is not. However, Goodman insists that there is no sense that does not beg the question in which “bent” predicates are more complicated than normal predicates. For if we initially regard “bent” predicates like “grue” as primitive, and define normal predicates like “green” accordingly, “bent” predicates such as “green” would become as “temporal”, or “positional” as normal predicates like “grue” (“green” could be defined as grue before midnight on December 2002 and bleen after midnight on December 2002).

Kripke himself comes up with an intuitive response to the “grue” paradox as well as to his own skeptical paradox: “Perhaps that sense can be brought out by the fact that ‘green’, but not ‘grue’, is learned (learnable?) ostensibly by a sufficient number of samples, without reference to time.”¹³⁴ This response is basically in agreement with a

¹³² See Kripke (1982), p. 58.

¹³³ This term was introduced by Blackburn (1984), *Spreading The Word*.

¹³⁴ Kripke (1982), n. 46., p. 59.

Dretskean approach to the error problem. Namely, it is the concept *green* that is possessed by a subject because it is the concept *green* that the subject picked up in the process of learning. But Kripke objects to the response as follows: "Who is to say that it is not 'grue' that others (or even, myself in the past?) learned by such ostensive training?"¹³⁵ The objection Kripke expresses clearly refers to the underdetermination problem. In the face of an infinity of concepts or rules that are mutually inconsistent but are equally compatible with the finite set of empirical data, which one does a pupil pick up?

The answer appears to lie in one of three possibilities. The first possibility is that a pupil does not pick up a concept at all. In the face of an infinity of logically possible concepts that are all compatible with the input stimuli, the cognitive mechanisms of a pupil may simply "freeze", or become dysfunctional. Thus, there may exist no concept learning at all. The second possibility is that a pupil does pick up a concept, but he picks one up *randomly*. The third possibility is that a pupil picks up a concept, and does not do so randomly. Instead, he picks up one from the possible infinity of concepts by virtue of having certain endogenous constraints imposed on his cognitive mechanisms.

In my view, the first possibility is very implausible. The reason is that it renders communication and mutual understanding between people impossible. If a subject acquires no concepts in the learning process, the subject could not entertain any thoughts. In that case, there would be no understanding and communication. KW's skeptic would not be able to present any skeptical paradox to us and we wouldn't be able to understand the skeptical paradox. Besides, in considering the underdetermination problem, KW does

¹³⁵ Kripke (1982), n. 46, p. 59.

not doubt the capacity of a subject to acquire a concept during the learning stage. What is doubted is that there is a determinate answer regarding *which* concept is learned (e.g., *green* or *grue*, *plus* or *quus*, etc.). Fundamental to KW's skeptical argument regarding the meaning of a word is that since there is no determinate answer regarding which concept is learned by a subject, there is no determinate answer regarding which concept is possessed by the subject. As a result, there is no fact that constitutes meaning something by a term or rule-following. It is apparent at this stage of the dialectic that KW has to grant that a concept can be learned; the problem is that we don't know for sure exactly which one.

What about the second and third possibilities? That is, could it be the case that a pupil picks up one concept from the infinite possibilities *at random*, or *with certain constraints imposed on his cognitive mechanisms*? If the former were the case, then the consequence would be no better than the first case. If everyone picks up a concept from the pool of a logical infinity of concepts at random, communication and mutual understanding would also be very implausible because it would be very likely that everyone uses the same term to convey different concepts. Even though it might be the case that two people by coincidence choose the same concept to attach to the same word, this would still be by coincidence, and no doubt rare. Communication would still be inexplicable.

In the past two decades or so, substantial results have been attained in research on infants that strongly suggests that the second possibility is false and the third possibility is very likely to be true. Numerous experimental results show that neonates are endowed with innately specified structures that enable them to respond to certain constrained

classes of perceptual, conceptual, numerical, and linguistic input (e.g., Anderson, 1988; Asch, 1952; Atran, 1990, 1995; Baillargeon, 1987, 1995; Berlin, 1992; Berlin, Breedlove, & Raven, 1973; Gelman, 1988; Carey & Spelke, 1994; Cohen, 1988; D'Andrade, 1987; Leslie, 1982, 1994; Mandler, 1988; Spelke, 1988, 1990; Wellman, 1990). For example, the experiments of Johnson et al (1989) showed that newborn infants (within the first hour of life) tracked certain types of face-like patterns further than other patterns. Spelke (1988) has also shown that very young infants are surprised when viewing a physically impossible event such as one solid object passing through another. It was also observed that all omnivorous mammals are likely to infer a causal link between distinctive-tasting food and gastrointestinal illness experienced many hours later (Garcia, McDowan, Ervin, & Koelling, 1968). Garcia et al. (1968) also established that rats were not able to notice correlations between intestinal distress and bright lights. This result shows that mammals are disposed to form some but not other concepts. Experimental results on language show that infants are already sensitive to word boundaries (Gleitman et al., 1988), to the distinction between linguistically relevant and other non-linguistically relevant sounds (Mehler & Bertoncini, 1988), to certain characteristics of their mother tongue (Mehler et al., 1988), and so on. These behavioral phenomena appear so early and are apparently so prevalent that it is very likely that infants are biologically, innately constrained in at least some aspects of their minds.

Especially interesting and relevant to our current discussion of the addition example are some empirical studies that show that the capacity to perform arithmetic calculations seems to be biologically given cognitive equipment in humans. In what follow, I will present some of the studies in some detail and discuss their implications.

In recent years, exciting and promising research has been done on infants and animals regarding the possibility of their possessing innate mathematical and physical knowledge (e.g., Butterworth, 1999; Dehaene 1997; Spelke et al, 1992; Wynn, 1990, 92). One prominent thesis claims, with strong empirical support, that humans as well as animals possess a rudimentary capacity to represent numbers. Dehaene (1997), one of the key figures in advocating this thesis, argues for the hypothesis that the rudimentary representational capacities of number is analogue in nature, present not only in humans, but also in other animals, ranging from rats, pigeons, parrots, raccoons, to monkeys and chimpanzees. This analogical power is a capacity to perform very fast approximate, rather than exact, numerical calculations on physical inputs carrying numerical information. For example, arithmetic relations such as that 9 is larger than 4, $7 + 15$ cannot possibly equal 146, and so on, are immediately known without much reflection or careful calculation. Dehaene argues that an analogical capacity is mainly responsible for such quick and approximate performances. Dehaene also believes that this analogical numerical capacity has a long evolutionary history, and has a specific cerebral basis.

The distance effect and the size effect are two major empirical data in support of the existence of an analogical numerical capacity in infants and animals. The distance effect is a systematic decrease in “numerosity” discrimination performance as the numerical distance between numbers decrease. For example, the fact that 25 is smaller than 50 is easier to detect than the fact that 45 is smaller than 50, because the distance between the two numbers in the former (25) is larger than the distance between the later (5). The size effect refers to the fact that even for numbers with equal numerical distance, performance also decreases with increasing number size. For example, it is easier to know that 3 is

larger than 2 than to know that 37 is larger than 36. Although the distance between the two pairs of numbers is the same, namely, 1, the latter pair (36 and 37) is larger in magnitude (more accurately, $37 > 3$ and $36 > 2$) and the difference is hence somehow more difficult to process. Both effects indicate that infants and animals possess only a fuzzy representation of numbers, in which imprecision grows proportionally to the number being represented.

This fuzzy numerical capacity is apparently different from the arithmetic ability of adult humans. Adult humans use linguistic and Arabic symbols, which make exact calculations possible. What, then, is the relation between the rudimentary, analogical numerical capacity shared by both infants and animals and the higher-level symbolic ability possessed only by trained humans? Dehaene argues that the connection in question is continuous, rather than discontinuous. That is to say, the analogue capacity of humans does not either disappear or get replaced in the process of maturing and learning arithmetic in symbolic form. Rather, the analogical capacity continues to exist, and constrains or coordinates with numerical information mentally encoded in verbal (e.g., twenty-four) and Arabic forms (e.g., 24). There are certain transformation procedures that enable information to be translated from one code to another. Experimental results in Ashcraft & Battaglia (1978) and in Dehaene et al., (1999), show the significance of the role played by the analogical capacity in some problem-solving situations. When presented with an arithmetic sentence like $34 + 47 = 268$, subjects readily rejected it as false, for 268 is instantly “sensed” to be too distant from the operands of the addition. This performance was made possible by the exercising of the analogical capacity, which compared magnitudes of numerosities that are transformed from linguistic/Arabic

information-input. Dehaene describes the interplay between the analogical, linguistic, and Arabic systems in an adult human mind as the “triple-code model”.

The existence and active functioning of the analogical capacity in an adult human linguistic/Arabic cognitive system of arithmetic thus shows that a subject does not learn a concept/rule at random. Rather, a subject learns a concept/rule through certain constraints imposed on his cognitive mechanisms. Given Dehaene’s triple-code model, a subject’s disposition to follow arithmetic rules is continuous with and constrained by our innately endowed numerical analogical capacities both during and before the learning stage. Human cognitive mechanisms possess such endogenous constraints which influence and shape human learning in a way that might be far more extensive than we ordinarily recognize. As Dehaene’s work shows, humans and a wide variety of animals are homogeneous in terms of sharing some very fundamental cognitive mechanisms that are innately present due to a long evolutionary history.

A related line of research also shows that newborn infants possess knowledge of very basic arithmetic relations. In Wynn’s experiments (1992), the violation-of-expectation paradigm was employed. In one group, 5-month-old infants were presented with scenes where two physical objects were successively placed behind a screen. These infant subjects evidently expected two objects when the screen dropped, and were surprised if only one object was observed. In another group, 5-month-old infants were presented with scenes where two objects were initially placed in a case and a screen came up. One of the two objects behind the screen was then taken away by an experimenter. Those infant subjects were able to observe this physical transformation clearly. Again, when the screen dropped, 5-month-olds expected one object, and were surprised if

another number of objects such as two were observed. These results showed that infants, as young as 5 months old, were able to perform simple computations with very small numerosities, such as $1 + 1 = 2$ and $2 - 1 = 1$. The same result was replicated by Koechlin et al., (1997) and Simon et al., (1995). They controlled and investigated variables such as location and identity of objects in their experiments, and demonstrated that infants' discriminatory behaviors toward numerosities were not explainable by either the location or identity of objects. Infants thus seemed to have abstract numerical representations of objects on the scene.

The fact that human cognitive mechanisms exhibit innate constraints has important implications for solving the underdetermination problem of learning in a dispositional account of rule-following. It implies that bent or wild rules, such as the quus rule according to which $68 + 57 = 5$, are very unlikely to be learned and followed as a matter of empirical fact. For, given the functioning of our rudimentary analogical mechanisms, we would spontaneously and instantly reject that arithmetic proposition as false in the process of learning addition. These spontaneous and very likely unconscious inclinations indicate that we possess innately endowed dispositions to acquire further generally more refined dispositions to follow only those rules that fall under certain range, not just any arbitrarily constructed rules of any unconstrained sort. KW's underdetermination problem with respect to learning, which is based purely on the logical/formal constructions of rule/concept predicates, is hence seriously undermined by the considerations of human/animal learning mechanisms which are revealed as being innately constrained in certain ways.

Note that we can, and do, take the facts about human endogenous constraints into account in forming hypotheses about dispositions. When we ascribe rule-following dispositions to people to explain and predict their behaviors in the way described by the DN model, we do strongly tend to reject the hypotheses that people are following wild rules, such as the quus rule. Our awareness of the facts about human innate constraints explains our rejection of KW. For based on the first-person experiences in our own learning histories, we know that it's extremely unlikely that we pick up the quus rule. We know that we are simply not built that way. Hence, there is a sense in which we don't seem to need Dehaene's and Wynn's experiments to know that we don't learn the quus rule. We, as laypeople, share the same insight with scientists about human's innate cognitive constraints.

This way of treating the underdetermination problem is similar to Chomsky's response to KW.¹³⁶ A Chomskyan approach posits certain innate linguistic rules that constitute knowledge of language and attributes them to a subject. This approach, however, is vulnerable to the objection that it simply begs the most important question against KW. For KW is claiming that there is no fact of the matter that can determine which rule a person follows. Instead of responding to KW by showing what such a fact is and how it determines a person's rule-following, a Chomskyan simply posits a rule and attributes it to a person. This move appears to be highly arbitrary, and is hence very unsatisfactory.

My approach, however, differs from the Chomskyan approach in an important way. That is, in my account, what are innate are not rules. Rather, what are innate are

¹³⁶ Chomsky, *Knowledge of Language*, 1986, esp. Chapter IV.

cognitive/biological mechanisms which create dispositions to follow/form certain rules. The fact about a person that determines (at least partly; as I will show shortly, another factor that influences the determination of rule-following is an individual's learning history) his rule-following is the person's dispositions which are innately constrained in a certain way. This answer does not beg the question in the way Chomsky does.

Nonetheless, there remains a gap between how innately endowed characteristics of human cognitive mechanisms influence human concept/rule learning and what concepts/rules humans actually learn at a later stage of their lives. The endogenous constraints that the human species share with other species do not dictate what unique rule/concept a person eventually acquires. A concept like *plus* is complex, so that its acquisition requires more than merely exercising the power of the innately endowed cognitive mechanisms. Thus, undergoing a certain period of training is also crucial in determining what concept a human being eventually acquires. Hence, wouldn't the skeptic argue that which concept a person acquires remains unfixed, and hence the original underdetermination problem of learning remains unresolved? The difficulty involved is clearly derived from the fact that the human species has a far greater degree of flexibility and creativity¹³⁷ in learning than other species such as rats, pigeons, or chimpanzees. The behaviors and also learning of non-human species are comparatively much more rigidly constrained by their innate cognitive endowments, whereas human constraints are subject to modifications and expansion through learning and reflection in the developmental process.

¹³⁷ See A. Karmiloff-Smith. "Beyond Modularity: Innate constraints and developmental change", in Carey & Gelman (1991).

Humans' acquisition of a concept/rule depends on their specific learning histories, in addition to their innately endowed mechanisms of learning. The influence of the learning history on a person's concept/rule acquisition is obvious. The way in which instruction and feedback are given to a pupil in a pedagogical setting is apparently crucial in the determination of the rule/concept that is being learned. For example, imagine a bizarre pattern of instruction is given to a pupil, so that " $2 + 2 = 4$ " is given at one occasion, " $2 + 2 = 3$ " at another occasion, and " $2 + 2 = 13$ " at still another occasion, and so on and so forth. Not only would it be very unlikely that the pupil picks up the standard basic routine rule $2 + 2 = 4$, but it would also be quite difficult to determine which rule the pupil will end up learning, if any. On the other hand, if " $2 + 2 = 4$ " is reliably and stably presented to a pupil, it would be nomologically plausible that the pupil learns this standard basic routine rule, due to the way the pupil's cognitive mechanisms (e.g., his working memory system) are constrained. Therefore, it's reasonable to hypothesize that innate constraints on human cognitive mechanisms and a person's specific learning history together determine the concept that the person acquires.

I do not intend to provide a detailed story about which specific rule/concept a subject acquires in his learning history. Various empirical approaches to the study of the mind and concept/rule learning have provided evidence to support theories of the mental processes and inner constraints that are causally operative in a human being's learning process. For example, VanLehn (1990)¹³⁸ actually proposes that analyses of "bugs", i.e., systematic errors, that occur during a pupil's history of learning arithmetic may reveal how human cognitive mechanisms work during learning. Vanlehn's data show that

¹³⁸ VanLehn, *Mind Bugs*, 1990.

children's errors in math learning are, more often than not, systematic and result from their carrying over incomplete procedural knowledge from previous learning. More specifically, possessing incomplete procedural knowledge leads to "impasses" in the process of solving a new addition or subtraction problem. Children nonetheless will always try to "repair" their impasses. Errors are thus the consequence of unsuccessful attempts to repair. Vanlehn then formulates a likely computational model that successfully demonstrates the "repair" mechanism which may explain and predict the occurrences of systematic errors in math learning.

Different versions of inner constraints have also been proposed by computational theorists of the mind such as Fodor (1975; 1981)¹³⁹, Chomsky (1967)¹⁴⁰, Kukla (1992)¹⁴¹, and so on, to delineate what the constraints might be on rule-generation and rule-evaluation that occur in a learning process. Chomsky's theory proposes that although a potentially infinite number of hypotheses could be generated regarding which concept is being taught in a learning situation, a human learner is capable of generating only a few concept-hypotheses given various cognitive constraints such as the limitation of short term memory. Fodor's theory posits certain "simplicity metrics" which human cognitive mechanisms employ to evaluate the relative simplicity of those generated concept-hypotheses, based on their syntactic and semantic properties.

Kukla's theory of inner constraints is even simpler than Chomsky's or Fodor's. According to Kukla's theory, a concept-hypothesis H gets selected if H is simpler, and H

¹³⁹ Fodor, *The Language of Thought*, 1975, *Representations: philosophical essays on the foundation of cognitive science*, 1981.

¹⁴⁰ Chomsky, Recent contributions to the theory of innate ideas: summary of oral presentation, in R.S. Cohen & M.W. Wartofsky (Eds), *Boston Studies in the Philosophy of Science*, Vol. 3 (Dordrecht Reidel).

¹⁴¹ Kukla, "Endogenous constraints on inductive reasoning", *Philosophical Psychology*, 1992.

is simpler than other concept-hypotheses if we think of H before we think of others.

Therefore, according to Kukla's theory, the basic routine rule R: $1 + 1 = 2$ would be simpler than other basic routine rules such as R*: $1 + 1 = 3$ at t1, $1 + 1 = 2$ at t2, $1 + 1 = 6$ at t3, and so on, if R is thought of earlier than R*. In such case, a learner will choose and hence learn the rule R, and look no further in a learning process.

Kukla's theory of inner constraints strikes me as likely because it fits well with the fact described in Dehaene's work that humans are very much like other animals in that they are cognitively homogeneous. It is nomologically plausible that the rule R* simply would not occur to the minds of humans any more than it would to the minds of rats when they face certain physical stimuli carrying numerical information. It is most likely that the symbolic representation of the rule R, rather than that of the bent rule R*, occurs first in the minds of human as well as non-human species. The reason is that the inner mental/brain mechanisms of human and non-human species are so innately constructed for evolutionary reasons that they are spontaneously and effortlessly disposed to pick up the rule R. R* or other non-standard rules may never get a chance to appear in their minds and get processed at all.

An abundance of empirical studies on bias in human reasoning (e.g., Evans, 1989) have shown that human thinking exhibits a certain systematic pattern, given certain human cognitive constraints. For example, in the by now classic experiments conducted by Wason (1968a, 1968b, 1977), Wason and Johnson-Laird (1970), Johnson-Laird and Wason (1970), subjects were given a rule T: "If there is a vowel on one side of the card, there is an even number on the other side of the card." They were presented with four cards facing up, corresponding to the following four cases: a) a vowel such as E; b) a

non-vowel such as K; c) an even number such as 4; and d) a non-even number such 7.

The rule T is a conditional of the form: "If P, then Q." Case (a) represents an instance of the conditional when given P, case (b) an instance when given $\sim P$, case (c) an instance when given Q, and case (d) an instance when given $\sim Q$. Subjects' task was to select a minimum number of those cards that are necessary to evaluate whether or not the rule T is true. The correct choice is to select cards (a) and (d), because according to logic, "If P, then Q" also implies T': "If $\sim Q$, then $\sim P$ ".

The results of experiments consistently showed that subjects, including very intelligent college students, performed poorly in the selection task. For example, in a group of 128 college students, only 5 of them correctly chose P & $\sim Q$ cards. Furthermore, subjects' errors exhibited some systematic patterns. Specifically, the two most common errors of the subjects' responses were that they chose both P and Q cards, or only P card. Moreover, it took enormous efforts for subjects to realize that "If P, then Q" logically implies "If not Q, then not P," and thus that (d) is a correct choice. These subjects are strongly inclined to reason one way rather than another way and that coming to appreciate certain alternatives as logically equivalent is very difficulty for humans.

A well-known cognitive mechanism that constrains human reasoning is the limitations of human short-term memory first illustrated by Miller's (1956) famous magic number 7 ± 2 . For example, in Johnson-Laird's (1983) mental model theory, it is proposed that people draw deductive inferences by trying to construct only one or a few, rather than comprehensive, sets of mental models of possible situations implied by the premises of an argument, due to the limited capacity of the short-term memory functioning as a *working memory*. As a consequence, subjects will inevitably err in

syllogistic reasoning when problems require more models to be constructed (Johnson-Laird and Bara, 1984; Johnson-Laird and Byrne, 1992). These empirical data thus show that human reasoning does not take place in any arbitrary or random way. Rather, it is seriously constrained and systematic. Non-standard rules like R^* mentioned earlier or *quus* in KW's discussion are very unlikely to occur in human reasoning, in this light.

For wild, non-standard rules like R^* or *quus* to actually occur in rule-following, R^* or *quus* must be explicitly considered and appreciated as a logical possibility. Psychological literature in bias in human reasoning suggests that rules like R^* may occur to a person only after standard rules like R are learned in a learning process. for the task of appreciating R^* or other bent rules/concept like *quus* or *grue* as a logically possible rule consistent with the finite data given in a learning environment typically requires much extra cognitive effort. Unless a person is explicitly reminded of such a possibility, the person would not even think of it. Even after much effort has been exerted on the task of entertaining a bent concept/rule under explicit instruction, it often remains quite difficult for a person to grasp the bent rule/concept firmly. In other words, it takes symbolic representation even to get an inkling of bent rules. In contrast, the processing of acquired standard rules/concepts is relatively effortless, unconscious, and spontaneous. Consequently, those bent rules could never be the rules that a natural organism would think of first. As a result, it can be determined which rule/concept a person learns in his learning history. Moreover, those theories offered by Fodor, Chomsky, Kukla, and so on, are put forth in a scientific spirit and can be tested. Therefore, the determination of concept/rule learning is tractable.

My claim is, therefore, that KW's skeptical doubt about the determinacy and reality of rule/concept learning and possession should not be regarded as a problem. The skeptical doubt is based on a purely logical consideration, namely, the sheer logical consistency of the bent rule with available data. It loses its hold in a broader context in which a human's ontogeneous learning mechanisms are treated phylogenetically, in terms of how animals are endowed with certain biological mechanism which enable them to react instinctively to environmental stimuli. The human species is endowed as well with innate constraints on cognitive mechanisms that significantly shape and influence their later acquisition of more sophisticated linguistic rules/concepts. Given a knowledge of inner constraints on human cognitive mechanisms and the knowledge of an individual's learning history, the concept/rule that he learns can, at least in principle, be unambiguously determined. And we have seen that hypotheses about them can be tested and confirmed. As a result, the reality of human dispositions to possess concepts and follow rules can be more clearly depicted and characterized by biologically given constraints and learning histories of particular individuals. I believe that this treatment from the learning perspective offers a plausible solution to KW's quus problem. Moreover, Goodman's grue problem may also receive a satisfactory treatment from the perspective of innate dispositions and learning histories of individuals. But a more extensive and detailed treatment of Goodman's grue problem has to wait for another occasion, due to the limitation of space and scope of my dissertation.

V. Solving the Finitude Problem

I have shown that KW's error problem for a dispositional account of rule-following derives partly from conflating a real disposition with its manifestations and

partly from failing to distinguish between distinctive learning stages. Attending to the former misconception helps reveal that a real disposition is an appropriate fact to constitute rule-following. Moreover, what a rule-following disposition is can be determined on a sound methodological and epistemological basis. Attending to KW's mistake of conflating distinctive learning stages helps us realize that a real disposition to follow a rule is acquired through learning and can be more clearly illustrated jointly by a subject's innately endowed constraints and by his particular learning history.

However, there remains the finitude problem. The finitude problem, as KW conceives it, is a problem arising from the fact that a disposition is finite. According to the dispositional account, a person's following a rule is constituted by his possessing a disposition to perform, say, calculations, in a certain way. KW points out, however, that a person's disposition can produce only finitely many calculations. However, a rule or meaning dictates infinitely many calculations. How then can possessing a disposition and following a rule be equated?

Martin & Heil, as we have seen in Chapter VI, argue that the finitude problem is an unfortunate result of KW's misconception of the metaphysical nature of a disposition. A disposition, according to Martin & Heil, is a real property and infinity is built into it at the most basic level. For example, a quark of type A can come into causal contact with a potentially infinite number of type-B quarks for possible manifest behavior of a certain sort. A type-A quark can also interact with a type-B quark potentially at any time. Thus, a type-A quark has a disposition of an infinite nature in the sense that the disposition can project itself to anywhere and at any time for a possible manifestation. Similarly, a subject's having a disposition P to follow a particular rule (e.g., to answer "4", given the

query “ $2 + 2 = ?$ ”, or even, to use “+” to mean plus) is infinite in the sense that it can project itself to any time in the future in any problem case. A subject’s rule-following disposition may be dissimilar to a quark’s disposition to interact with a quark of another type in that the subject’s disposition can be repeatedly manifested while the latter disposition can perhaps be manifested only once. A quark’s disposition to interact with other quarks may be compared to the disposition of a bomb to explode: once the bomb is exploded, it loses its explosive power. On the other hand, a subject’s rule-following disposition is analogous to a bridge’s having a supporting dispositional power which is not lost after a car passes on it. Nevertheless, there is a clear sense of infinity in both cases: a subject’s or object’s disposition is projectible to an potential infinity of times for a manifestation under a potential infinity of input stimuli.

Although Martin & Heil’s view of dispositions is insightful and convincing, a skeptical worry remains about their treatment of the finitude problem. The skeptical worry can be explicated more clearly in Shogenji’s dispositional framework (1993).¹⁴² Suppose we accept that an agent’s disposition to follow the basic-routine rules can be non-circularly specified (e.g., by appealing to certain justifiable criteria of classifying rule-following behaviors into correct and incorrect ones), and that a disposition is infinite. A skeptic would still ask, will those basic routines, already unproblematically fixed, be applied *recursively* or *reqursively*? A recursive application is defined as repeatedly or uniformly applying basic routines in cases which a subject has not previously encountered, while a reqursive application can be defined as applying basic routines non-uniformly in cases which a subject has not previously encountered. Both ways of

¹⁴² Please see Chapter V of this dissertation on Shogenji’s ideas.

applying basic routines are compatible with a disposition's being infinite. Thus, it remains to be determined which disposition to apply basic routines a subject possesses. Because of the limited capacity of a subject, the subject's disposition to follow a (basic-routine) rule can only have a limited number of manifestations. Then, how can it be determined on a sound epistemic ground whether it is recursion or reqursion that is built into an agent's rule-following disposition? The finitude problem thus remains unresolved, due to the epistemological difficulty it involves.

My diagnosis of the finitude problem that remains unresolved within Martin & Heil's account is that it is derived from a sort of Humean inductive skepticism. A Humean skeptic about induction would ask us why, given the observation that the sun rose in the past, we have any reason or justification to believe that the sun will continue to rise tomorrow. In other words, why are we justified in projecting an observed regularity in the past onto the future and believing that the future will resemble the past? Such a skeptical doubt about induction underlies the skeptic's doubt regarding whether it is recursion or reqursion that is built into an agent's disposition. If an agent has a disposition to recursively apply basic routines, then the agent's future behaviors will resemble his past behaviors. In such a case, the agent would be following the plus rule. On the other hand, if an agent has a disposition to reqursively apply basic routines, then the agent's future behaviors will not resemble his past behaviors, and he would hence be following a non-standard rule. It is clear that underlying the finitude problem is a Humean problem about induction.

My overall response to the problem is to adopt a reliabilist position about the inductive skepticism involved. I do not plan to, and cannot, embark on the whole

discussion of reliabilism. The main idea of reliabilism is that a subject's belief that *p* is justified if *p* is produced by some reliable process of belief-formation. The subject does not have to be aware of the factors that are relevant to how he comes to possess the belief that *p* in order for his belief that *p* to be qualified as knowledge. In this regard, reliabilism is often called an "externalist" approach to knowledge.¹⁴³ I believe that such an externalist view of knowledge (and justification) will offer us a proper response to KW's skeptical view that every use of a term or every act to follow a rule is merely "a leap in the dark." Our intuitive idea, contrary to KW's, is that we just know what we do when we follow a rule or mean something by a term. My claim is that reliabilism can do justice to our intuitive idea in question.

A reliabilist solution to the finitude problem is, roughly, this. When I watch you add, I form, on the basis of a finite sample of your behavior, a hypothesis about what infinitary disposition you possess inside you. However, KW's skeptic would ask: How do I, and how can I, know this. My reliabilist response is: simple. We can, and do know what disposition a person has, and hence what rule the person is using, even if we ourselves cannot articulate a logically decisive reason for generalizing about the person's disposition one way rather than any number of other ways logically consistent with the data. For so long as our tendency to generalize about the person's disposition happens to be right, we are epistemically warranted in believing that the generalization is right. If such a reliabilist approach to the Humean skepticism about induction has a chance of

¹⁴³ For major proponents of reliabilism, see A. Goldman, 1979: "What is justified belief," in G. Pappas, ed.; 1986: *Epistemology and Cognition*; 1988: "Strong and weak justification," in J. Tomberlin, ed.; F. Dretske, 1981: *Knowledge and the Flow of Information*; R. Nozick, 1982: *Philosophical Explanations*; D. Armstrong, 1973: *Belief, Truth, and Knowledge*. For a recent attempt to offer a reliabilist solution to the problem of induction, see M. Levin, 1993, "Reliabilism and induction." *Synthese*.

succeeding, we would be justified in believing that a subject possesses a disposition, acquired through the interplay of his innate cognitive constraints and his past learning history, to apply basic routines recursively.

Let me elaborate the reliabilist solution a bit further. A standard form of reliabilism, forcefully advocated by A. Goldman (1979, 1986, 1988) and others, states that a belief is justified if and only if it is produced by a reliable cognitive process. A reliable process is one that is capable of producing a sufficiently high ratio of true beliefs. According to reliabilism, a subject need not know that a belief-forming process is reliable for the belief produced by that process to be justified. All that is required for a belief to be justified is that the belief-forming process is reliable. So, for example, you believe the sun will rise tomorrow, because of your tendency to project past regularities and the observation that the sun has risen in the past. You will be justified in believing that the sun will rise tomorrow, reliabilism says, if the sun will rise and if your belief-producing mechanism of inductive inference is reliable (i.e., if your inductive extrapolations will continue to be mostly correct). Similarly, since a competent adder will most likely apply basic routines repeatedly in the next new case and the belief-forming mechanism of inductive inference is reliable (if it has so far been very successful in producing true beliefs about a subject's calculations and will continue to be so in the future), we are justified in accepting the inductive belief that people will repeatedly apply those basic routines they acquired through learning and therefore follow the recursive rule.

It's worth repeating here that the strategy I take in justifying our future rule-following behaviors is essentially third-person. KW's skeptic may jump in and object to my approach again that our having justification of our rule-following behaviors does not

consist in a third-person factor. Rather, rule-following, KW's skeptic would continue, is essentially first-person in the sense that we simply know what to do in a new case without having to rely on inductive justification for the new case. It is my main task in the next chapter to address the first-person problem from a reliabilist point of view.

Let me repeat that I do not intend to discuss reliabilism as a form of theory of knowledge and the theoretical difficulties it might face in any further detail. My aim is to point out that underlying KW's finitude problem is a Humean, epistemological problem about extrapolation, and that a reliabilist theory of knowledge offers a way to provide the epistemological basis for the dispositional account of rule-following. The skeptical doubt about the finitude problem on a realist dispositional account of rule-following would consequently disappear.

Chapter VIII. Proposing a Plausible Dispositional Account of Rule-Following (II): Solving the First-Person Epistemic Problem

I have shown in the previous chapter my solution to the error and finitude problems. However, there remains one intriguing problem for a realist dispositional account of rule-following, namely, the first-person epistemic problem. When we mean something by a term or follow a rule, we seem to be *guided* by the meaning or the rule that we firmly grasp¹⁴⁴. When (or right before) we perform a calculation based on the rule that we acquire and grasp from our learning history, we are usually fairly certain about what rule we are following, without relying on any inference based on the observation of our behaviors or investigation of our neurophysiology. Our performances may happen to be incorrect sometimes due to various factors. However, in many cases, our performances are correct and their correctness is guaranteed by the rule or meaning to which we commit ourselves. In other words, when pressed with the question of why our performances are correct, we are often able to cite the rule or meaning to which we seem to have direct access to justify those performances.

In contrast, a disposition that constitutes meaning and rule-following is a theoretical posit in a subject and could be known only through the inferences and hypotheses based on subject's behaviors and other empirical evidence. In fact, I have argued this in the previous chapter. A disposition is hence something opaque to us. Given that meaning/rule-following is crucially first-person and subjective, whereas a disposition is fundamentally third-person and objective, how then can these two items be the same thing? A difficulty for the realist dispositional account is, therefore, to show how a

¹⁴⁴ Kripke (1982) is explicit and emphatic on this observation of rule-following and meaning. See, for example, p. 17.

realistic disposition may accommodate the first-person epistemic dimension of rule-following and meaning, and the opaque nature of a disposition. The hint of a solution to the first-person epistemic problem lies in showing how an underlying disposition could be consciously engaged in a manner similar to the direct and non-inferential way by which a person is consciously acquainted with the meaning of a term or a rule.

The first-person epistemic problem for a dispositional account of meaning and rule-following can be formulated schematically by the following argument (E):

(E)

1. The rule that an agent follows (or the meaning of a term that an agent understands) is directly accessible to the agent who grasps the rule.
 2. An agent's underlying disposition to follow a rule (or to apply a term) is not directly accessible to the agent who possesses the disposition.
-
3. Therefore, following a rule (or understanding the meaning of a term) is not (constituted by) a disposition.

The argument is clearly valid. Given that the two premises are true, the conclusion that rule-following or meaning cannot be constituted by a disposition has to be accepted. Consequently, the dispositional account of rule-following and meaning must be false.

In this chapter, I will try to save a realist dispositional account from the first-person epistemic problem by refuting the argument (E). In particular, I will try to show that the second premise of (E) is false, and hence the argument does not go through. As a result, KW's conclusion that meaning or rule-following cannot be constituted by a disposition is not sustained. More importantly, by pointing out how and why premise 2 is false, I hope

to offer a more adequate and clear depiction of the phenomenological dimension of rule-following and meaning within a realist dispositional framework.

My position will be along the same line as Martin & Heil's realist dispositional account and Yalowitz's expressivist dispositional account that we discussed earlier in Chapter V. Both of these two accounts offer useful conceptual resources to address the first-person dimension of the rule-following problem. For example, Martin & Heil's employment of the notion of "cue-manifestation" brings our attention to the role played by the "feeling of knowing" experiences in rule-following. Yalowitz makes the idea clear that the possessor of a disposition comes to know about the disposition by directly expressing it. However, as I will point out later, both accounts ignores the importance of the epistemic role played by judgments and beliefs of a propositional form such as "I mean plus by '+'" or "I am following the plus rule". My claim will be that my version of dispositional account does more justice than both of those two accounts in explicating the epistemology and phenomenology of rule-following.

In what follows, I will first offer a functionalist characterization of a disposition, then provide a plausible functionalist account of introspective consciousness, and apply it to explicate how a possessor of a disposition to rule-following may get access to the disposition in a direct and non-inferential manner. Finally, I will also show how the introspective account of a functionally characterized disposition can be used to explain the Wittgensteinian puzzle that a verbalized rule apparent in our introspection is seemingly self-interpreting but does not fix the rule being followed. All this may seem a bit too ambitious. But it's a project worth undertaking. I hope that the result is fruitful.

A Functionalist Account of Dispositions

The crucial question for the first-person epistemic problem is: How do we come to know the disposition to rule-following that we possess. In the context of the realist account of dispositions I presented in the previous chapters, a disposition is real in a minimal sense, namely, a disposition ascription cannot be identified in terms of or reduced to a counterfactual statement. A disposition ascription, such as “the piece of glass is fragile” or “John has a disposition to add”, attributes a real property. The ascribed disposition is responsible for its possible manifested behaviors under various conditions (e.g., breaking when dropped on a hard floor or answering “4” when queried “ $2 + 2 = ?$ ”). However, a disposition may not manifest itself due to the absence of suitable conditions or the presence of intervening conditions. In such a case, the disposition remains intact, though unobservable. This is a strong realist intuition about dispositions. We may ask, then, what kind of real entity or property a disposition is. The answer to this question has a direct bearing on responding to the crucial question of how we come to know the dispositions we possess. So far, I have only presented a negative depiction of a disposition, namely, it is not to be “reduced” to a counterfactual conditional. A more positive and informative characterization of a realistic disposition is needed.

A characterization of dispositions that I find plausible is the functionalist account, which is described by McLaughlin as “the leading theory of dispositions today.”¹⁴⁵ Functionalism is well known as a theory of the mind in which states of mind and mental properties are construed as functional states and properties. Human dispositions to rule-

¹⁴⁵ B. McLaughlin (1995), “Disposition”, in J. Kim and E. Sosa (eds.), *A Companion to Metaphysics*, p. 123.

following or meaning can be characterized by a functionalist account of mind as being realized in a human biological and neurological system.

Though there are a variety of functionalist theories¹⁴⁶, the basic and common idea of a state's possessing a functional property is for the state to occupy a certain causal role. For something to occupy a certain causal role is in turn for it to produce a certain output when receiving a certain stimulus input. Let's take an artifact, such as a thermostat, for example. Any system would be a thermostat if it performed the following task: when there is temperature change, the system can produce a representation of the change in a suitable way. So characterized, what a thermostat is is determined by what it does. More specifically, the sort of causal role that the state of an object occupies qualifies the object as a thermostat, as a system that possesses a functional property. Note that the exact physical mechanism that constitutes a thermostat is irrelevant and inessential to whether or not it is a thermostat. Two objects might be constructed of completely different materials and have different mechanisms, but could both be thermostats, insofar as they perform the same task under similar circumstances. What is crucial to a functional property of a state is, therefore, the state's having a certain causal role, which is to be construed at a higher level (the functional level), above the object's physical constituents¹⁴⁷.

The distinction between a higher-level functional state and a lower-level physical state is thus crucial to the functionalist thesis in its original formulation. It is the higher-

¹⁴⁶ A range of papers about functionalism can be found in Block (1980), Lycan (1990), Rosenthal (1991), Beakley and Ludlow (1992), Goldman (1993), etc.

¹⁴⁷ Note that this characterization of a functional property excludes a version of functionalism proposed by D. Armstrong (1968) and D. Lewis (1966) according to which a functional property is identified as its realizer. The exclusion shall not matter much, for my subsequent discussion will not hinge on this.

level functional property, rather than the lower-level physical property, that captures what is shared in common by all thermostats. The same insight is employed not only to apply to artifacts like thermostats, but also to other kinds of items, such as biological organs like hearts. The essential feature of a heart, according to a functionalist, can be best captured by what it does, or by what its functional role is, namely, to pump blood and enable blood circulation. Though a heart is typically constituted by a bio-neurophysiological mechanism of a certain structure, it can also be constructed from different material as far as the material realizes a particular sort of causal role (e.g., artificial hearts).

Similarly, mental properties, such as having pains, are believed by functionalists to be properties of a higher-level kind, which does not have to do with the kinds of material that realize them. For example, pains are what they are not because of the physical or neurophysiological features of the people who possess them, but because of their occupying a particular sort of causal role in people. It's both imaginable and empirically possible that an alien is capable of being in pain while nothing of his physical make-up is remotely similar to ours. Nonetheless, the alien would still be in pain insofar as he exhibits certain patterns with distinctive kinds of input-output causal relations, such as: when the alien is hit by a stone on his head, he screams.¹⁴⁸

However, there is a crucial difference between a mental property such as pain and the property of being a thermostat or a heart that we just mentioned. While the functional role of being a thermostat is characterizable purely in terms of an input-output causal relation, a mental property or state has to be analyzed not merely in terms of its input

¹⁴⁸ Though this is the current consensus in philosophy of mind, John Searle is a prominent dissenter. See, for example, Searle, *The Rediscovery of the Mind*, 1992.

cause and output effect, but also by reference to other mental properties or states. A person's pain, caused by grasping a hot handle, would produce not only a behavioral output such as removing his hand from the handle, but also a desire to wash his burned hand with cold water. Such a desire would not come to exist unless the person also believed that cold water helps ease the pain. A state of pain thus has to be viewed as occupying a place in a causal network involving other inner mental states, in addition to certain input stimuli and behavioral outputs. Those other mental states in the network such as believing that P or desiring that Q are to be in turn characterized by other causal networks of suitable sorts.

The above sketch of functionalism is very rough. But it's similar to Martin & Heil's realist characterization of a disposition. According to functionalism, different tokens of the same mental state type share something essential in common, i.e., the same causal role. By analogy, objects may be made of different sorts of material or exhibit different categorical structures, but they may nonetheless possess the same disposition. More importantly, the same disposition is characterized as having a functional essence, which is in turn defined as an occupant of the same causal network in objects of potentially different material make-up. Take fragility for example. All fragile objects are fragile by virtue of their having a certain input-output causal relation, e.g., when they are dropped on a hard floor, they break. The exact physical structures of these fragile objects are not in the business of explaining why they are fragile. There may be a rich and interesting story at the lower physical level which would help shed light on why all fragile objects behave the way they do. However, the way fragile things behave can be directly

explained by the functional role that they possess without necessarily having to appeal to their physical structures.¹⁴⁹

A disposition to follow a rule or mean something by a term can be similarly construed as a functional state. That is, the disposition's nature can be captured by the causal role it has in a subject that possesses it. The functionally characterized disposition to rule-following is very complex, both because of its having a potentially infinite pairings of input-stimuli and output-responses and its interrelations with other inner dispositional/functional states.

First-Person Access to a Functionally Characterized Disposition to Rule-Following

Given that a rule-following disposition is characterized as a functional state, we may continue to ask how its possessor comes to know about such a functional state. How does a subject know, or have conscious access to, such a complex causal network that that he embodies?

Dennett¹⁵⁰ (1978) provides a plausible functionalist account of introspection that would help illuminate how an agent may be introspectively aware of such a functionally characterized dispositional state. In what follows, I will give a highly compressed depiction of Dennett's account. The main purpose is to see how the introspection of a rule-following disposition may be illuminated within a naturalistic functionalist framework.

¹⁴⁹ There is a problem regarding how or whether or not a disposition, construed as a higher-level functional property, can have causal power, if it is a real property in the world. For discussions and a solution on this, see Mumford (1998), esp. chapter 9. I will leave further investigation of this problem for another occasion, since the problem does not directly hinge on my main concern of rule-following.

¹⁵⁰ D. Dennett, "Toward a Cognitive theory of Consciousness", in *Brainstorms*, 1978.

Dennett's account is given in terms of a computer model language as follows. Suppose Control (a higher executive component realized in the brain) 'decides' for various reasons to 'introspect':

- “(1) it goes into its introspection subroutine, in which
- (2) it directs a question to M (short-term or buffer memory);
- (3) when an answer comes back (and none may), it assesses the answer; it may
 - (a) censor the answer
 - (b) 'interpret' the answer in the light of other information
 - (c) 'draw inferences' from the answer, or
 - (d) relay the answer as retrieved direct to PR (a “public relations” or printout component).
- (4) The outcome of any of (a-d) can be a speech command to PR. The point of a buffer memory M is that getting some item of information in M is a necessary but not a sufficient condition for getting it accessed by PR in the form of the content of some speech act command.”¹⁵¹

There are three important points to note about this model of introspection. First, introspection is the series of processes by which Control, a higher executive component in the brain, directs questions to the buffer memory M, gets the answers after some possible editions and transformations, and sends them to the speech center, which if so commanded, publishes them in a written or verbal form. Second, the process of introspection is best described as asking questions and receiving answers, rather than a direct monitoring or tracking of one's first-order inner processes. Third, Dennett

¹⁵¹ Dennett (1978), p. 156.

proposes that what a person is conscious of depends on what the person can introspect, and what the person can introspect is explained as a routine by means of which he gain access to and reports on the contents of his buffer memory.

In Dennett's account, therefore, whether or not a mental state is introspectively available depends on whether or not the mental state is reportable by its possessor under certain conditions. For example, suppose a person possesses the intention to go to the supermarket. One way in which this intention may be introspectively accessed by the person is for him, when being queried where he is going, to verbally express a thought by saying that "I am going to the supermarket." The content of the thought somehow (via the operations of the underlying cognitive sub-systems) becomes accessible in the agent's buffer memory and, when facing certain external stimuli such as being queried, gets "published" by a speech act performed by the agent.¹⁵²

¹⁵² In Dennett's account, consciousness and introspective consciousness are undistinguished. This is undesirable, for a mental state's being introspectively noticed occurs in a more attentive and focused manner than the same state's being conscious in an ordinary, casual manner. D. Rosenthal's (1986; 1990; 1993; etc.) higher-order thought (HOT) theory of consciousness offers a unified account of both. In Rosenthal's theory, when a mental state (first-order) is being thought about, the mental state becomes conscious. The thought that is about the first-order mental state is a second-order thought (or a higher-order thought, a HOT), that can get expressed in a speech act under certain normal conditions such as when the subject's tongue is not cut and functions normally. However, the second-order thought is itself not a conscious thought. It becomes conscious when it is itself thought about, i.e., when it becomes the target of a third-order thought. Such a third-order thought may be expressed by the subject's reporting that he is thinking about the first-order mental state. When this happens, the first-order state is being introspectively accessed. In my view, Rosenthal's HOT theory does not necessarily conflict with Dennett's cognitive model of introspection and consciousness, but rather the former offers a conceptually more elaborate and coherent account of the same phenomena. In what follows, I will not make a theoretically exact distinction between introspective consciousness and consciousness. But it's better to keep the distinction in mind that Rosenthal's HOT theory makes plausible and clear.

Dennett's account of introspection can be applied to explain how a functionally characterized disposition to rule-following appears in one's introspective consciousness. According to Dennett, the process of introspection is depicted as asking questions and receiving answers. Therefore, when being asked " $2 + 2 = ?$ ", a person having a disposition to add normally utters or writes down the answer "4". On another occasion, the same person would respond with "6" either by uttering it or writing it down when being queried with the question " $3 + 3 = ?$ ", and so on for numerous other occasions. Thus, the objects of introspection when one exercises his dispositional power to rule-following are the contents of those utterances or written sentences such as $2 + 2 = 4$, $3 + 3 = 6$, and so on. These contents are usually of a propositional form that figure in a person's speech-acts performed by the speech center mechanism of the person. These propositional contents are exactly "rooted" in a disposition which get "published" or expressed in utterances or writing under certain conditions. This is one way by which an underlying disposition may be consciously accessed by its possessor.

There are other ways by which a rule-following disposition can be consciously accessed by its possessor. When a person follows an arithmetic rule, he does not merely possess a disposition to produce numerous performances of calculation in accordance with the rule and become conscious of those arithmetic performances by virtue of expressing them in speech acts. The person also exhibits certain intriguing qualitative features that accompany his arithmetic performances. These qualitative features enter into a person's realm of introspective consciousness by virtue of being expressed with contentful thoughts, such as "I got it!", "Now I can go on", "The answer has got to be correct!", and so on, which Wittgenstein (1953) also observes when he reflects on the

role they play in rule-following. These thoughts, which are intimately associated with those qualitative feelings and which may be expressed in speech acts, not only compel the person to perform in a certain way, but also “inform” the person with a force of assurance that what he is disposed to produce will be correct. Even though the person’s actual performance might be incorrect, these meta-cognitive responses provide a subject with information about his underlying dispositional conditions. In Martin & Heil’s terminology, these meta-cognitive features are the “cue-manifestations” of a rule-following disposition, which serve as reliable, though fallible, indicators of the readiness of a subject’s underlying disposition.

Various forms of cue-manifestations are known as “feeling-of-knowing” or “feeling-of learning” experiences, which have become objects of vigorous studies in cognitive psychology (see, e.g., Forrest-Presley, Mackinnon, & Waller 1985; Flavell, Green, & Flavell 1995a, 1995b; Metcalfe & Shimamura 1994; Nelson 1992, 1996; Nelson & Narens 1990, 1994; Weinert & Kluwe 1987). For example, Reder (1987, 1988), Reder & Ritter, (1992), Schreiber & D. Nelson, 1993), etc., define the feeling of knowing as a “rapid, preretrieval stage during which individuals judge the expected retrievability of a queried piece of information.”¹⁵³ Feeling of knowing is observed to be an automated process that precedes the cognitive performance in a problem situation and requires

¹⁵³ See Miner & Reder, “A new look at feeling of knowing: its metacognitive role in regulating question answering,” in Metcalfe & Shimamura (1994). Note that this definition modifies on and differs from the classic definition given by Hart (1965) where feeling of knowing is “...the state of believing that a piece of information can be retrieved from memory even though that information currently cannot be recalled.” (see again Miner & Reder, “A new look...” for this definition.) The modified definition of feeling of knowing is, in my view, closer to the spirit of Martin & Heil’s notion of cue-manifestation and captures the central role better played by meta-cognitive feeling of knowing experiences in our cognitive activities.

minimum mental effort. Nelson & Narens (1990) labels such a process as a “monitoring” process that exists at a higher (or meta) level that is about the lower (object) level cognition. Feeling of knowing not only enables an agent to monitor his own first-order cognitive process, but it also serves a “control” function. Empirical evidence shows that feeling of knowing affects strategy choice (e.g., Reder, 1988) and search duration (e.g., Gruneberg, Monks, & Sykes, 1977; Lachman & Lachman, 1980; Nelson et al, 1984; Reder, 1987, 1988).

The speculations about what mechanisms might underlie feeling of knowing experiences are diverse. A dozen or so mechanisms have been proposed, which were then subsumed by Nelson et al. (1984) under two main categories: *trace access mechanisms* and *inferential mechanisms*. The “trace access hypothesis” suggests that subjects have partial access to the target item in a problem situation when they make feeling of knowing judgments (See Blake, 1973; Koriat & Lieblich, 1974 for empirical support). The “inferential hypothesis”, on the other hand, suggests that subjects base their feeling of knowing judgments on related episodic information in their personal memories, on their presumed expertise on the topic of the question, or on their familiarity of the cue in a problem case (See Nelson, Leonesio, Landwehr, & Narens, 1986; Nelson et al. 1984; Koriat & Lieblich, 1974, 1977; Bradley, 1981, etc.) It remains unclear and undecided which mechanism underlies feeling of knowing experiences. However, empirical research in neuropsychology and cognitive neuroscience, typically the study of impairments in various aspects of “knowing” as a result of brain injury or disease, have shed light on what the neurophysical basis may be that realizes the meta-cognitive function of feeling of knowing experiences (for a useful review, see Shimamura, in

Metcalf & Shimamura, 1994). Therefore, it should be uncontroversial to conclude from these empirical studies that the notion of “cue-manifestation” does capture a genuine cognitive (or better, meta-cognitive) phenomenon that not only is grounded in a neuro-physiological basis but also plays an important role in our ordinary mental activities. As to the reliability of such meta-cognitive mechanisms, it remains unsettled. Psychologists tend to be mildly positive about it. As Leonesio & Nelson (1990) commented, the accuracy of feeling of knowing judgments is well above chance yet far from perfect.

Central to my dispositional account of human rule-following is that such a disposition is an innately constrained biological capacity that is shaped by an agent’s learning history. For example, a disposition to add is acquired through learning by organisms of higher sorts, who are controlled by certain innate biological and phenotypical constraints. The human system itself imposes severe constraints on how such a disposition (or function) will be realized. Humans are born with certain innate capacities such that when they receive external, physical stimuli (such as light) which exceed a certain “threshold”, they “sense something” (such as color). Being able to sense something is an innate capacity of organisms of higher sort.

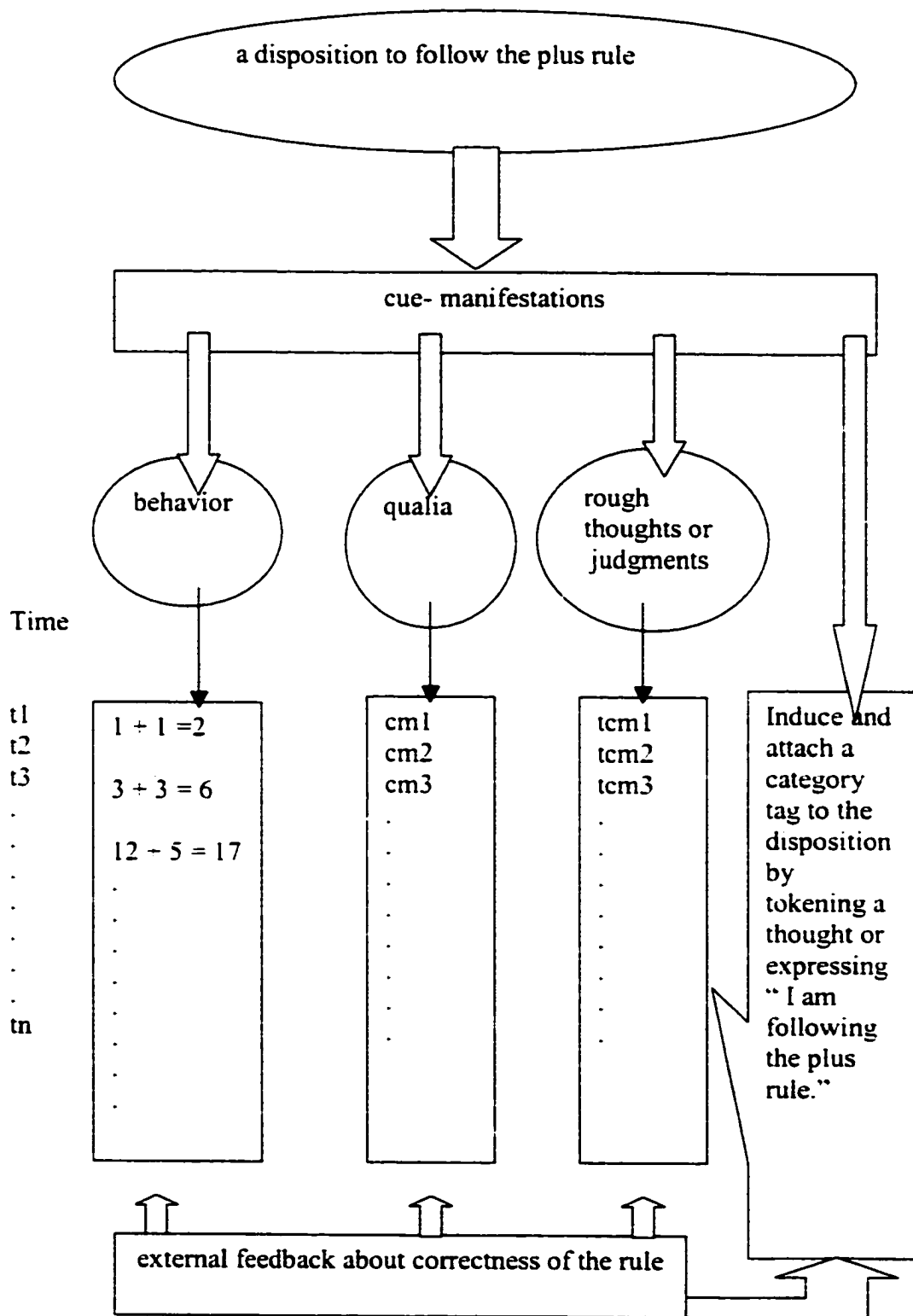
When following a rule, based on the qualitative feelings one experiences, one is able to *assess* his own rule-following capacities for producing relevant behaviors. As Martin & Heil observe, we may not know of the exact fine-grained web-structure of our underlying dispositional capacities to rule-following because we have not yet conducted a thorough empirical investigation of our behaviors and neurophysiology. Nevertheless, we may know (or sense) something about the underlying disposition on the basis of certain qualitative sensations. By turning those sensations into objects of our cognitive

analysis, we can have certain corresponding thoughts about them. Furthermore, by categorizing these inner qualitative states, we may eventually “label” them with some “verbal tag” such as “I mean plus by ‘+’” in a sense consistent with people in a similar dispositional state. Thus, in a direct and non-inferential manner, we, possessors of certain dispositions, know something about the readiness conditions of our underlying dispositions from a subjective viewpoint.

Therefore, we have qualitative feelings (or cue-manifestations in Martin & Heil’s language) accompanying our rule-following activities, and are capable of having thoughts about those qualitative feelings, such as a thought like *Now I can go on!* or a very rough one like *A-ha!* More importantly, we, as competent speakers, can further “attach a verbal tag”, such as “I am following the plus rule” or “I mean plus by ‘+’”, to our rule-following activities such as uttering “ $2 + 2 = 4$ ” and the qualitative and meta-cognitive features that go with it. Articulating such a verbal tag is to express a conscious thought that one “grasps” a certain rule or “understands” a certain meaning. This ability is apparently acquired through a learning process where a subject associates a meaning or rule-tag with his rule-following behaviors. However, one’s actually grasping a rule or understanding the meaning of a word does not consist merely in his issuing a verbal tag. As Wittgenstein has shown, one’s claiming or believing that he follows a rule is not equivalent to his actually following it. Rather, one’s grasping a rule or understanding a meaning lies in one’s possessing a disposition, which is capable of being manifested in behaviors and meta-cognitions, in addition to the ability to give a verbal tag to capture the rule. In short, although attaching a meaning or rule-tag to one’s behaviors is not equivalent to constituting meaning or rule-following, contrary to the claim of Wright’s

judgment-dependence account, this meta-cognitive act of thinking or judging reflects a subject's knowledge about his being in a dispositional state which helps guide and affect his decisions and actual actions in following a rule.

The structure and content of the rule-following phenomenology within a realist dispositional account can be depicted as on the next page.



Note: "cm" stands for cue-manifestation; "tcm" stands for thought cue-manifestation

Figure 1: A schematic illustration of the cue-manifestations of a disposition that are open to introspection

Resolution of the First-Person Epistemic Problem and Explanation of Some Puzzling Phenomena of Rule-Following

We now return to the argument (E) in the beginning of this chapter. Let me restate it here again.

1. The rule that an agent follows (or the meaning of a term that an agent understands) is directly accessible to the agent who grasps the rule.
 2. An agent's underlying disposition to follow a rule (or to apply a term) is not directly accessible to the agent who possesses the disposition.
-
3. Therefore, following a rule (or understanding the meaning of a term) is not (constituted by) a disposition.

My contention with regard to (E) is that the second premise is false. A rule-following disposition is, in my view, directly accessible to the person who possesses the disposition. I have just explained a way in which this access might occur. I grant that a disposition to rule-following is an objective property of a subject. The dispositional property's being capable of manifesting itself in the form of behavior such as uttering "2 + 2 = 4" renders it subject to third-person, theoretical investigations. This, however, should not mislead us into regarding a disposition as an entity whose identity condition can only be accessed through an inferential process. Given a functionalist notion of dispositions and the Dennettian functionalist account of introspection, a functionally characterized disposition of a subject is manifested in the form of having thoughts or performing speech-acts. The subject explicitly or implicitly expresses $1 + 1 = 2$, $2 + 2 = 4$, $45 + 23 = 68$, and so on. The way by which a subject thinks about or verbally expresses those propositional contents is exactly the way by which the disposition to

follow the plus rule gets manifested. A subject, by virtue of having thoughts about the propositional contents or expressing them, thus gains a direct conscious access to the disposition. Therefore, KW's second premise is false. We do have a direct and non-inferentially conscious access to our dispositions.

Nevertheless, a subject's having a direct conscious access to his underlying disposition does not mean that his access to the disposition is complete. Each time (at t_1 , or t_2 , etc.) a disposition manifests itself in the subject's speech act or thought, the subject's acquaintance with the disposition is limited to one instance of an infinity that constitutes the rule-following disposition. Hence, there is no way that a subject's single (or a limited number of) occurrent thought(s) or speech act(s) can encompass all the instances of a rule. Does the direct-but-incomplete access to one's disposition then imply that one does not possess knowledge that he follows a certain rule? My answer is no.

The reason, in my view, is that, a person's attaching a meaning or rule-tag such as "I mean plus by '+'" or "I follow the plus rule" to his own overall dispositional condition provides the person with knowledge about what he means by a sign or which rule he follows. It will take two steps to argue for this point. The first step is to notice that there exists a two-level operation when a person follows a rule. The first level is the person's performing a particular calculation, such as adding 2 and 2 and getting an answer 4. The second (higher) level is the person's explicitly or implicitly judging or believing (a meta-cognitive act) that he is adding, or following the plus rule by attaching a "tag" such as "I am following the plus rule" to his own rule-following behaviors.

The second step of my argument is that given an externalist theory of justification such as the reliabilist theory of the sort outlined at the end of the last chapter, the belief or

judgment that one follows a particular rule would be justified as far as that belief is produced by a reliable belief-producing mechanism. As we have pointed out, a complex cognitive system of a higher sort is capable of producing reliable though not infallible information in the form of qualitative experiences about its own underlying processes. We can reasonably hold that man is a species born with such a high-level cognitive complexity which is capable of reliably producing meta-cognitive beliefs or judgments that are about what rules he follows. Therefore, provided that a person does, ontologically, possess a disposition to follow the plus rule, his reliably produced belief that he follows the plus rule would constitute knowledge for the person. As a result, despite the fact that a person only has an incomplete access to his own disposition via the direct manifestation of a disposition in the form of tokening a (higher-order) thought (such as having a thought that $2 + 2 = 4$) or performing a speech act (such as uttering “ $2 + 2 = 4$ ”), the person is also able to possess knowledge about what rule he follows by virtue of having an indirect meta-cognitive judgment or belief about his own disposition.

Therefore, I agree with Wittgenstein’s insight, and stand in opposition to Wright’s judgment-dependence view, that having such a belief or judgment of a propositional form does not constitute rule-following or meaning: rule-following and meaning are more likely to be constituted by some kind of dispositional capacities. Even though in the metaphysical sense attaching a meaning/rule tag does not constitute meaning or rule-following, in the epistemological sense it constitutes a piece of knowledge for the person who possesses the disposition.

It has to be noted that the phenomenology of rule-following has to be in turn distinguished from the epistemology and metaphysics of rule-following. In our first-

person experiences of rule-following, we do not feel as if we perform a two-step operation when we follow a rule. We don't first perform a particular calculation and then think that we are following a particular rule when we perform it. On the contrary, in our phenomenology, the higher-level meta-cognitive judgment or belief that I am following, say, the plus rule comes first in our introspective consciousness. The higher-level meta-cognitive acts usually precede my setting out to perform a calculation. My claim is that a reliabilist theory of justification delivers the status of knowledge for those meta-cognitive acts.

We may now re-examine the first premise of the argument (E) in this light. The first premise states, based on KW's intuitive understanding of meaning and rule-following, that the rule or meaning that a person follows or grasps has to be something that is directly accessible to the person. I must concede this premise because in my dispositional account of rule-following a disposition as a fact that constitutes rule-following is expressible in a direct manner and its possessor becomes consciously acquainted with it. But the fact that the rule-following disposition may be only partially accessed at any given time does not fit well with our commonsense picture of rule-following, that we seem to have a complete grasp of the rule that we follow. This apparent discrepancy between partial access to a disposition and complete grasp of a rule can be resolved by the fact that we have another way of accessing a disposition than expressing it directly. That is, we may have an implicit or explicit meta-cognitive belief or judgment, such as "I am following the plus rule," that points indirectly to the disposition of adding. The meta-cognitive judgment or belief corresponds to the complete grasp of the rule that we follow. Thus, once we pay sufficient attention to the

distinction between the metaphysics and knowledge involved in rule-following, we are able to see that this kind of meta-cognitive belief or judgment, though not constitutive of rule-following, not only provides one with knowledge about one's own rule-following disposition, but also explains that subjectively a rule (or meaning) seems to appear to one's introspective consciousness as an object grasped as a whole.

Finally, my dispositional account of rule-following can be further applied to explain the Wittgensteinian puzzle of why the verbal formula does not need further interpretation to fix the rule being followed. When the disposition underlying following a rule manifests itself, a salient occurrence in our introspective consciousness is a variety of thoughts or judgments, including the two kinds that we have talked about: 1) thoughts about particular performances such as $2 + 2 = 4$, $3 + 3 = 6$, *and so on*, and 2) thoughts regarding labeling a rule or meaning tag to one's overall disposition and behaviors, such as *I am following the plus rule*. There is another kind of thought or judgment that may occur in our conscious awareness when we follow a rule: we may think of a rule we follow not just as a tag such as "I am adding," but in the form of a formula, such as " $f(x) = x + 2$ " or " $f(x) = x + 2$ is the rule being followed." We perform calculations by making judgments regarding the rule $f(x) = x + 2$ accordingly, such as $f(x) = 4$ given x is 2, $f(x) = 6$ given x is 4. Our commonsense intuition seems to be that the rule-- $f(x) = x + 2$ --is sufficient to fix the rule that is being followed. Those particular performances are merely instances dictated by and deduced from the formula. But as Wittgenstein has convincingly shown us, the formula is unable to determine the rule being followed, because the signs in the formula such as "+" can be interpreted in various different ways and any chosen interpretation would allow further competing interpretations. It thus

leads to an interpretation regress. This appears to be paradoxical, for the formula seems to be self-interpreting but does not fix the rule being followed.

In my dispositional account, an agent's following the formula $f(x) = x + 2$ is determined by his disposition to perform in a certain way. The rule $f(x) = x + 2$ makes the agent feel that it commands him to answer 4 when x is set to 2 and 6 when x is set to 4. That is how the disposition is perceived: it is directly manifested in an agent's thinking or answering 4 when x is 2, 6 when x is 4, and so on. Therefore, we do introspect the procedures of applying the formula by virtue of thinking or uttering the procedures that the answer is 4 when x is 2 and the answer is 6 when x is 4, and so on. So, phenomenologically, there is no gap between the verbalization of the formula and its interpretation. Here my dispositionalist treatment does justice to McDowell's (see Chapter IV of this dissertation) intuition that rule-following and meaning are transparent to both the speaker and listener in the sense that they are understood without being mediated by interpretation. Moreover, my account goes a step further than McDowell's, in offering a deeper and more illuminating analysis of what people do when they follow a rule in the form of formula.

Moreover, in the Dennettian picture, we may not know exactly how we add 2 and 2 or 4 and 2, which is achieved by a complex first-order process that is not accessible to our introspective consciousness. In a functionalist account of cognitive processes and introspection, what we introspect is that *there is* a mechanism which produces 4 when 2 is put in. However, we cannot introspect what that mechanism is. Therefore, KW is right that *the rule is not the mechanism*. The verbal report of the introspected rule is merely the report that there is some mechanism.

The same phenomenological account of following a rule can be applied to the case of the counting proposal that we discussed in chapter II. KW's skeptic considers it as an intuitively plausible candidate that constitutes rule-following, because when we follow the plus rule, we seem to be guided by a finite set of instructions. For example, when we add, we first count one group of things one by one and then count another group also one by one. We then put two groups of things together and count, again one by one, the totality. No one, based on common sense, would deny that this procedure of counting is a legitimate way of adding and some of us may use this procedure to add on some occasions. KW's skeptic, however, rejects it on the ground that it is circular, because the set of instructions involves a term "counting" that already contains and presupposes a concept of rule. As a result, the set of instructions or counting procedures does not fix the rule being followed. This result appears quite counter-intuitive, but it is a solid point.

From the perspective of my dispositional account, the counting proposal fails because it wrongly takes a set of counting-instructions to constitute one's following the plus rule. Rather, what constitutes an agent's adding is his possessing a disposition to apply the set of counting procedures. An agent's disposition to apply the set of procedures is acquired through a process of learning, which is constrained in a certain way by the nature and structure of the agent's innately endowed biological mechanisms. Each particular step of applying those procedures is a manifestation of the counting-disposition which the agent is conscious of by having a thought about the application of a procedure or by directly expressing one particular step of the counting-procedures. Thus, the way in which the set of counting-procedures or instructions is propositionally embedded in a disposition and the way by which it gets "published" or exhibited explain

why phenomenologically the set of instructions feels like the fact that constitutes following a rule (like adding).

My account of the first-person phenomenology of rule-following is thus built on Martin & Heil's realist dispositional framework, especially heavily relying on their notion of cue-manifestation. The problem with Martin & Heil's account, however, is that it does not offer an adequate depiction of the content and structure of the phenomenological dimension of rule-following. Their account merely addresses the role played by the "feeling of knowing" experiences. That is, the feeling of knowing or feeling of certainty that arises in actual performances offers a subject a piece of direct information about his own underlying rule-following disposition. But this is far from adequate in depicting the phenomenology and epistemology of rule-following. There are also various forms of judgments and thoughts that occur in our introspective consciousness and play important roles in rule-following. My account brings in a functionalist characterization of a disposition and a Dennettian account of introspection that results in what I hope a more illuminating and adequate explication of the epistemic role played by judgments and beliefs of various propositional forms in rule-following.

My version of the dispositional account is similar to Yalowitz's dispositional account that we discussed in chapter VI. Yalowitz's account shows how a subject can know in a direct manner his own rule-following disposition without relying on inferences. According to Yalowitz, a subject has a direct access to his own underlying disposition by virtue of directly expressing it in speech acts. The Dennettian functionalist model of introspection can be viewed as an account that further elaborates Yalowitz's expressivist dispositional account. However, Yalowitz's account and mine differ in an

important way. In Yalowitz's account, an agent's performing a particular speech act by uttering, say, $2 + 2 = 4$, is sufficient for his having knowledge about which rule he follows, even though the content of the knowledge is merely partial relative to the whole range of the infinite instances of the rule being followed. Yalowitz argues for this position by emphasizing the distinction between the two kinds of knowledge: knowing-how knowledge and knowing-that knowledge. The former is a kind of procedure knowledge, such as being able to ride a bicycle, whereas the latter is a kind of propositional knowledge. The capacity to ride a bicycle consists in, as we may all agree, one's being able to actually perform relevant skills in coordinating his body with the bicycle. Being able to give detailed propositional knowing-that descriptions of how to ride a bicycle is neither sufficient nor necessary for doing it. Following a rule is thus analogous to riding a bicycle: both activities essentially involves a capacity to master and exercise certain skills without necessarily having verbal articulation of those skills. On the other hand, knowing-that knowledge in Yalowitz's account rule-following is the meta-cognitive belief or judgment that *I mean plus by "+"* or *I follow the plus rule*, and is propositional in nature. It unavoidably leads to an interpretation regress, and hence cannot constitute meaning and rule-following.

I agree with Yalowitz that there is a distinction between knowing-how and knowing-that knowledge and that meaning or rule-following consists in the dispositional capacity to know how to apply certain procedures. However, there are two things that are missing from Yalowitz's view. First, there is an element of propositional knowing-that knowledge involved in the knowing-how dispositional capacity that constitutes rule-following. When an agent's rule-following disposition manifests itself through

expressing a speech-act such as uttering “ $2 + 2 = 4$ ”, the content of the utterance is of a propositional form, that *2 plus 2 equals 4*. Yalowtiz’s claim that rule-following is a kind of knowing-how is thus inaccurate. The Dennettian functionalist model of introspection allows us to have a better understanding of the knowing-that aspect of rule-following.

Second, Yalowitz underplays the importance of meta-cognitive beliefs or judgments of a propositional form in human rule-following: I mean plus by “+” or I follow the plus rule. It’s correct to hold that, metaphysically, such propositional judgments do not constitute rule-following and meaning. However, it’s wrong to regard them as irrelevant to rule-following. Those meta-cognitive judgments or beliefs not only function as a guide in an agent’s rule-following activity, but they also display his knowledge about the rule he himself follows. Yalowtiz ignores the fact that these judgments or beliefs are an agent’s meta-cognitive acts of attaching certain “tags” to his own rule-following experiences and activities. The mental mechanism responsible for producing these meta-cognitive beliefs and judgments is apparently innately embedded in a complex biological system like man and gets developed during the agent’s learning history. Such a mental mechanism would be, under normal conditions, reliable and hence the beliefs it produces are justified. Thus, propositional judgments or beliefs in their very abstract form (e.g., I mean plus by “+”) can give us semantic or rule knowledge.

One of the major objections KW’s skeptic raises to a dispositional account of rule-following is the first-person epistemic problem. What I have shown is that the argument (E), which embodies the problem in question, contains a false premise, i.e., premise 2. As a result, the conclusion of the argument that rule-following or meaning is not constituted by a disposition does not follow. The key to rejecting the second premise

is to recognize the ways in which a disposition is manifested and the various forms the manifestations take. Once the structure and content of the phenomenological dimension of a disposition is uncovered and depicted by suitable cognitive model of introspection, various puzzles concerning rule-following may also be resolved.

IX. Conclusion and the Future Directions

The whole discussions on rule-following began from KW's skeptic who challenges us to cite a fact that can determine which rule we follow so that the fact can justify our particular applications of the rule. KW refuted various plausible candidate facts, such as an agent's past usage, a limited set of instructions or algorithm, a disposition, qualitative feelings, and so on, as adequate fact to ground the correctness condition that is essential to rule-following and meaning. KW therefore concluded that there is no such thing as rule-following or meaning, because there is no fact of the matter about an individual that can determine which rule a person follows and justifies the individual's relevant behaviors.

My major purpose in this study was to propose a naturalistic dispositional solution to the skeptical paradox. Specifically, I argued that KW's three major objections to the dispositional account, i.e., the finitude, the error, and the first-person epistemic problems will disappear if the misunderstanding regarding the metaphysical, methodological, and epistemological nature of human disposition can be cleared.

My first attempt was to show that from the metaphysical dimension, a disposition, following the insight of Martin & Heil, is best construed realistically, rather than counterfactually. A rule-following disposition is a subject's robust property, and has to be distinguished from its manifested rule-following behaviors under various conditions. Once the distinction is made, erroneous behaviors are those occurrences where the subject's realistic disposition fails to manifest itself due to the presence of its inhibitory dispositional partners or the absence of its reciprocal dispositional partners. The error problem KW raised against dispositional account of rule-following does not exist any

longer. The fact that a disposition is capable of only finitely many manifestations leads KW to hold mistakenly that the disposition is finite. However, a disposition, when construed realistically, is infinite in nature. A subject (or object) may be finite for possible manifestations of a disposition due to his limited capacity and lifespan, his disposition to follow a certain rule is “there-and-ready” for a possible manifestation under any one of a potential infinity of conditions. The finitude problem of the dispositional account raised by KW’s skeptic is then unsubstantiated.

My second attempt was to show that from the methodological dimension, given his manifested behaviors, which rule-following disposition a person possesses can be specified via a process of scientific hypothesizing. On a first glance a certain degree of vacuity or seeming circularity is unavoidable in such an enterprise of hypothesizing (but with a positive methodological purpose) due to its implicit employment of *ceteris paribus* clauses or normality conditions. Nevertheless, a sufficient condition for the non-vacuity of a hypothesis invoking a cp-law can be offered. As a result, a suitable criterion to distinguish legitimate from illegitimate disposition-ascribing hypotheses can be given, and disposition-ascribing practice is thus epistemologically grounded. The epistemic worry that we are not sure whether the rule a person’s realistic disposition of an infinitary character embodies is *recursive* or *requusive* is resolved by adopting a reliabilist theory of justification. I do not have to wait to see what my future performance is in a not yet encountered problem case in order to know whether or not the performance falls under the same pattern of instances dictated by the rule I used to follow in the past. As far as it is true that my future performance will be correct and I believe that it will be correct, my belief that it will be correct would be justified, given that this belief is produced by a

reliable mental mechanism and given a reliabilist theory of justification. Consequently, a suitable epistemological ground can be delivered for our inductive belief that the person's disposition embodies the recursive, rather than reclusive, rule.

Lastly, from the epistemological dimension, I tried to explicate how a subject is capable of having direct and non-inferential knowledge about his own underlying rule-following disposition. There are two ways for a subject to know his own underlying rule-following disposition. One way is to conduct, as we have just seen from the methodological perspective, scientific theorizing and inductive reasoning based on subject's manifested behaviors and relevant empirical evidences. This way of knowing one's disposition is essentially third-person, indirect, and inferential. However, one may also know his own rule-following disposition in an essentially first-person, direct, and non-inferential manner. The key to see how first-person knowledge about one's own rule-following disposition is possible is to pay sufficient attention to the way in which a rule-following disposition is manifested and the variety of forms in which the manifestations take. One way a rule-following disposition manifests itself is by virtue of the disposition's being directly expressed by its possessor in a speech act or being directly thought about in a propositional form such as *2 plus 2 equals 4*, *3 plus 3 equals 6*, and so on. Therefore, a subject does have a direct conscious contact with his underlying disposition—i.e., with the fact that constitutes rule-following. However, this way of directly accessing a disposition only brings an incomplete piece of knowledge to a subject regarding what his own underlying disposition to rule-following is. For those limited manifested speech-acts or thoughts still under-determine the rule that dictates an infinity of instances. Consequently, a subject may still remain ignorant exactly what

rule-following disposition he possesses in such instances. My claim is that the cue-manifestations of a rule-following disposition play an important role here. When following a rule, one has various feeling of knowing and feeling of certainty experiences which serve as a reliable, though fallible, indicator of a subject's own underlying dispositional condition. These feelings inform the subject about whether or not his own underlying sub-systems are functioning normally in producing relevant rule-following behaviors. More importantly, a subject is able to issue verbal tags, such as "I am following the plus rule", that give the subject a very explicit piece of information about what his own rule-following disposition is. Even though, as KW pointed out, labeling such a verbal tag to oneself does not amount to the fact that constitutes rule-following, it constitutes a piece of knowledge about rule-following. Humans are endowed with complex cognitive mechanisms. Being able to issue a verbal tag is a certain kind of meta-cognitive act that is produced by some reliable mechanism to bring about a belief about the rule that he follows. Since the belief in question is produced by a reliable mental mechanism, it suffices as knowledge, given a reliabilist theory of knowledge. The person thus has a direct and non-inferential knowledge about the rule/meaning he is following.

I have thus shown that all the three major objections, i.e., the error, finitude, and first-person epistemic problems, that KW's skeptic raises against the dispositional account of rule-following can be suitably rejected. What may we then conclude? Can we consequently conclude that rule-following or meaning is constituted by a disposition? Two different versions of conclusion regarding dispositional account of rule-following, a strong one and a weak one, can be drawn as follows:

(S) The strong conclusion: rule-following or meaning is constituted by a disposition, namely, a disposition is both necessary and sufficient for rule-following and meaning.

(W) The weak conclusion: rule-following or meaning may not be constituted by a disposition; a disposition is necessary but not sufficient for rule-following and meaning.

Logically speaking, the refutation of KW's three major objections against a dispositional account of rule-following and meaning at most leads to the acceptance of (W), rather than (S). According to KW's skeptical reasoning, for a disposition to constitute meaning, it has to be an infinite entity of some sort, can ground the correctness condition, and be known in a direct and non-inferential manner. That means, all the three constraints in question are necessary conditions for a disposition to be a fact that constitutes rule-following and meaning. However, even if all the three constraints are fulfilled by a dispositional fact about an individual, it does not necessarily imply that a disposition is sufficient as a fact that constitutes meaning and rule-following. There might be other elements that are crucial to meaning and rule-following but KW's skeptic has failed to consider. A suitable conclusion to draw is therefore in a weak sense that KW's skeptical claim that there is no fact about an individual regarding whether he follows one rule or another does not sustain. In other words, the reality of rule-following and meaning is vindicated despite KW's skeptical argument.

My position is, at the moment, to accept the weak conclusion (W), that a disposition is a necessary and indispensable factor for rule-following and meaning. But disposition alone may not be sufficient to constitute meaning and rule-following. My

position thus amounts to offering a non-reductionist dispositional solution to KW's skeptical paradox. For to make a reductive claim that meaning or rule-following is constituted by a disposition, a disposition has to be both necessary and sufficient for meaning and rule-following.

An alternative to the weak, non-reductive view of dispositional account of rule-following is to accept the strong conclusion (S). That is to claim that a disposition is a necessary and sufficient condition for rule-following and meaning. Any objection to the strong conclusion will invariably point out the missing properties or conditions that a disposition has to possess in order to constitute rule-following and meaning. I will leave this as an open question for the future studies. However, I will point out some possible plausible directions for future investigation below. If a disposition is necessary but not sufficient for rule-following and meaning, what feature might a disposition fail to encompass or be coupled with?

Teleological property

One feature I can think of to be such a candidate is a teleological property. Recall that we have discussed Millikan's teleological account of meaning and rule-following in Chapter VI. In Millikan's view, the normative element of meaning and rule-following, i.e., the correctness condition, has to be grounded in the biological function or purpose fulfilled by the operations of an organism's mechanism such that the chance of its reproductive success may be enhanced. I argued that such a biological property is not specific enough to determine the meaning of one's word or the content of one's mental state. Nevertheless, the teleological notion in Millikan's account may capture something crucial that helps characterize meaning, mind, and rule-following. That is, rule-following

behaviors and acts of similar sorts are produced by certain cognitive mechanisms of an organism to fulfill certain biological function such that the fitness of the organism may be improved. Moreover, the existence of the rule-following mechanisms can be explained by its having brought about a reproductive success in the evolutionary history of the organism's species. The notion of a biological function appears unable to be completely captured by the concept of a disposition, which is characterized in terms of a causal network in an organism. Therefore, if we accept the teleological view that the notion of biological purposiveness is a key element in understanding meaning and rule-following, a disposition would be insufficient to constitute meaning and rule-following.

Two other open questions that might pose a threat to my naturalistic realist dispositional account of rule-following and meaning are listed below.

The problem of self-knowledge

When we follow a rule or mean something by a term, we seem to know the rule or meaning in a direct, non-inferential, and authoritative manner. How can this special way of knowing be explained? As we saw in Chapter IV, Wright's judgment-dependence account was specifically offered to account for the first-person dimension of the rule-following and meaning phenomena. My criticisms of Wright's account focused on the metaphysical aspect of the same phenomena. That is, I argued that even if we grant that Wright's judgment-dependence account provides us with a good account of self-knowledge, it is not satisfactory as a metaphysical account of rule-following and meaning. Nonetheless, it remains an open question that Wright's judgment-dependence account may serve as a good account of the epistemological dimension of the rule-following and meaning problem. Wright's account is a constitutive account in which

there exists an a priori and constitutive relation between one's meaning something by a term and one's judging that he means something by the term. That the relation between one's meaning something by a term and one's judging that he means something by the term is constitutive implies that the relation is "unbreakable". Specifically, there cannot be a judgment about one's own meaning without one's actually possessing that meaning, and there cannot be any case of one's having certain meaning without judging that he has that meaning. Contrary to this kind of constitutive account of self-knowledge is the cognitive-model approach such as the one taken by Dennett or Rosenthal that I discussed in Chapter VIII. On the cognitive-model approach, the way by which one comes to know or introspect one's own intention or meaning state is subserved by certain cognitive or functional processes. Hence, the relation between one's meaning something by a term and one's judging that he means something by the term is empirically independent. If Wright's constitutive approach turns out to be a better account of self-knowledge than the cognitive-model approach, then the credibility of my dispositional account of rule-following and meaning would be endangered, for my dispositional account relies on the cognitive-model approach. It would thus be an important task to compare the plausibility of each approach to self-knowledge.

The objectivity and necessity of rules

My naturalistic dispositional account attempts to account for the *ontology* of *rule-following*. In particular, as far as a realistic disposition can accommodate the three important elements of a rule, i.e., the infinite feature of some sort, the correctness condition or normativity, and first-person accessibility, it is ontologically sufficient to account for rule-following. However, the realist dispositional account ignores some

crucial questions regarding the *ontology* of *rules*. Rules are, on our intuitive idea about them, objective. They dictate an infinity of instances which seem to exist independently of human dispositions and judgments. Rules, in other words, are objective. On our commonsense intuition, rules provide objective standards by which human rule-following behaviors are measured. Moreover, the relation of rules to their instances and each other also appear to be necessary. We cannot think of a possible world in which the add-2 rule $f(x) = x + 2$ fails to dictate $f(2) = 4$ or $g(f(x))$ fails to dictate $g(f(6)) = 24$ where $g(x)$ is the multiply by 3 rule $g(x) = x * 3$. A problem for my naturalistic realist dispositional account of rule-following is then how it can accommodate the fact that rules are objective and their relations to their own instances or other rules are necessary. The kind of objectivity that human rule-following dispositions possess appears may be different from the kind of (more rigid) objectivity that rules enjoy. Besides, human dispositions may be too contingent to ground the feature of necessity in the relation between rules and their instances. We may easily imagine a possible world in which all human beings possess the dispositions to respond 3 to “ $2 + 2 = ?$ ”. However, the mathematical fact that $2 + 2 = 4$ in the plus rule remains true in every possible world. There thus appears to be a gap between the dispositional account serving as a metaphysical depiction of rule-following and the ontology of rules. The question is how the gap might be bridged.

We may use an example to illustrate this point. Consider the add-2 rule $f(x) = x + 2$. Once its definition is laid down, there follows an infinite number of pairs $(x, f(x))$ which are the outcomes of applying this rule. This seems to be an objective and necessary fact about the $f(x)$ rule. On the other hand, whether or not a rule, like $f(x) = x + 2$, is followed by a subject is to be determined by the subject's dispositions, according

to a naturalistic dispositional account. It does not seem plausible that we may identify any objective properties for the $f(x)$ rule merely from the subject's dispositions. It's equally difficult to see that the necessity of the $f(x)$ rule can consist in the fact that the subject, or everyone else is disposed to agree that $f(2)$ equals 4 under ideal conditions. How can the fact that I have a "real" disposition to compute the $f(x)$ rule explain that $f(2) = 4$? The naturalistic dispositional account of rule-following seems insufficient to answer this question.

A tentative solution to the problem is to rigidly fix the references of a rule by past stable performances a subject is disposed to produce under ideal conditions. In this approach, a meta-statement such as " $f(x)$ denotes the add-2 rule" is necessary but a posteriori, and the objectivity of the rule is no more than people's dispositions to add numbers. More thoughts, of course, are needed to examine whether or not this approach will work successfully. Due to the scope of this dissertation, I will have to leave it to the further research project. More broadly construed, it will be a task to investigate the relation between reliabilism in epistemology and Platonism in metaphysics.

To summarize, my unique contribution to the debate on KW's rule-following paradox is showing how a plausible dispositional solution can be offered. Consequently, we can derive the weak conclusion that the existence of rule-following and meaning is not threatened by KW's skeptical argument. Moreover, by virtue of giving a unified treatment of a rule-following disposition, human rule-following phenomenon is explicated more clearly. In the treatment in question, a disposition is a theoretical entity that is ascribed to both physical and psychological realms for the explanation and prediction of behaviors. A human rule-following disposition is specifically innately

constrained in a certain way and is significantly shaped by an individual's learning history, which can together, at least in principle, determine the rule an individual follows. Moreover, a rule-following disposition is something that can be known directly by its possessor. I offer what I hope an illuminating depiction of the first-person phenomenology of rule-following by exploring some plausible functionalist theories of dispositions and introspective consciousness. In my version of the dispositional account, the conceptual resources in reliabilism are exploited to account for the justification and knowledge we have about the rule we follow and the meaning we grasp.

However, there exists a gap between having a plausible dispositional solution to KW's skeptical paradox and accepting the strong conclusion that a real disposition is both necessary and sufficient for rule-following. Therefore, it remains unclear whether or not a real disposition is sufficient to constitute rule-following or meaning. Further research is needed to answer this remaining question.

Bibliography

- Armstrong, D.M. 1968: *A Materialist Theory of the Mind*, New York: Humanities Press.
- 1973: *Belief, Truth, and Knowledge*. Cambridge: Cambridge University Press.
- Asch, S. 1952: *Social psychology*. Englewood Cliffs, Nj: Prentice Hall.
- Ashcraft, M.H. and Battaglia, J. 1978: Cognitive arithmetics: evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory*, 4, p. 527-538.
- Atran, S. 1990: *Cognitive Foundations of natural history*. New York: Cambridge University Press.
- 1995: Causal constraints on categories and categorical constraints on biological reasoning across cultures. In D. Sperber, D. Premack, & A.J. Premack (Eds.) *Causal cognition: A multidisciplinary debate* (pp. 205-233). Oxford, England: Oxford University Press.
- Baillargeon, R. 1995: Physical reasoning in infancy. In M.S. Gazzaniga (Ed.) *The cognitive neurosciences* (pp. 181-204). Cambridge, MA: MIT Press.
- Beakley, B., & Ludlow, P., eds. 1992: *Philosophy of Mind: Classical Problems/Contemporary Issues*. Cambridge, MA: MIT Press.
- Berkeley, G. 1948-57: *The Works of George Berkeley, Bishop of Cloyne*, eds. A.A. Luce and T.E. Jessop, Edinburgh: Thomas Nelson.
- Berlin, B. 1992: *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.
- Berlin, B., Breedlove, D., & Raven, P. (1973). General principles of classification and nomenclature in folk biology. *American Anthropologist*, 74, 214-242.
- Blackburn, S. 1984: The Individual Strikes Back. *Syntheses* 58, pp. 281-301.
- 1984: *Spreading the Word*, Oxford: Oxford University Press.
- Blake, M. 1973: Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 12, 311-319.
- Block, N. 1980: *Readings in Philosophy of Psychology*. Cambridge, Mass.: Harvard University Press.

- Boghossian, P. 1989(a): The Rule-Following Considerations, *Mind* 98, pp. 507-49.
- 1989(b): Review of McGinn's "Wittgenstein on Meaning", *Philosophical Review*, 98, pp. 83-92.
- 1990(a): The Status of Content, in *Philosophical Review* 99, pp. 157-184.
- 1990(b): The Status of Content Revisited, *Pacific Philosophical Quarterly*, 71, pp. 264-78.
- 1991: Naturalizing Content, *Meaning in Mind: Essays for Jerry Fodor*, B. Loewer & G. Rey (eds.). Cambridge: Mass.:MIT Press.
- Bradley, J.V. 1981: Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society*, 17, 82-84.
- Butterworth, B. 1999: *What Counts: How Every Brain Is Hardwired for Math*. New York: Simon and Shuster, The Free Press.
- Carey, S., and Gelman, R. (eds.) 1991: *The Epigenesis of Mind*. Lawrence Erlbaum Associates, Inc.
- Cartwright, N. 1983: *How the Laws of Physics Lie*. Oxford: Oxford University Press.
- Chomsky, N. 1986: *Knowledge of Language*, New York:Prager.
- Coates, P. 1986: Kripke's Sceptical Argument: Normativeness and Meaning, *Mind* 45, pp. 77-80.
- Cohen, L.B. 1988: An Information-Processing Approach to Infant Cognitive Development. In L.Weiskrantz (ed.) *Thought without Language*. Oxford University Press.
- Crane, T. (ed.) 1996: *Dispositions: A Debate*, London. Routledge.
- D'Andrade, R. (1987). A folk model of the mind. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 112-148). New York: Cambridge University Press.
- Davidson, D. 1984: *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- Dehaene, S. 1997: *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.
- 2001: *Precis of The Number Sense, Mind and Language*. Vol 16.

- Dehaene. S., Spelke, E., Stanescu, R., Pinel, P. and Tsivkin, S. 1999: Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284, p. 970-974.
- Dennett, D.C. 1969: *Content and Consciousness*, London: Routledge and Kegan Paul.
- 1978: *Brainstorms: Philosophical Essays on Mind and Psychology*, Bradford Books.
- 1991: *Consciousness Explained*, Boston: Little, Brown and Company.
- Devitt, M. 1990: Transcendentalism about Content, *Pacific Philosophical Quarterly*, 71, pp. 247-63.
- 1991: *Realism and Truth*. Princeton: Princeton University Press.
- Devitt, M., & Rey, G. 1991: Transcending Transcendentalism, *Pacific Philosophical Quarterly*, 72, pp. 87-100.
- Dretske, F. 1981: *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.
- Dummett, M. 1991: *The Logical Basis of Metaphysics*. Cambridge, Mass.: Harvard University Press.
- 1993: *The Seas of Language*. Oxford: Oxford University Press.
- Edwards, J. 1992: Best Opinion and Intentional States, *The Philosophical Quarterly*, 46, pp. 21-33.
- Evans, J. St.B.T. 1989: *Bias in Human Reasoning: Causes and Consequences*. Lawrence Erlbaum Associates Ltd.
- Flavell, J.H., F.L.Green, & E.R.Flavell, 1995a: Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development*, 60, 1 (serial no. 243): 1-95.
- Flavell, J.H., F.L.Green, & E.R.Flavell, 1995b: The development of children's knowledge about attentional focus. *Developmental Psychology*, 31, 4: 706-712.
- Fodor, J. 1975: *The Language of Thought*. New York: Thomas Y. Crowell, Co.
- 1981: *Representations*. Cambridge, MA: MIT Press.
- 1987: *Psychosemantics*. Cambridge, Mass.:MIT Press.

- 1990: *A Theory of Content and Other Essays*. Cambridge, Mass.:MIT Press.
- Forrest-Presley, D.L., G.E. Machinnon, & T.G. Waller, (Eds.) 1985: *Metacognition, Cognition, and Human Performance*. Orlando: Academic Press.
- Gampel, E.H. 1997: The Normativity of Meaning, *Philosophical Studies*, 86, pp. 221-242.
- Garcia, J., McGowan, B.K., Ervin, F., & Koelling, R. (1968). Cues: Their relative effectiveness as reinforcers. *Science*, 160, 794-795.
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychologist*, 20, 65-95.
- George, A. (ed.) 1989: *Reflections on Chomsky*. Oxford: Basil Blackwell.
- Ginet, C. 1992: The Dispositionalist Solution to Wittgenstein's Problem about Understanding a Rule, *Midwest Studies in Philosophy* 17, pp. 53-73.
- Gleitman, L.R., Gleitman, H., Landau, B., & Wanner, E. 1988: Where Learning Begins: Initial Representations for Language Learning. In F. Newmeyer (ed.), *The Cambridge Linguistic Survey* (pp. 150-193). New York: Cambridge University Press.
- Goldfarb, W. 1985: Kripke on Wittgenstein on Rules, *Journal of Philosophy* 85, p. 471-488.
- Goldman, A.I. 1979: What Is Justified Belief? In G. Pappas.
- 1986: *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- 1993: *Readings in Philosophy and Cognitive Science*. Cambridge, MA: MIT Press.
- Goodman, N. 1954: *Fact, Fiction, and Forecast*. Atlantic Highlands, NJ: Athlone Press.
- Grice, P. 1957: "Meaning", *Philosophical Review*, 66, pp. 377-88.
- Gruneberg, M.M., Monks, J., & Sykes, R.N. 1977: Some methodological problems with feeling of knowing studies. *Acta Psychologica*, 41, 365-371.
- Heal, J. 1990: *Fact and Meaning: Quine and Wittgenstein on the Philosophy of Language*. Oxford: Basil Blackwell.
- Hempel, C. 1988: Provisoos: A Problem Concerning the Inferential Functions of Scientific Theories, *Erkenntnis*.

- 1977: Dispositional Explanation, in *Dispositions*, Tuomela, R. (ed.)
- Hoffman, P. 1985: Kripke on Private Language, *Philosophical Studies*.
- Holton, R. 1993: Intentional Detecting, *Philosophical Quarterly*, 43, pp. 298-318.
- Holtzman, S.H. and Leich, C.M. (eds), 1981: *Wittgenstein: To Follow A Rule*. London: Routledge and Kegan Paul.
- Hume, D. 1952: *An Enquiry Concerning Human Understanding*. Chicago: Encyclopedia Britannica.
- Johnson, M.H. & Karmiloff-Smith, A. 1989: Neural Development and Language Acquisition: Parcellation and Selective Stabilization.
- Johnson-Laird, P. 1983: *Mental Models*. Cambridge, Mass.: Harvard University Press.
- Johnson-Laird, P., and Wason, P. 1970: A Theoretical Analysis of Insight into a Reasoning Task, and Postscript—1977, in J Johnson-Laird, P., and Wason, P. eds. (1977)
- Johnson-Laird, P., and Wason, P. eds., 1977: *Thinking*. Cambridge: Cambridge University Press.
- Johnson-Laird, P., and Bara, B. 1984: Syllogistic Inference. *Cognition* 16: 1-61.
- Johnson-Laird, P., and Byrne, R. 1992: Modal Reasoning, Models, and Manktelow and Over. *Cognition* 43: 173-82.
- Katz, J. 1990: *The Metaphysics of Meaning*. MIT Press.
- Kim, J., & Sosa, E. 1995: *A Companion to Metaphysics*. (Eds.) Basil Blackwell.
- Koechlin, E., Dehaene, S. and Mehler, J. 1997: Numerical transformations in five month old human infants. *Mathematical Cognition*, 3, p. 89-104.
- Koriat, A., & Lieblich, I. 1974: What does a person in a "TOT" state know that a person in a "don't know" state doesn't know. *Memory & Cognition*, 2, 647-655.
- 1977: A study of memory pointers. *Acta Psychologica*, 41, 151-164.
- Kripke, S. 1982: *Wittgenstein on Rules and Private Language*. Oxford and New York: Blackwell.
- Kukla, A. 1992: Endogenous Constraints on Inductive Reasoning. *Philosophical Psychology*, 5(4), pp. 411-25.

- Lachman, J.L., & Lachman, R. 1980: Age and actualization of world knowledge. In L.W.Poon, J.L.Fozard, L.S.Cermak, D.Arenberg, & L.W.Thompson (Eds.), *New Directions in Memory and Aging* (pp. 285-311). Hillsdale, NJ:Erlbaum.
- Leonesio, R.J., & Nelson, T.O. 1990: Do different metamemory judgments tap the same underlying aspects of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 464-470.
- Lepore, E. (ed.) 1986: *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.
- Leslie, A.M. 1982: The perception of causality in infants. *Perception*, 11, 173-186.--- 1994. ToMM, ToBY, and agency: Core architecture and domain specificity. In L.A. Hirschfeld & S.A. Gelman (Eds.) *Mapping the mind: Domain specificity in cognition and culture*. Cambridge, England: Cambridge University Press.
- Levin, M. 1993: Reliabilism and Induction, *Synthese*, 97 (3), pp. 297-334.
- Loar, B. 1985: Review of Kripke (1982). *Nous* 19.
- Loewer, B & Rey, G (eds.) 1991: *Meaning in Mind*, Cambridge: Mass.: MIT Press.
- Lycan, W. 1990: *Mind and Cognition: A Reader*. Oxford: Basil Blackwell.
- Mandler, J.M. 1988: How to Build a Baby: On the Development of an Accessible Representational System. *Cognitive Development*, 3, 113-136.
- Mehler, J. & Bertoncini, J. 1988: Development: A Question of Properties. Not Change? *International Social Science Journal, Cognitive Science*, 115, 121-133.
- Martin, C.B. 1994: Dispositions and Conditionals, *Philosophical Quarterly*, 44, pp. 1-8.
- Martin, C.B & Heil, J. 1998: Rules and Powers, *Philosophical Perspectives* 12. Blackwell.
- McDowell, J. 1984: Wittgenstein on Following a Rule, *Synthese*, 58 (3), pp. 325-63.
- 1992a: Meaning and Intention in Wittgenstein's Later Philosophy. *Midwest Studies in Philosophy* 17, pp. 40-52.
- 1992b: Intentionality and Interiority in Wittgenstein. In *Meaning Scepticism*, K. Paul (ed.) (Dordrecht: Reidel, 1992)
- McGinn, C. 1984: *Wittgenstein on Meaning*. Oxford: Basil Blackwell.

- McLaughlin, B. 1991: *Dretske and His Critics*. Cambridge, Mass.: Basil Blackwell.
- Miller, A. 1989: An Objection to Wright's Treatment of Intention. *Analysis*, 49, pp. 169-173.
- 1998: *Philosophy of Language*. McGill-Queens University Press.
- 2000: Horwich, Meaning, and Kripke's Wittgenstein. *The Philosophical Quarterly*, 50, pp. 161-174.
- Miller, G.A. 1956: The magical number, seven plus or minus two: some limitations on our capacity for processing information. *Psychological Review* 63: 81-97.
- Millikan, R. 1987: *Language, Thought, and Other Biological Categories*. Cambridge: MIT Press.
- 1990: Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox. *The Philosophical Review*, 99, pp. 323-354.
- 1993: *White Queen Psychology and Other Essays for Alice*. Cambridge: MIT Press.
- Mumford, S. 1998: *Dispositions*. Oxford: Oxford University Press.
- Nelson, T.O. 1984: A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109-133.
- Nelson, T.O., Gerler, D., & Narens, L. 1984: Accuracy of feeling of knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, 113, 282-300.
- Nelson, T.O., Leonesio, R.J., Landwehr, R.S., & Narens, L. 1986: A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 279-287.
- Nelson, T.O., & Narens, L. 1990: Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol.26). New York: Academic Press.
- Nozick, R. 1982: *Philosophical Explanation*. Cambridge, Mass.:Harvard University Press.
- Pappas, G., ed. 1979: *Justification and Knowledge*. Dordrecht: Reidel.
- Pettit, P. 1990(a): The Reality of Rule-Following, *Mind*, 99, pp. 1-21.

- 1990(b): Affirming the Reality of Rule-Following, *Mind*, 99, pp. 433-449.
- 1993: *The Common Mind: An Essay on Politics, Psychology, and Society*, Oxford: Oxford University Press.
- Pietroski, P. & Rey, G. 1995: When Other Things Aren't Equal: Saving *Ceteris Paribus* Laws from Vacuity, *British Journal for the Philosophy of Science*, 46 (1), pp. 81-110.
- Puhl, Klaus (ed.) 1991: *Meaning Scepticism*. Berlin: de Gruyter.
- Quine, W.V.O. 1953: *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.
- 1960: *Word and Object*. Cambridge, Mass.: MIT Press.
- 1970: *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice-Hall.
- Reder, L.M. 1987: Strategy selection in question answering. *Cognitive Psychology*, 19, 90-138.
- 1988: Strategic control of retrieval strategies. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22). San Diego, CA: Academic Press.
- Reder, L.M., & Ritter, F.E. 1992: What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 435-452.
- Rey, G. 1997: *Contemporary Philosophy of Mind*. Cambridge, Mass.: Blackwell.
- Rosenthal, D.M. 1986: Two Concepts of Consciousness, *Philosophical Studies* 49.
- 1990: Why Are Verbally Expressed Thoughts Conscious? Report No. 32: Center for Interdisciplinary Research (ZiF). University of Bielefeld.
- 1991: *The Nature of Mind*. (ed.) Oxford University Press.
- 1993: Thinking That One Thinks. In *Consciousness: Psychological and Philosophical Essays*, M. Davies & G. Humphreys (eds). Oxford: Basil Blackwell.
- Ryle, G. 1948: *The Concept of Mind*. Chicago University Press.
- Sartorelli, J. 1991: McGinn on Concept Scepticism and Kripke's Sceptical Argument. *Analysis*, 51, pp. 79-84.
- Schiffer, S. 1972: *Meaning*. Oxford: Clarendon Press.

- Schreiber, T.A., & Nelson, D.A. 1993: Feelings of knowing and retrieval processes. Unpublished manuscript.
- Searle, J. 1992: *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Seligman, M.E.P. 1970: On the generality of the laws of learning. *Psychological Review*, 77, 127-190.
- Sellars, W. 1963: Empiricism and the Philosophy of Mind. in *Science, Perception, and Reality*.
- Shogenji, T. 1992: Boomerang Defense of Rule-Following, *The Southern Journal of Philosophy*, 30, pp. 115-122.
- 1993: Modest Skepticism about Rule-Following, *Australasian Journal of Philosophy*, 71, pp. 486-500.
- 2000: The Problem of the Criterion in Rule-Following. *Philosophy and Phenomenological Research*, 60 (3), pp. 501-525.
- Simon, T.J., Hespos, S.J. and Rochat, P. 1995: Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development*, 10, p. 253-269.
- Soames, S. 1997: Skepticism about Meaning: Indeterminacy, Normativity, and the Rule-Following Paradox, in Meaning, Reference, and Truth: *The Canadian Journal of Philosophy, Supplementary Volume*, 23.
- 1998: Facts, Truth Conditions, and the Skeptical Solution to the Rule-Following Paradox, in *Philosophical Perspectives* 12.
- Spelke, E.S. 1988: Where perceiving ends and thinking begins: The apprehension of objects in infancy, In A. Yonas (Ed.), *Perceptual development in infancy. Minnesota Symposim on Child Psychology* (Vol. 20, pp. 191-234). Hillsdale, NJ: Erlbaum.
- 1990: Principles of object perception. *Cognitive Science*, 14, 29-56.
- Sullivan, P. 1994: Problems for A Construction of Meaning and Intention, *Mind*, 103, pp. 147-168.
- Tomberlin, J., ed., 1988: *Philosophical Perspectives*, 2: Epistemology. Atascadero, California: Ridgeview.
- Tuomela, R. (ed.) 1978: *Dispositions*, Dordrecht, Reidel.

- VanLehn, K. 1990: *Mind Bugs: The Origins of Procedural Misconceptions*. Cambridge, Mass.: MIT Press.
- Wason, P. 1968a: Reasoning about a Rule, *Quarterly Journal of Experimental Psychology* 20: 273-81.
- 1968b: On the Failure to Eliminate Hypotheses...---A Second Look, in Johnson-Laird and Wason, eds., 1977.
- 1977: Self-Contradiction, in Johnson-Laird and Wason, eds., 1977.
- Wason, P., and Johnson-Laird, P. 1970: A Conflict between Selecting and Evaluating Information in an Inferential Task, *British Journal of Psychology* 61: 509-15.
- Weinert, F.E. & Rainer, H.K. (Eds.) 1987: *Metacognition, Motivation, and Understanding*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Wellman, H.M. 1990: *The child's theory of mind*. Cambridge, MA: MIT Press.
- Wittgenstein, L. 1974: *Philosophical Investigation*, Oxford: Basil Blackwell.
- 1978: *Remarks on the Foundations of Mathematics*. 3rd edn. Oxford: Blackwell.
- Wright, C. 1984: Kripke's Account of the Argument against Private Language. *Journal of Philosophy*, 81 (12), pp. 759-78.
- 1987: On Making Up One's Mind: Wittgenstein on Intention, in Weingartner and Shurz (eds.), *Logic, Philosophy of Science and Epistemology. Proceedings of the 11th International Wittgenstein Symposium*, Kirchberg. Nienna: Holder-Pichler-Temsky.
- 1988: Moral Values, Projection, and Secondary Qualities, in *Proceedings of the Aristotelian Society* 62, suppl. Vol, pp. 1-26.
- 1989a: Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics, *Reflections on Chomsky*, A. George (ed.) Oxford: Basil Blackwell.
- 1989b: Critical Notice of McGinn 1984, *Mind*, 98, pp. 289-305.
- 1992: *Truth and Objectivity*, Cambridge, Mass.: Harvard University Press.
- 1993: Eliminative Materialism: going concern or passing fancy? In *Mind and Language*, 8 (2), pp. 316-26.
- 1995: Can There Be A Rationally Compelling Argument for Anti-Realism about Ordinary ("Folk") Psychology? *Philosophical Issues*.

- Wynn, K. 1990: Children's understanding of counting. *Cognition*, 36, p. 155-193.
- 1992: Addition and subtraction by human infants. *Nature*, 358, p. 749-750.
- Yalowitz, S. 2000: A Dispositional Account of Self-Knowledge, *Philosophy and Phenomenological Research*, 111, pp. 249-278.
- 2001: Individualism, Normativity, and the Epistemology of Understanding, *Philosophical Studies* 102, pp. 43-92.
- Zalabardo, J. 1997: Kripke's Normativity Argument, *Canadian Journal of Philosophy*, 27, pp. 467-488.