Order Number 9510715

# Following a rule: A critique of algorithmic models of meaning and morality

Schoelkopf, Constantin, Ph.D.

City University of New York, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

# Following A Rule:

## A Critique of Algorithmic Models of Meaning and Morality

by

Constantin Schoelkopf

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

1994

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.

9/20/94 — Date

Chair of the Examining Committee

9/26/94 — Date

Executive Officer

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

**Abstract**

**Following A Rule:**
**A Critique of Algorithmic Models of Meaning and Morality**

by

Constantin Schoelkopf

Adviser: Bernhard Baumrin

The dissertation attempts to show why the semantic paradox as developed in Kripke's book on Wittgenstein has received so much philosophical attention. I argue that philosophy of language since Frege has adopted a certain model of language, maintaining that language is a calculus and that the meaning of a word is essentially rule-like. I also attempt to show that Kant's deontological moral theory, and rule-utilitarianism can be understood to consider rules as essential for morality. The rule-following paradox Kripke develops apparently shows that rule-following is impossible, jeopardizing the views mentioned. I try to show that Kripke's meaning skepticism and the view of language and morality as calculi share the same presuppositions, and rest on the fact that philosophers misunderstand the terms they use in order to create philosophical theories.

**Acknowledgments**

There are many people to thank for their support and interest, but I hope I shall be forgiven if I mention only those whose advice has been invaluable. I am especially indebted to my adviser, Bernhard Baumrin, without whose generosity and kindness I would have not been able to finish this dissertation. I would also like to thank Juliet Floyd whose advice was essential for writing this dissertation. Both philosophers have spent much of their time helping me, and I am very grateful.

## Table of Contents

## Introduction

Over the last decade, philosophical commentaries have taken up the task of understanding the implications of Wittgenstein's *Philosophical Investigations* by way of commenting on Kripke's skeptical paradox which states that "[t]here can be no such thing as meaning anything."[1] Kripke argues for this paradox (which he neither supports himself, not attributes to Wittgenstein) by way of investigating rule-following. Using an example from elementary arithmetic Kripke shows that any actions that I pursue claiming to follow a rule can be re-interpreted as in fact following an infinite number of different rules; since all my rule-following is thus compatible with an infinite number of rules, there is no one particular rule that I am following. Since meanings, so Kripke, are rules determining the application of a term, the rule-following paradox entails that there is no meaning.

Wittgenstein never held the views that there is no meaning, or that one can follow rules any which way one pleases. Wittgenstein himself acknowledged that he had trouble putting down in writing what he wanted to convey, and his style is often cryptic, and some passages are apparently written for the *insider*. Other passages, and the

---

[1] Kripke (1984), p. 55. All references to this work will be abbreviated by *Kripke*, followed by the page number.

general drift of the *Philosophical Investigations* are, however, very lucid, which in turn seems to have provoked commentators to read all sorts of rather obscure philosophical theories into the *Philosophical Investigations* siding him with — or comparing him to — just to mention two of the more extreme examples, Derrida[2] and Davidson.[3] This is a saddening trend in the interpretation of a philosopher mostly critical of philosophical endeavors. But one should not think that Wittgenstein can be interpreted any which way.

As a matter of fact, the *Philosophical Investigations*, especially their first 250 sections, attempt to cure philosophers of the picture of language as a calculus whose essence is to refer to objects by means of words, a picture of language which Wittgenstein claims arises from common sense observations about language, and whose explicit formulation goes as far back as Leibnitz.[4] The method Wittgenstein chooses is not only quite different from, for example, Frege, or contemporary writers, but his view of analytic philosophy of language is also not very flattering; after all, he attempts to show that philosophical statements

---

[2] Cf. Wheeler (1986) "Indeterminacy of French Interpretation: Derrida and Davidson" in LePore (1986) pp. 477-495

[3] Cf. Cavell (1986) "Metaphor, Dreamwork and Irrationality" in *op. cit.*, p. 495-507.

[4] I show this connection in the first chapter.

(propositions?) are disguised nonsense.

> My aim is: to teach you to pass from a piece of disguised nonsense to something that is patent nonsense.[5]

In the *Tractatus*, Wittgenstein argued that mathematical propositions and propositions about philosophical logic have no sense. This position, and the remark quoted above (and there are plenty more in the *Philosophical Investigations*)[6] make the *Philosophical Investigations* probably hard to swallow for those who think that analytic philosophy of language has made progress evidenced by, for example, the fact that "Frege solved [the hitherto intractable problem of] generality"[7], an issue medieval philosophers abandoned as too complex. How can philosophy be nonsense? Especially philosophers in the analytic tradition are thus not too eager to embrace the charge that they have been puzzling over nonsense for millennia.

Arguably, however, the skeptic's statement that there is no fact of the matter as to what one means <u>is</u> patent nonsense. The paradoxical statement is ultimately derived from a combination of considerations that represent the core of traditional as well as contemporary analytic philosophy

---

[5] Wittgenstein, *Philosophical Investigations*, Section 464. All references to this work will be abbreviated by *PI*, followed by the section number.
[6] Cf. *PI* 119, 134, 246 etc.
[7] Dummett (1967) p.227.

of language. According to this view of language,[8] using a word is employing a referential expression according to certain rules. According to the skeptical line of argument, there are, however, no constraints on the interpretation of such rules governing the employment of the expressions; and precisely this consideration allows the skeptic to derive the proposition that there is no fact of the matter as to what someone means. Kripke's skeptic thus follows Wittgenstein's method to the letter: starting from the core assumptions of contemporary philosophy of language, the skeptic ends up drawing a conclusion that is patent nonsense. The very thesis, however, that an expression can be interpreted any which way is one that has enjoyed an enormous influence in this century. Quine's indeterminacy of translation, and Goodman's argument for his version of anti-realism use arguments that bear a strong similarity to the rule-following paradox which they historically preceded.[9] I need not point out that these philosophers have an outstanding reputation, and their following should be legion.

I think that philosophy of language, however, is not the only area in which the rule-following paradox may be relevant. I will focus on the impact of the rule-following

---

[8] The Model M, as I label this view, will be discussed in the first chapter.
[9] Cf. Kripke, p. 55 ff.

considerations not only in philosophy of language, but also in moral philosophy. In both areas, the rule-following paradox seems to indicate that following rules is no better than whimsical behavior. I think that this threat can be defused. I will try to show that the rule-following paradox derives its plausibility and force from a picture of language and morality as quasi-mechanical, rule-driven calculi. Once we abandon the view that language and morality are something like mechanical calculi, we can also abandon the view that rules need interpretation. Meaning and morality are thus not endangered by the rule-following paradox, merely certain theories of meaning and morality.

Ματεματα (and with it the idea of a calculus), 'the things that can be learned', have had a tremendous impact on philosophy. Since mathematical physics has replaced the telos-driven physics of Aristotle, philosophers have been convinced that mathematics is the interface between mind and matter, a language common to both. Objectivity seems to be the hoped-for byproduct of this model. This may be most apparent in Kant, who thought of the rules of the understanding as quasi-functions, and the applicability of formal geometric and arithmetic properties of space and time as a priori *constitutive* for anything that could be called nature, and therefore objectively known. The early 20th. century has both shifted and expanded this model of

knowledge and method. While Kant tried to establish a theory of knowledge by giving the a priori conditions for any judgment, 20th century philosophers replaced this epistemic theory with a theory of language that attempted to state the logical conditions for reference and meaning as elements of knowledge. Although the explanatory aims thus shifted from a theory of knowledge to one about language, the aim of giving a formal theory remains in place. Language is seen as a system of symbols that represent objects, and those symbols are combined into larger units by rules.

Much of Wittgenstein's later philosophy is intended to debunk this picture of language as a calculus for reference and the status of rules in philosophy. The *Philosophical Investigations* are designed to show that 'rule', 'meaning', 'language', 'understanding' etc. do not name essences, are not what Wittgenstein calls metalogical concepts.[10] His investigation into the 'nature' of language via, among other things, an investigation of our concept of what it is to follow a rule is intended to show that rules (of a calculus) are neither sufficient nor necessary for language. It also shows that there is no essence to rule-following a philosophical investigation could find. The insight that there is no essence to rules is thus directed against

---

[10] Cf. Wittgenstein, *Philosophical Grammar*, Section 8.

compositional theory of language[11] and Russell's program to reduce the laws of arithmetic to set theory,[12] and can also be used to present considerations against the views held by, for example, rule-deontologists and rule-utilitarians, who hold that rules and rule-following are the essence of morality.

I will first sketch a certain conception of language that is rather influential in contemporary philosophy of language. I think that this conception consists of a cluster of ideas, which are traditionally associated with Frege. This theory of language sees language as a type of calculus, a mechanical employment of words to refer to objects, an employment which can be explicated by purely formal and precise rules.[13]

In the second chapter I will outline a view in ethics that makes essential use of the concept of rules as well. Rule-utilitarianism, as well as Kant's rule-deontology seem to have a similar view of morality, which is seen as a body of rules that determines, in a prescriptive sense, which moral actions are right and therefore obligatory.

---

[11] Frege is often credited with having started this project. This issue will be dealt with in more detail in the first chapter.
[12] I won't explicitly treat the issue of logicism as a philosophical thesis.
[13] The ideas will be further developed in the first chapter.

In the third chapter I present the rule-following paradox which states that there is no fact of the matter as to which rule we are following, entailing that rule-following is impossible. The theories canvassed in chapters one and two thus seem to be doomed. Kripke's paradox has force against the view that language is a system of rules; in particular, I will discuss the impact the paradox has on versions of Platonism, Mentalism and Dispositionalism. The argument Kripke presents will then be expanded to make the paradox applicable to the kind of moral theories that see morality as a system of rules, that is, in very similar terms as the philosophers of language depict language as described in the first chapter. The overall argument in the chapter is that Kripke is right in suggesting that a certain view of language leads to a pseudo-paradox, namely, that there is no fact of the matter as to what someone means, or , *mutatis mutandis*, that there is no fact of the matter as to what is moral. The paradox is thus a *reductio ad absurdum* of the pictures leading to such a paradox, whether this be in the area of theories of meaning or morality.

The fourth chapter offers a Wittgensteinian critique of the rule-following paradox, showing that rule-following is a mundane activity, and that rule-following is not central either to language or morality. The argument of this

chapter will have to be detailed enough to show what's wrong with the rule-following paradox without thereby establishing a particular philosophical account of how and in what sense rules do or should determine our actions. This chapter will also explore certain issues and suggest certain consequences of the rule-following considerations for moral philosophy. The Wittgensteinan view of rules does not impugn the fact that rules may be acted upon as a matter of fact, and that they are extremely useful. I will not end the dissertation by suggesting yet another philosophical view, or by presenting yet another philosophical thesis. I think that there is nothing in general to be said about rules. What one cannot say anything about, one shouldn't.

**Reference, Rules, and Interpretation:**
**A Theme in the Philosophy of Language**

In the opening paragraph of his essay Truth and Meaning, Donald Davidson makes the following claim:

> It is conceded by most philosophers of language, and recently by some linguists, that a satisfactory theory of meaning must give an account of how meanings of sentences depend upon the meaning of words. Unless such an account could be supplied for a particular language, it is argued, there would be no explaining the fact that we can learn the language: no explaining the fact that, on mastering a finite vocabulary and a finitely stated set of rules, we are prepared to produce and to understand any of a potential infinitude of sentences. I do not dispute these vague claims ... . (Davidson (1967), p. 17.)

Calling these "vague claims" concessions may at first be surprising; concessions - in philosophy as elsewhere - are made after arguments or facts force a person to accept a view which is significantly different from the one she previously held. What, however, one might ask, can be more obvious than the fact that the meaning of a sentence depends on the meaning of the words it contains, and the fact that there are rules that express how the meaning of the sentence depends on the meaning of the words? Davidson's opening remarks apparently claim that philosophers previously rejected, and recently started conceding, some rather obvious and trivial truths about language, and this is not an altogether trivial opening line for this rather

influential essay in the philosophy of language.

Philosophers of language like Davidson very often begin their theorizing with apparently trivially true observations. These observations seem so trivially true that there is apparently no need to investigate the assumptions underlying those claims. But they are intended to motivate a certain framework for theories of language. Wittgenstein, in the *Philosophical Investigations*, has indicated that such apparently platitudinal observations can set the stage for a philosophical view or theory, and suggests that one can dissolve the perennial problems apparently endemic to philosophy and philosophers only if the initial observations are very carefully examined. Thus, in Section 308 of the *Philosophical Investigations*, Wittgenstein points out a 'conjuring trick' by which the philosopher describes a phenomenon. The way the phenomenon has been pre-theoretically described forces us to look for an explanation, which turn out to be pseudo-explanations and pseudo-answers. When those fail we do not carefully investigate the pre-theoretical description of the 'phenomenon', but seem to be driven to deny that the phenomenon exists at all:

> How does the philosophical problem about mental processes and states and about behaviourism arise? — The first step is the one that altogether escapes notice. We talk of processes and states and leave their nature undecided.

> Sometimes perhaps we shall know more about them — we think. But that is just what commits us to a particular way of looking at the matter. For we have a definite concept of what it means to learn to know a process better. (The decisive movement in the conjuring trick has been made, and it was the very one that we thought quite innocent.) — And now the analogy which was to make us understand our thoughts falls to pieces. So we have to deny the yet uncomprehended process in the yet unexplored medium. And now it looks as if we had denied mental processes. And naturally, we don't want to deny them.[14]

Similarly, the *Philosophical Investigations* open by quoting an autobiographical passage by St. Augustine describing his upbringing and, apparently unwittingly, reveals a "picture of the essence of human language" which Wittgenstein claims is at the roots of a philosophy of language:

> Every word has a meaning. This meaning is correlated with the word. It is the object for which the word stands.[15]

Words are thus considered to be names of objects, and rules of composition combine them into sentences.

In what follows I will try to argue that the Wittgensteinian suspicions regarding method and substance of analytical philosophy of language are well-taken, since I will argue that contemporary and traditional analytic philosophy of language share a picture of language, a

---

[14] *PI*, 308.
[15] The quoted passages are from *PI*, Section One.

picture motivated[16] by subtly changing ordinary and pedestrian observations about language. By placing explanatory weight on ordinary words like *meaning* and *understanding*, and by giving a certain interpretation to certain ordinary notions like *rule* and *language*, philosophers perform the 'conjuring trick' that allows them to motivate the belief in a need for general explanatory concepts in a theory of language. I will try to show that a number of theorists share a picture of language, a picture I will call the Mechanical Model of Meaning (Model M). According to Model M, *meaning* and *rules* are the central explanatory concepts of the theory, and essential properties of language. The fundamental semantic element is the word (or some basic expression) whose meaning is an object. Language is comprised of words and a system of rules for their combination. A theory of language, accordingly, starts by giving a theory of meaning for the basic semantic elements. The paradigm for meaning is reference, an interpretation of a symbol by the object it refers to. Successful reference to an object requires using a term in accordance with this referential relation that exists

---

[16] I will use this term in order to avoid speaking of 'arguing' or 'reasoning'. Since I contend that philosophers turn every day observations into philosophical pictures, I think that 'arguing' is not quite correct. I thus speak of 'motivating' in order to retain the compelling element of the picture, without giving it the credit of being justified by 'arguments' in the classical sense of the term.

between symbol and object. The relation between the object (the interpretans) with the word (the interpretatum) is established by means of a mapping relation, which is a type of rule assigning objects to symbols. Furthermore, the combination of simple symbols into complex symbols is also a rule-driven mechanism, in which rules determine the permissible combinations of the simple symbols. Consequently, according to Model M, understanding language is a process of following the rules of this calculus. If this is correct, then Wittgenstein's first section of the *Investigations* is thus essentially a correct analysis of contemporary analytic philosophy of language.

To show that Model M is the underlying picture for much of current analytic philosophy of language, I present a couple of contemporary philosophers of language who, by their own lights and those of their colleagues, hold very different theories of language. I will not describe the individual theories in too much detail; instead I will try to show that regardless of their differences the philosophers in question all reason for their theoretical framework by starting with the type of trivial observations Davidson makes in the quotation above. Starting from those apparently trivial observations, they quickly move on to theory-laden statements regarding universal properties of language, properties which particular theories then purport

to explain. This manner in which philosophers go back and forth from the apparently trivial initial observation and the first rung of the ladder leading to a philosophy of language is a philosophical maneuver I will try to reveal.

I have labeled the picture of language Davidson sets out to motivate in the quote above the "Mechanical Model of Meaning". Following Davidson's outline of the "vague claims", Model M has the following elements:

**The Mechanical Model of Meaning (Model M)**

1. Natural language has an essential, constitutive property, meaning.

2. Meaning is best represented in terms of a system of rules mechanically correlating symbols directly to objects ('reference').

3. A system of rules of composition recursively[17] generate (potentially infinite)[18] semantic complexes from

---

[17] What is meant here is the view that a theory of language develops a model that orients itself toward Recursive Function Theory, or the Theory of Algorithms. This is the notion of a 'mechanically calculable function'. For the importance of this topic for the theory of language, see Kripke (1984), Fodor (1990). For the notion of recursion, see Davis (1967).

[18] For the most part, I will ignore this element in the first chapter. It is, however, of pivotal importance to the argumentative situation contemporary dispositionalism faces (causal and informational theories of language as defended by, among others, Fodor (1987,1990), Dretsky (1981) and Millikan (1986)). The dispositionalist face a version of Descartes' argument that we do not know the wax through the faculty of imagination: since the imagination is finite, but the wax can take up infinitely many forms, the imagination cannot be the source of the idea of wax.

(continued...)

semantic simples.

4.
Understanding is a mental process best represented in terms of the referential and compositional structure of language.

I will start the discussion with the first three theses which together make up the core of the view of language as a calculus. The fourth thesis pictures understanding a language as operating this calculus; this issue will be discussed in the second part of this chapter.

---

[18](...continued)
Similarly, since language is assumed to have potentially infinite power of expression (see Davidson's quote above on page 11) a finite mind cannot be the source of meaning. This is the key argument Kripke uses to reject all but a 'skeptical solution' to the paradox. The issue will be discussed in detail in chapter 3.

## The Mechanical Model of Meaning (I): Meaning, Interpretation, and Language as a Calculus

One of the claims Davidson makes is that the meaning of words is essential to a satisfactory theory of meaning, since the meaning of words is taken to be the essence of language. The terms, — *word* and *meaning* — seem familiar enough; in an essay published a few years after the one from which the above quotation was taken, Davidson apparently makes the same point using a different terminology:

> An acceptable theory should ... account for the meaning ... of every sentence by analyzing it as composed ... of elements drawn from a finite stock.[19]

The elements of a meaning theory mentioned here (as in the first quote) seem to be the meanings of words as they are listed in a lexicon or dictionary.[20] In fact, dictionaries are used for precisely that purpose; a person consults a dictionary to help her understand the meaning of a word in order to understand the meaning of a sentence. For example, the sentence "brambles rubbed against the car

---

[19] Davidson (1970) p.56.

[20] Lexicons are, I take it, not part of any theory of language. They are simply useful lists explaining the meaning of terms. But the term is also used, interestingly enough, to name the list of symbols in formal languages. This is further discussed on page 38 *infra*.

doors" will not be understood unless the meanings of the words the sentence contains is known, and the meanings of those words are listed in a dictionary, which is, after all, a finitely long list. It seems thus perfectly natural to think that words are the elements of meaning, and that the meaning of a sentence is determined by the meanings of the words it contains. Again, one may wonder how anyone could have disagreed with this account; one may especially be surprised that the majority of linguists could have been inveterate enough to believe otherwise. Surely they must be convinced by now?

Davidson is apparently not alone in thinking that these trivial observations motivate a theory of language. Jerry Fodor, certainly in many respects opposed to Davidson's truth-theoretic holism, agrees with Davidson's view practically *verbatim* (*ceteris paribus*):

> The question [concerns how the content of a] ... complex expression ... is determined ... by the content of their ... parts.[21]

---

[21] Fodor (1987) p.138. This argument is intended to be universal in scope. Fodor writes:

> The classical argument that mental states [1] are complex adverts to the productivity of the attitudes [2]. There is a (potentially) infinite set of - for example - belief state [3] types, each with its distinctive intentional object [4] and its distinctive causal role [5]. This is immediately explicable on the assumption that belief states [6] have combinatorial
>
> (continued...)

Much in the line of trying to convert others to accept the obvious lies in the second part of the claim alluded to by Davidson, that language has a finite set of rules of meaning composition. The rules of composition are used to generate complex units of meaning — sentences — from the elements of the sentence. 'Rules of composition', again, seems a rather ordinary notion. The claim about rules of meaning composition seems to be the same claim as the one about parts. If a sentence has parts that have meaning, then there must be some way in which they are put together to make the whole. We don't just throw words together when we formulate a sentence. Thus it seems trivial to acknowledge that we have to learn both the words of a language as well as the rules by which they are put together. Although one may be vaguely aware of the fact that we usually learn rules of grammar, not of meaning composition, it may be accepted as somehow true that there have to be some rules that are responsible for complex meanings. "Planes fly" may be either geometrical nonsense, or considered a piece of aeronautical information, depending

---

[21](...continued)

[7] structure. (Numbers in brackets added; see below.)

[1] could stand for sentences; [2] for "that-clauses"; [3] for "belief-sentences"; [4] for "object", etc. The point of the argument is thus entirely general and argues for combinatorial rules and semantic/psychological simples. Fodor, in fact, believes in a Language of Thought, thus more or less identifying the principles of mind with those of language.

on what meaning we assign to the noun. In this fashion, we might think, rules are employed ensuring that the meaning of the sentence depends on the meaning of the parts, just like Davidson reports philosophers now concede. We might even start thinking of formulating such rules, such as "never use a verb describing a physical activity to predicate something of an abstract noun", which would guarantee that above sentence about planes would get the proper interpretation. Davidson does not tell us which rules control the semantic composition of a sentence, nor are we told which ones have been selected by philosophers. This is the reason Davidson refers to the shared claims in the philosophy of language as 'vague'. But it seems not altogether farfetched to suppose that there are such rules, and the claim itself sounds just as trivial as the one that states that the meaning of a sentence depends on the meaning of its parts.

These two types of claims regarding the elements and mechanics of meaning are apparently specific enough to identify a particular *program* in the *philosophy* of language, at least to those who pursue such a program or follow its spirit. Stephen Schiffer, in his *Remnants of Meaning*, reviews the status of this program two decades after Davidson published his influential essay. Schiffer, who devoted much of his book to showing why Davidson's particular theory (an empiricist, truth-theoretically

oriented, sententialist semantic holism) is all wrong,[22] cites Davidson as one of the many proponents of a widely held set of theses in the philosophy of language, viz. that

> *every natural language has a compositional meaning theory* — a compositional meaning theory for a language L being a finitely statable theory of L which specifies the meaning of all the primitive vocabulary in L and specifies compositional rules showing how these meanings determine the meaning of the infinitely many complex expressions in L. [23]

I selected this passage to exhibit a certain philosophical maneuver. Davidson's quote states that philosophers agree about some trivial observations about language (and so they should), and thinks that those observations can be cited as supporting Model M (which is a different story). With the slight of hand of a trained philosopher, the trivial observations are translated into a general theoretical statement that describes a widely accepted programmatic perspective in the philosophy of language. Schiffer, who (like many others) quotes the opening passage of *Truth and Meaning* to support his thesis that Davidson supports Model M, thinks that no additional arguments are required to change the trivial observations

---

[22] Schiffer's *Remnants* argues that Davidson's theory cannot give a compositional account of propositional attitude sentences. But then, according to Schiffer, no theory can.
[23] Schiffer (1987), p. 2.

about language into something different.[24] Schiffer's formulation no longer clothes the claims in the pedestrian terminology Davidson or Fodor prefer, but it still maintains that the pedestrian observation Davidson makes lend support to the compositional meaning theory, and Davidson and Fodor concur, as their respective writings confirm.

Schiffer's statement, if compared to the trivial observations in Davidson's quotation, is characterized by terminological allusions to model theory and formal semantics, and apparently nothing in Davidson's statement is even remotely an allusion to such theoretical representations. Stating that philosophers agree that an explanation is needed for how the meanings of sentences depend on the meanings of words they contain is a far cry from talking about compositional meaning theory in Schiffer's sense. This 'agreement' about a program, then, is the bridge across the big theoretical chasms that separate Schiffer, Davidson, Fodor, and others. Schiffer,

---

[24] Notice, for example, that the term 'depend' is not at all specified in the initial observation; the notion of "finitely statable theory of L which specifies the meaning of all the primitive vocabulary..." makes a specific claim as to the type of dependence that is acceptable. The problem, from my perspective, is that no one can even give a hint as to what the theorems of such a theory of meaning would look like. Davidson was the first who made a proposal as to what the <u>form</u> of the theory should be, and this <u>form</u> gives an idea as to what <u>type</u> of dependence he was looking for.

for example, is not a holist like Davidson[25], nor does he believe, as Fodor does, in intentional realism, not to mention Fodor's Language of Thought Hypothesis.[26] All three, however, agree that the innocent remark — the meaning of a sentence depends on the meaning of words — motivates a program, can be used as a starting point in a philosophy of language.

How does Model M shift the trivial observation that the meaning of a sentence depends on its component words into a theoretical framework? Model M supposes that natural language is essentially characterized by meaning, and also a certain mechanism, a calculus, which together are the theoretical entities needed for a theory of language. Instead of seeing *meaning* and *understanding* as one of the many other terms that occur in language like *table* and

---

[25] "[T]he holistic method ... starts with the complex (sentences, at any rate) and abstracts out the parts." Davidson (1977), p. 221

[26] Intentional Realism is the thesis that
> there are mental states that have associated intentional objects, (for example, the state of having a symbol that means 'I raise my left hand' in my intention box) and that these mental states that have associated intentional states also have causal roles (for example, my being in one of these states causes my left hand to rise). What makes the story a Language of Thought story ... is the idea that these mental states that have content also have syntactic structure — constituent structure in particular — that's appropriate to the content they have. (Fodor (1987), p. 136f.

*door*,[27] *meaning* is taken to be the essential property of language that needs to be explained to accomplish a theoretical understanding of language, while understanding is the mental process that processes language by means of a calculus and by correlating symbols with objects. The notion of 'meaning' thus gets a particular treatment due to the fact that the basic bearer of meaning — according to Model M as presented by Davidson, Schiffer, and Fodor — are words, or the 'primitive vocabulary' Schiffer talks about.[28]. This is due to the fact that language is seen as a calculus that forms the meaning of a sentence from the meaning of words. Meaning, together with rules, are thus the properties of language that need to be explained. Once it is decided that words are the primary bearers of meaning, it is natural to assign words name-like properties. Kripke, in his book on Wittgenstein,[29] introduces a new version of meaning skepticism, which he attributes to Wittgenstein. Using an example from elementary arithmetic Kripke shows that any actions that I pursue claiming to follow a rule can

---

[27] See *PI*, 97.
[28] Philosophers in the analytic tradition, of course, realize that there are different type of words, like quantifiers, pronouns, names etc. But the general drift is to show that those words function very similarly in that they refer to some object or other, be that object concrete or abstract. Grice (1957) and Austin (1961) seem to be an exception, but both try to reduce meanings to intentions. This reductive model has the same scheme in that the hearer is to associate the word with a meaning act of the speaker. Meaning is thus a relation between a word and the speaker's intention.
[29] Kripke (1982).

be re-interpreted as in fact following an infinite number of different rules; since all my rule-following is thus compatible with an infinite number of rules, there is no one rule that I am following. Since meanings, so Kripke, are rules determining the application of a term, the rule-following paradox entails that there is no meaning. I suggest that Kripke's skeptic can be most naturally introduced in connection with a theory of meaning that is designed as a theory of word reference. This theory of reference tries to explain meaning as a relation between a symbol and an object.

The idea of language as a calculus goes back at least to Leibnitz' idea that there should be an ideal language (*characteristica universalis*) in which the basic signs would represent simple and not further analyzable concepts, and all complex signs would be defined in terms of the simple ones.[30] In the 19th century, Boole went a step further and wrote that an underlying calculus is actually in place:

> [T]he immediate subject of examination is language, with the rules which govern its use; while in making the internal processes of thought the direct object of our inquiry, we appeal in a more immediate way to our personal consciousness, - it will be found that in both cases the results are formally equivalent. Nor could we easily conceive, that the unnumbered tongues and dialects of the earth (could not)

---

[30] Leibnitz (1951), p. 10 (cited in Baker, G. and Hacker, P.M.S. (1984)) p.22

> have preserved (...) so much that is common and universal, were we not assured of the existence of some deep foundation of their agreement in the laws of the mind itself.[31]

It was only in the late 19th century that Frege tried to present a theory as to how the general idea of logic, — and hence rational language use — as a formal system in our contemporary sense should be fleshed out. Contemporary philosophers thus see Frege as having broken the ground for modern semantics, and he is very often interpreted as being primarily a philosopher of language.[32] The idea of language as some form of calculus seems to have been accepted well before Frege showed what a theory of such a calculus would look like. In this sense, contemporary philosophers see themselves as following a program for which Frege gave the general directions, but the idea of language as calculus is in fact older. The exact content of the notions of meaning and compositionality are of course, once they form part of a theory of language, dependent on the theory. But the trouble is, that there is no theory available.[33] All that

---

[31] Boole (1854) ch.2, § 1 (cited in Baker, G. and Hacker, P.M.S. (1984)) p.25

[32] This may not be accurate. Ricket (1986) argues that Frege's talk of meaning need not be interpreted as invoking a semantical concept. Frege's chief motivation was his interest in the foundations of mathematics, for which he had to clear away the logical imperfections of language, for which purpose he developed the concept script, thus moving from mathematics to logic. Cf. Frege (1919), p 253f.

[33] Here are two samples of how experts in the field see the situation: "There is amazingly little agreement, even about the basics. It is often not clear what problems a theory is

(continued...)

philosophers have is Model M, and that is not a theory.[34]

Regardless of the fact that philosophers seem not able to move from Model M to a theory of language, they nevertheless credit Frege with what are claimed to be theoretical insights accepted in Model M. Michael Devitt and Kim Sterelny, for example, claim in *Language and Reality* that the

> elements of sentences are words. What is it about a word that affects the truth conditions of sentences containing it? What part of its meaning plays the role? The obvious answer is its *reference*. ... The sentence's property of being true if and only if (such and such is the case) is seen to depend, in part, on two distinct referential relations holding between parts of the sentence and objects: *designation*, holding between the name [...] and an object [...]; and *application*, holding between the predicate ... and many objects.
>
> [This] approach to meaning ... was started in the nineteenth century by the logician, Gottlob Frege. ...[35]

The sketch leaves out much of Frege's theory by

---

[33](...continued)
trying to solve. It is often not clear whether theories are in competition and, if not, how they relate to each other." (Devitt and Sterelny, (1987) p. ix). "[I]t is time for me to say what [theories of language] are true. None." Schiffer (1987) p. 265. I might add that "theory" is used merely as a metaphor by the authors I quoted: there is not a single axiom that is even guessed at in any book of contemporary philosophy of language.

[34] I understand theory here in a very wide sense. A theory in this sense is a set of sentences that allows us to derive, from basic axioms, laws that allow us to explain phenomena. Model M has no such axioms, and no such laws.

[35] Devitt and Sterelny (1990) p. 18.

emphasizing that a theory of meaning is first and foremost a theory of reference. But the authors are not in any case pursuing some exegetical issue. Crucial parts of Frege's theory are therefore simply dropped from their account of Frege's importance. Frege's notion of coloring, and his notion of sense are not even mentioned.[36] Nevertheless, the claim presented by Devitt and Sterelny is two-fold: Frege started a project, and contemporary philosophy of language continues that project. Another contemporary philosopher, while discussing Frege's theory, again stresses primarily the notion of reference, even though he briefly introduces

---

[36] There is, of course, a huge literature on Frege's notion of sense. It cannot be the task of this dissertation to give an account of Frege's theory. It may suffice at this point to cite Frege who must have thought that the gloss that follows explains what he meant by 'sense':

> It is natural, now, to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the reference of the sign, also what I should like to call the *sense* of the sign, wherein the mode of presentation is contained. (Frege (1892) p.24.)

Notice that he, too, thinks that the notion of sense is something that is 'natural', i.e. he also uses a pedestrian motivation for Model M. The sense of an expression is that semantic property that determines its reference. Frege introduces the notion of sense in order to follow through with his logicist program. Part of this program is the consideration that the word 'true' specifies the goal of logic. 'True' is applied to thoughts, which contain names which name objects. The notion of sense allows Frege to show that a thought can have constituents, since the meaning of moon is the thing Armstrong stepped on, and this cannot be part of a thought. The distinction between sense and reference also allows him to clarify that the notion of concept does not apply to the sense of a predicate, but to the entity named by the predicate. (see Frege (1919) p. 255)

the notion of *sense*:

> On Frege's theory, any meaningful expression, whether a sentence component or a complete sentence, semantically refers to (designates, stands for) something, if anything, as its referent (Bedeutung), but it only does so by semantically expressing something else: its sense (Sinn). The sense of an expression is a purely conceptual representation, and the referent of the expression is whatever uniquely fits the representation.[37]

In this account, Frege's distinction between concepts and names is not found important enough to mention.[38] Furthermore, Frege's theory may not even be a theory of language in the customary sense, for Frege rejected the idea that there can be a coherent theory of natural language. Many sentences in natural language have no truth value,[39] and although, according to Frege, sense, the mode of representation of the object, determines reference, not everybody associates the same sense with the same sign.[40]

---

[37] Salmon (1986) p. 47.
[38] Frege's account of concepts is that they express functions. He also draws a sharp distinction between concepts and objects. A concept is predicative, that is, it is the meaning of a grammatical predicate. A proper name, according to Frege, is incapable of being used as a grammatical predicate. cf. Frege (1892a) p. 43.
[39] All sentences that contain names that have no reference have (according to Frege) no truth value.
[40] For some general hint as to Frege's misgivings about the sense of ordinary proper names, see his *On Sense and Reference loc.cit.*, p. 24, esp. footnote to p. 24. Frege remarks that ordinary proper names fail to have the *sense* required for a scientific notation. Furthermore, according to Frege, a sentence containing (what he termed) a proper name (definite description) which lacked reference also lacked truth value. This deficiency of natural language
(continued...)

Frege's notion of 'meaning' is thus very different from what either Salmon, or Devitt and Sterelny take meaning to be. Frege is setting out a rigorous, scientific presentation of logic, i.e. the general laws of rational language use. The notion of meaning as understood by authors like Davidson develops in his work, which is extensionalist[41] and therefore quite different from Frege's, which is intensionalist.[42] The differences between the semantic theories of Frege, Salmon, Devitt and Sterelny, Davidson, Fodor, and others are quite significant. But those philosophers seem to be in agreement about the fact that symbols of whatever complex structure stand for something, and that symbols are concatenated by rules. None of the authors seems to rely on what is meant with the pedestrian

---

[40](...continued)
(that it contains myriad sentences that have no truth value) he considered one of the reasons why we cannot give a coherent account of natural language.

> [T]he *Bedeutung* of a singular term is identified with the object for which it stands ... [which] leads Frege to regard a sentence containing a proper name that lacks a bearer devoid of truth value ... . Dummett (1981) p. 161 (cf. also pp. 31, 32, 102).

See also Dummett (1967) and Resnik (1981) p.83-103, esp. p. 99 ff.

[41] See below for a more detailed account of Davidson's program.

[42] An intensionalist compositional semantics aims at theorems which correlate the sentence of the language with an intension, that is, an abstract meaning that determines a truth value for the sentence. These entities would be the "contents" of the sentences mapped onto them.

view that the meaning of a sentence depends on the meaning of its component words. Language is thus held to have a semantic property, namely meaning, and meaning is construed as the reference of a word or symbol to an object of one type or another. Reference becomes a relation between a symbol and an object, and language as a whole is compositional, *viz.* language is a system of rules determining the combination of such symbols. I think it is worth a philosophical investigation as to whether or not such a program is viable for a theory of natural language.

There is a further type of semantic theory that is a strong supporter of Model M, and it may not be implausible to believe that Davidson's paper has been given so much attention because it gathers indirect support and credibility from this type of theory: model theory. Model theory may well be characterized as the purest form of the type of theory that I claim is the core of the Model M. Model theory is semantics *more geometrico*.[43] Here is how Davidson's project is characterized by a fellow philosopher:

> Donald Davidson is the creator of two justly famous theories. The first is about the form

---

[43] The following few paragraphs may presuppose a familiarity with contemporary semantics. The task of these paragraphs is to 1) make a connection between model theory and the Davidsonian program and 2) emphasize the view that a semantic theory makes essential use of the notion of rules, and 3) maintain that this all is the same picture of language which I dubbed Model M.

> that a compositional semantics, or meaning theory, for a particular language should take; and the second ... is [a sententialist theory] about the logical form of propositional-attitude ascriptions. ... *Sententialist* theories of propositional-attitude go hand in hand with extensionalist accounts of compositional semantics.[44]

Davidson is an extensionalist.[45] A compositional semantics for a language is extensionalist if its compositional component is a finitely axiomatized truth theory for L, where L is simply the whole of natural language. For a truth theory, Davidson relies on the work of A. Tarski. Tarski states that

> Semantics is a discipline which, speaking loosely, deals with certain relations between expressions of a language and the objects ... 'referred to' by those expressions.[46]

The heart of Tarski's views is the definition of truth

---

[44] Schiffer (1987) p. 111.

[45] Davidson abandons the notion of reference altogether:

> The theory will have done its work if it provides, for every sentence *s* [in the scheme *s means that p*] ... a matching sentence p ... . [L]et us try treating the position occupied by 'p' extensionally ... (Davidson (1967) p.5).

But he merely abandons it for the notion of 'satisfaction', preserving the basic semantic relation between an expression and an object:

> The theory gives up reference ... as part of the cost of going empirical. It can't, however, be said to have given up ontology. For the theory relates each singular term to some object or other, and it tells us what entities satisfy each predicate. (Davidson Reality without Reference in Davidson (1984) p. 223)

[46] Tarski (1944), p. 345.

in terms of satisfaction.[47] Semantics in this model sees expressions as referring expressions. Words, as referring elements are component of expressions that themselves refer to, in Davidson's modified program, the absolute truth.[48] Davidson thus can rely on the rigorized theory of Tarski behind the general idea of expressions referring to objects, or being satisfied by objects, when he introduces his particular theory of (natural) language. Even though Davidson ultimately abandons the notion of reference,[49] the theory retains the character of a relation between two entities, one linguistic, and one non-linguistic, and the expressions are concatenated by rules. The theory is still considered to be motivated by the "vague claim" that "a satisfactory theory of meaning must give an account of how meanings of sentences depend upon the meaning of words".

It may thus be useful to highlight some features of non-natural languages and model theory. A formal language is identical with the set of its well-formed formulae; a well-formed formula is simply a formula that conforms to the principles of the language as determined by some arbitrary set of symbols used in the language, and the recursive set of rules used to generate formulae from the set of symbols,

---

[47] For a readable account of Tarski's program see Putnam (1978), Lecture I.
[48] See page 224, *op. cit.*
[49] He replaces it with the notion of satisfaction. See footnote 45.

which have no intrinsic semantic properties, do not refer. The only properties they have are syntactic — they can be recognized by their shape. The set of well-formed formulas of a particular formal language is thus completely determined by the arbitrary settings of those elements:

> The set of well-formed formulas of a particular formal language is determined by a fiat of its creator, who simply lays down what things are to be (well-formed formulas) of his language. Usually one does this by specifying
>
> a set of symbols, (the alphabet of his language)
>
> and
>
> a set of formation rules determining which sequences of symbols from his alphabet are (well-formed formulas) of his language.[50]

The set of rules and symbols of formal languages must be defined without any reference to an interpretation. By contrast, natural language is, among other things, a means of communication, and a theory of language must account for the fact that symbols have meaning. Words, this is the main tenet of Model M, have meaning, refer, and can be true of something. The philosophical question of how noises, sounds, and marks can have meaning, is, of course, a matter of dispute. Different kinds of explanations for meaning are admitted by some, and rejected by others. The philosophers subscribing to Model M agree that language has an essential property, meaning, and that this property can be explained by a theory of reference. Yet applying the notion of a

---

[50] Hunter (1971) p. 4.

formal system in an account of natural languages cannot take for granted that it will do justice to this element of reference and meaning in natural languages.

Philosophers have not neglected this issue. Formal languages may be interpreted by developing a model, or interpretation, for such languages:

> In rough and very general terms, an interpretation of a formal language is an assignment of meaning to its symbols and/or formulas. Model theory is the theory of interpretation of formal languages (a model of a formula of a language is an interpretation of the language for which the formula comes out true).[51]

Once we have constructed the symbolism, that is, the 'language' in its syntactic aspects, we describe how this symbolism is to be interpreted by giving a model for this symbolism. The interpretation of a set *A* of well-formed formulas consists of a so-called domain, and a function which assigns symbols some element in that domain. Just as with the formal system itself, which is arbitrary, the interpretation of a formal system is an arbitrary assignment of objects to symbols. A formal system can be given any interpretation, so long as the domain is not the empty set. There are infinitely many functions that can assign an interpretation, and thus a model, to a system. A model of a formal language has thus the following characteristics: a

---

[51] *Op. cit.*, p. 6.

set of functions maps the formulas of the language onto a domain of objects. This assignment of meaning is called an 'interpretation', and the formulas as well as the elements of the language have a name-relation to the objects in the model.

The notion of *meaning*, in its ordinary usage, may actually support the idea that meaning is to be understood in relational terms; the picture here is that there are two elements, the symbol and a further element which assigns something to that symbol, and makes it meaningful. The term *meaning* seems to indicate that there is a further, underlying element to a communication, even after the words in their literal sense have been understood; *meaning*, in the way we commonly understand the term, is linked to interpretation, something beyond the literal use of a term. Interpretation (this is even etymologically the case) suggests that two different entities have to be connected. Philosophy of language seems eager to apply this picture in a theory of natural language. *Meaning* thus becomes an object of a certain type (physical, mental or abstract) which interprets a symbol. A second entity is needed to give a symbol, a word, or sentence, *meaning*. The meaning of 'meaning' thus comes to mean reference to this object, function or structure. Model theory has a similar character: two elements, the symbol and the element of the

model are associated by mapping one element onto the other, which is understood to be a function. This function-theoretic understanding may also be seen in terms of rules. There is a rule that uniquely assigns a value to a symbol, and whenever the symbol or appropriate group of symbols occurs in the language, a rule may be seen to assign the appropriate value. The character of meaning as understood in semantics and the notion of a model for a symbol share thus a structure — allowing that theory of language gets its theoretic inspirations from model theory — which in turn explains the reason that philosophers quickly move from trivial observations about language to a certain theoretical understanding of language.

The above considerations, I believe, show that there is a strong connection between Frege, or at least what contemporary philosophers attribute to Frege[52], contemporary semantics, and model theory. The core element of this connection is what I called the Mechanical Model of Meaning - a picture which starts with certain trivial observations about language. Model M holds that language contains

---

[52] The underlying thesis for this chapter is, of course, that Wittgenstein attributed Model M to Frege as well. This dissertation cannot answer the question whether Frege is really a philosopher of language, or whether he has been misunderstood by contemporary philosophers of language. Given the notorious misunderstandings and misinterpretations in the history of philosophy, this is not at all unlikely. In this case I restrict myself to claiming that Frege is currently held to support Model M.

sentences which contain symbols that refer to objects. The objects they refer to are what gives those words an interpretation; assigning the symbol an object, and combining the symbols to form complex expressions is accomplished by fixed, determinate, and mechanically applicable rules. This view is shared by model theory which maps symbols onto a domain which is the reference, the interpretation, or the value of those symbols. Frege, in the contemporary interpretation, is credited with inventing semantic theory: he is interpreted as holding that meaningful words and sentences always determinably refer to something. According to the commentators who have been examined above, Frege holds that words in combination make up sentences and those sentences refer to other objects, *viz.* thoughts or truth values. By starting with the rather pedestrian formulation Davidson chose for the opening line of his essay *Truth and Meaning*, I intended to show that this innocent view about language quickly evolves into a theory-driven account of the nature of language.[53] Wittgenstein, for example, whom I tried to follow in spirit in this section, finds this picture expressed in an autobiographical remark by St. Augustine, quoted in the first section of the

---

[53] It is perhaps important to mention again that there really is no theory of language available. No-one knows a single theorem of this theory. On the other hand, the theoretical apparatus from model theory and formal languages, and the idea about what it is that is to be explained (reference) are firmly in place.

*Philosophical Investigations*. For reasons I have already discussed, Wittgenstein could as well be taken to comment on Davidson:

> These words, it seems to me, give us a particular picture of the essence of human language. It is this: the individual words in language name objects -- sentences are combinations of such names. --- In this picture of language we find the roots of the following idea: Every word has a meaning. This meaning is correlated with the word. It is the object for which the words stands.[54]

## The Mechanical Model of Meaning (II): Understanding

This philosophical picture of language as a means for referring to objects, lingering on uncritically despite the *Philosophical Investigations*, has a complimentary picture about how we employ this language, a view expressed in the second part of Davidson's quote. The idea Davidson alludes to is that language has to have a certain structure, because it must be understood by human beings. As a consequence, so its supporters argue, language understanding has to be of a certain type, viz. language production and language understanding must conform to the structural principles of the Model M view of language. In line with the theoretical

---

[54] *PI* 1.

moves mentioned above, contemporary philosophy of language supposes that understanding language is a mechanical process. Although many contemporary philosophers do not specify what the mental apparatus is like, and vague metaphors like "intention box" are common coinage,[55] there is an underlying basic commitment to some assumptions about the psychological processes underlying language understanding and production. Davidson makes explicit references to such psychological processes, which serve the purpose of helping narrow down the choices for the model that serves best as a theory of natural language.[56] Most prominently, M. Dummett has argued that "a theory of meaning is a theory of *understanding*."[57] Although this is not a universally accepted approach to a theory of meaning, a good many philosophers share the belief that a theory of mind and understanding is pivotal for a theory of meaning.[58] Philosophers are quite adamant that they are not making an empirical claim about the nature of the mind; instead, they consider it merely a theoretical necessity that the mind has a certain role for understanding and processing language.[59]

---

[55] Cf. e.g. Fodor (1987) p. 136.
[56] See the opening quote of this chapter, where he makes a connection between the notion of understanding and the need for compositional semantics. See also Footnote 57.
[57] Dummett (1973) p.92. His student J. McDowell defends a similar view in McDowell (1993), p. 111-136.
[58] See, among others: Foster, J. (1976) and Platts (1979).
[59] This is not the case for Chomsky, who makes metaphysical claims:

(continued...)

David Lewis, for example, writes:

> I distinguish two topics: first, the description of possible languages or grammars of abstract semantic systems whereby symbols are associated with aspects of the world; and second, the description of the psychological and sociological facts whereby a particular one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics.[60]

There are (mathematically speaking) infinitely many formal systems and models, and extending such systems to theories of natural language must address the problem that one of the many systems has to be selected. But a philosopher should not assume without argument what kind of system natural language is. Davidson and Lewis make what they consider untendentious assumptions by suggesting that the fact that natural language has to be learnable by human beings with finite psychological abilities requires that the process of understanding has to be of a certain type. Davidson, in the familiar quotation, justifies this

---

[59](...continued)

> [W]e may regard the language capacity virtually as we would a physical organ of the body and can investigate the principles of its organization, functioning, and development in the individual and the species. Personally, I feel that this is just the right way to approach the study of human language. (Chomsky (1980) p. 185).

Chomsky simply conflates different philosophical categories in this passage (a physical thing (organ) with a capacity). This is the reason I won't discuss him much in this connection.

[60] Lewis (1972) p. 170.

assumption by arguing that language could not be learned unless it had certain features, presumably because the mental processing is of a particular nature. Lewis, on the other hand, makes the assumption without giving an explanation at all. The view seems to him somehow self-evident:

> [W]e have a lexicon wherein finitely many expressions - word or word-like morphemes - are assigned to categories.[61]

Although the application of model theory in Model M holds fast to the assumption that natural language is somehow essentially referential, formal systems need not be finite nor need they be compositional; a formal language may consist of merely an infinitely long formula, or a single symbol. As far as model theory goes, a language can consist of an infinitely long formula which has only one object in the model of the language as an interpretation, or can consist of an infinitely large set of formulae that get their interpretation from infinitely many different objects (objects in the model). Such a language — if the basic argument were correct — would not be learnable. Davidson, however, derives the thesis about compositionality of natural languages from the fact that he thinks that the explanandum (language) has certain properties; theories of natural languages must account for the fact that natural

---

[61] *Ibid.* p. 171.

languages have meaning, have infinite power of expression, and need to be used by beings with finite abilities. In what sense language actually has such properties is a different story. Davidson's motivating statement, however, is not the only one making a connection between the structure of language and understanding. Frege maintained that the

> possibility of our understanding [sentences] which we have never heard before rests evidently on this, that we construct the sense of a [sentence] out of parts that correspond to the words. If we find the same word in two sentences, [...] then we also recognize something common to the corresponding thoughts, something corresponding to this word. Without this, language in the proper sense of the word would be impossible. We could indeed adopt the convention that certain signs were to express certain thoughts , like railway signals ('The track is clear'); but in this way we would always be restricted to a very narrow area, and we could not form a completely new [sentence], one which would be understood by another person even though no special convention has been adopted beforehand for this case.[62]

This is a rather complex inference, and Frege makes reference to some theoretical entities of his own creation.[63] For the purposes at hand it may be sufficient to note that Frege claims that there is an inference from the nature of mental properties to one about language. This inference from properties of the mind to one about the subject of our theoretical interest has a long tradition.

---

[62] Frege (1979) p. 79.
[63] Frege also moves from trivial observations to Model M.

In other areas of theoretical interest, e.g. mathematics or physics, epistemological arguments of this nature certainly don't find universal approval.[64] One cannot without further ado conclude from epistemic considerations to properties of mathematical systems or objects. Mathematics is commonly thought to study systems that are independent of our knowledge of them. No constraints for mathematical objects themselves follow obviously from epistemic considerations. But if natural languages are said to have compositional structure since they need to be understood, understanding needs to be a mechanical process employing the rules that Model M postulates as being constitutive of language. This is the reason why Wittgenstein's and Kripke's rule-following considerations — to be considered below — have attracted so much attention. Understanding or producing a sentence, so it is assumed, is gathering the meaning of a sentence by deriving it from the primitive vocabulary and the rules of composition. Human beings process language, following rules

---

[64] Of course, there are philosophers like Dummett who think that epistemological considerations put a constraint on theories of mathematics:

> Any justification for adopting one logic rather than another as the logic for mathematics must turn on questions of *meaning*. (Dummett (1978) p. 215)

(Dummett's view of meaning is discussed *supra*, note 57). My claim here is not that there are no arguments for this position, although I have my doubts about them. I am merely pointing out that there are no <u>direct</u> inferences from epistemological considerations to substantive claims about mathematics.

of composition and decomposition, and gathering the meaning of words by assigning objects to the symbols. Although the theoretical issues are complex and allow for various interpretations, philosophers agree that understanding language is a process of following rules.

## The Mechanical Model of Meaning (III): Infinity

There is an additional thesis Davidson mentions almost in passing. And this thesis turns out to be what Kripke's skeptic maintains is the stumbling block for Model M type theories of language. Davidson and others attribute to language an infinitary element: language, so the thesis, can actually refer to infinitely many objects with words, and a language user has the capacity to express infinitely many thoughts. This conviction is for some philosophers somehow self-evident:

> We *know* that mental states with general contents are states with infinitary normative character; it is precisely with that observation that the entire discussion began.[65]

I think that this is gem of an example of theory-laden observation. Of course, it is not easy to understand how a manifestly finite apparatus has infinite powers. One can easily see that a theory assuming this will need recursive

---

[65] Boghossian (1989) p. 542.

rules to build up infinitely many expressions from finite elements. Such a theory will need rules for the representation of infinitely large sets of referents to be referred to by a primitive symbol.

Model theory uses functions to assign interpretations to symbols and formulae of a language. But functions are abstract mathematical entities.[66] Given that words are the basic element of language as understood by philosophers of language, the theory must give an explanation of the relation between its basic elements and objects. And since non-natural languages are the paradigm, we might expect that the paradigm for giving an explanation of the meaning of words will be some form of rule that assigns a value to symbols. Kripke's skeptic implicitly puts the preconceptions of Model M to a test, specifically with respect to the claim of Model M, *viz.* that language is potentially infinite. Before I turn to Kripke's skeptical investigation of Model M, I will discuss whether Model M has a cousin in another area of philosophy. In particular, I

---

[66] This is, of course, just as provoking a statement as 'the meaning of a sentence depends on the words it contains'. That this topic of mathematical functions lends itself to similar investigations as those about meaning is the reason that there is some amount of discontinuity of discussion regarding mathematics and rule-following in the *Philosophical Investigations* around section 201; much of the rest of the rule-following considerations were, in the first draft of the *PI*, branched off into the *Remarks on the Foundations of Mathematics*; cf. Floyd (1991) p. 146.

believe that certain moral theories have used rules for very much the same purpose as philosophers of language have: they are held to be objective, rigorous and the very foundation of rationality.

## Kant's Deontology and Rule-Utilitarianism: Morality as a System of Rules

Early education is not only concerned with teaching language but — among many other things — also with moral behavior. After an initial period in which we encourage and discourage specific actions or omissions we switch to a more general form of education by teaching rules of grammar and conduct which children then apply to novel situations. The first chapter intended to show that truistic observations like the one above about rules are used by Frege, Davidson and other contemporary philosophers of language to motivate a picture of language as a calculus of rigid and recursively specifiable rules — a picture I called Model M. In this chapter I will sketch two theories of morality which I think share important features with Model M — a rule-deontological and a rule-utilitarian version of morality. Both versions of a moral system can be found in Kant's *Foundations*, even though Kant subscribes only to the former.

Kant's rule-deontological moral philosophy is based on the view that the practical principles, concepts and basic laws (*Grundsätze*) of morality are founded in the idea of freedom of the will of the noumenal self. Although this metaphysical aspect of Kant's system is fundamental for his system, I will not discuss the notion of freedom or

autonomy.[67] I will restrict my discussion of Kant's theory of morality to the fact that it is a paradigmatic rule-deontological moral theory. For Kant, morality is not seen as a practice of generating rules from particular cases, restricting the validity of rules to specific areas and circumstances from which they are gathered. Instead, Kant's theory develops a picture of morality as a system of strict, abstract, and a priori rules of practical reason, basic principles which determine the maxims of a (good) will and thence the moral character of an action independently of human interest and circumstances. According to Kripke's skeptical argument, which I will present in the next chapter, such an a priori connection between rules and determination of use and action may not be straightforward. The proposed solution to Kripke's paradox I present in chapter 4 attempts to show that the paradox is intimately connected with the view of rules as rigid, abstract mechanisms that determines an application independently of our practices.

---

[67] I am not concerned with the issue whether persons who act upon rules behave autonomously. I doubt that this can be made coherent at all unless we a) either give the notion of autonomy an ordinary notion (and then we neither need such things as the veil of ignorance, or the transcendental notion of freedom) or b) if the Kantian system can be made coherent. I doubt anybody seriously considers this a real possibility. Although I am sympathetic to the former 'solution', I think it is not in Kant's interest. For a defense of such a view, see Baumrin (1976).

The moral theories I will present in this chapter continue the theme I outlined for philosophy of language in the first chapter. Historically, Kant's philosophy has had a strong influence on Frege. In particular, there are two aspects in which the Kantian system of morality is similar to Frege's contribution to a philosophy of logic and language. Kant's system of morality has, similar to Fregean semantics, two elements: a supersensible notion (freedom/sense), and a system of strict principles or rules (determining the moral value of a will/the meaning of an utterance). While Frege's *Begriffsschrift* is a systematic, rigorous calculus-language expressing all of logic and mathematics, Kant's moral philosophy is an attempt to establish ethics as a system of rigorous, a priori rules determining the will and the moral evaluation of an action.[68] Kant's theoretical work which inspired Frege who in turn inspired many contemporary philosophers of language, and his ethics continue to have a strong impact on philosophy, even though the transcendental aspect of his philosophy is considered less important for contemporary moral philosophy.[69] Kant held that without the principles

---

[68] Principles, categories, schemata, laws and concepts are all rules, according to Kant. The notion of a rule seems to have been rather unproblematic for Kant.

[69] One might note, however, that Frege's theories are much more influential now than they were when Quine's empiricism was dominant. It may thus come about that Kant's transcendental moral theory will have a comeback as well.

of theoretical reason and the pure forms of intuition, the notion of a cause (the necessary nexus between events) would be the illusion Hume claimed it was. Kant also maintained that morality in general, and the notion of duty in particular would be an illusion if morality cannot be given a foundation in a priori practical principles. He states:

> Everyone must admit that a law, if it is to hold morally, i.e. as a ground of obligation, must imply absolute necessity; he must admit that the command "Thou shalt not lie" does not apply to men only, as if other rational being had no need to observe it. The same is true of all other moral laws properly so called. He must concede that the ground of obligation here must not be sought in the nature of man or in the circumstances in which he is placed, but sought a priori solely in the concepts of pure reason, and that every other precept which rests on principles of mere experience, even a precept which is in certain respects universal, so far as it leans in the least on empirical grounds ... may be called a practical rule, but never a moral law.[70]

Moral philosophy must establish the foundations of this law, and thus discover the basis for the a priori rules of the will:

> The moral law can be found in its purity and genuineness (which is the central concern in the practical) nowhere else than in a pure philosophy; therefore, this (i.e. metaphysics) must lead the way, and without it there can be

---

[70] Foundations of the Metaphysics of Morals, p. 389. I will abbreviate reference to this work by *FMM*, followed by the page number of the "Academy Edition". CPR will refer to the *Critique of Pure Reason*, CPrR to the *Critique of Practical Reason* and CJ to the *Critique of Judgment*. All page references refer to the "Academy Edition".

no moral philosophy. (FMM 390)

Kant — this is the second point of similarity with Frege's work — abandons the psychological and phenomenological aspects for his theory. When Frege writes "I believe that the sense of a name is not something subjective",[71] Kant's 'practical' counterpart is that the "good ... determines the will ... not by subjective causes."[72] The first part of this chapter will thus investigate the Kantian foundations of morality from the perspective of a system of a priori principles - independent of the subjective (empirical) part of human nature - which are held to determine the will and all moral action *qua* moral.[73] Kant's reason for devising a system of a priori rules or principles of morality are in line with a long tradition in philosophy that sees rules as somehow the expression of

---

[71] Frege *Letter* in Moore (1993), p. 45.
[72] FMM 413.
[73] Kant held that a priori principles, categories and concepts are absolutely rigid, while acknowledging that rules in the empirical arena are open to exceptions. (Cf. N 5234 and N 5237). Kant's attempts to give a "Deduktion" of the various theoretical and practical principles - a justification that and how the categories are related to their domain - are notoriously difficult to understand. This holds for the transcendental Deduktion, the schematism chapter, and for the third section of the *Foundations*. (cf. Henrich (1975)). These parts of the Kantian system have the task to show the connection between the a priori rules and empirical rules or laws, and I know of no-one who thinks that Kant succeeded in detailing the connection between those different kind of rules. I will thus not attempt to explain how empirical rules can be open to exceptions. There are no exceptions, however, for the a priori, necessary rules of pure reason and pure practical reason.

rationality or as eternal principles that are independent of the mind and fundamental for the laws of nature and the universe. The nature and origin of the principles and rules is thus open to much discussion. But Platonism — Frege is a Platonist — and Kantianism see the need to give a metaphysical picture of the status of those rules and principles precisely because experience gives no foundation for what is seen to be their metaphysical character: the fact that they are primary and necessary.[74] Kant took rules to be the signature of reason and rationality, so that his system of morality could liberate us from the dogmas of religion and natural law theories without abandoning the notion of a moral law. He was looking for a system of morality, a totality of rules governed by strict principles, much like Frege after him was looking for a system of rules for the foundations of logic.

Although some contemporary writers — Rawls is an example

---

[74] While Kant held that empirical rules are open to exceptions, Dworkin (1977) (following Rawls (1955)) distinguishes between rigid rules ("black letter rules" (*loc. cit.* p. 46)) and principles which have a different "force" from rules in arguments (*loc. cit.*, 71) and are not rigid; principles, according to Dworkin can contradict each other without making either of them valid. I think that the fact that philosophers vary in their claim regarding rules may be a symptom of how easily we get confused about the function and status of rules. I will expand on this topic in the fourth chapter.

— see themselves in the Kantian tradition,[75] this rests on what one might call a 'pragmatic' interpretation of Kant's moral philosophy.[76] Such a pragmatic interpretation can take several different routes to incorporate some of Kant's writing which I will not discuss. The second part of this chapter will, however discuss a version of morality that has a pragmatic or empirical element to it. The version of rule-utilitarianism I shall present in the second part of this chapter also relies heavily on a picture of rules which will come under a skeptical attack in the next chapter. Kant discusses a version of rule-utilitarianism in the *Foundations*, and although Kant rejects the view that the principles of morality are founded on something other than pure practical reason, he sketches a version of rule-utilitarianism — a moral theory which regards the principles and rules of morality as a posteriori, but nevertheless ideal and systematic principles of morality. That is, Kant's theory discusses a version of rule-utilitarianism as well as rule-deontology. I will

---

[75] Rawls' approach would not be acceptable to Kant who claims that morality has to be founded on the a priori principles of practical reason which he attempts to do in the *Foundations*. Rawls makes quite clear that his 'foundation' of 'fairness' is political, therefore empirical, not metaphysical. (Compare Rawls (1985))

[76] I will not discuss the virtues of such a pragmatic version of Kantian moral philosophy. Despite Rawls' allegiance to Kantian moral philosophy (Rawls 1980), there are good reasons to think that Rawls is Hegelian as well. See Schwarzenbach (1991).

first address Kant's system, and then discuss rule-utilitarianism in the second part of this chapter.[77] In the next chapter, the view of rules discussed here as well as in the first chapter will confront Kripke's skepticism about rule-following.

## Kant's pure practical philosophy

Kant begins the *Foundations* with the following famous claim:

> Nothing in the world — indeed nothing even beyond the world — can possibly be conceived which could be called good without qualification except a **good will**. ... [T]he other talents of the mind [and] qualities of temperament are doubtless in many respects good and desirable. But they can become extremely bad and harmful if the will ... is not good. ... [H]appiness makes for pride but then also for arrogance if there is not a good will ... to make it universally conformable to its end. ... Thus the good will seems to constitute the indispensable condition even of the worthiness to be happy. (FMM 392-3)

He continues:

> The good will is not good because of what it effects or accomplishes or because of its

---

[77] Before I turn to Kant's theory, I would like to put up another sign-post: The reason I discuss these rule-oriented moral philosophers is that Kripke's rule-following paradox states that we cannot know which rules we follow, since any action is compatible with an infinite number of rules. Apparently, then, any type of theory that sees rules at the center of its subject matter seems thereby doomed. In the fourth chapter I will try to resolve this paradox, presenting a view of rules and rule-following which suggests that we revise some of our philosophical views regarding rules and rule-following.

> adequacy to achieve some proposed end; it is good only because of its willing, i.e., it is good of itself. (FMM 393)

The claim that the only thing in the world that is good without qualification could be a good will may at first seem not so obvious, and Kant seems to have a very different methodological starting point from Davidson and his philosophy of language.[78] Kant's remarks may be read as a *précis* of a critique of what seems a much more obvious approach to morality. Aristotle — Kant arguably refers (here as in the *Critique of Practical Reason*) to Aristotle's *Nicomachean Ethics* — held that ethics is concerned with the flourishing of human beings. Human flourishing is leading a life that can be described as perfecting the right combination of mental, psychological and societal virtues.[79]

---

[78] In the first chapter I try to show that philosophers of language - including Frege - try to motivate a picture of language as a system of rigid, recursively specifiable rules by appealing to truistic statements like "the meaning of a sentence depends on the meaning of the words it contains". The point I am concerned with here is that Kant's notion of a good will as the only really good thing may appear truistic to those (it clearly would not be truistic to Aristotle, who maintains that ethics is about action, and has no notion of the will — at least not the one Kant as in mind) who grow up in a culture dominated by Christian religious views.

[79] Plato, in the *Republic*, distinguishes between three parts of the soul, and determines that justice can only be attained if reason governs the 'lower' parts (emotions and vegetative parts) which have to obey. Although Aristotle seems to say that the life of contemplation is the highest good, he also states that "such a life would be too high for man" and his composite nature. (NE 1177b27). Although he thus ranks the parts of the soul much like Plato, he attributes intrinsic moral excellences to the 'lower' parts

(continued...)

This seems the more plausible view (to me, at least) about what ethics is about: it teaches a life worth living. One may thus think that Kant intended to revolutionize ethics.[80]

But Kant wrote at a time when Europe was practically exclusively Christian. His audience, so Kant's view, accepted as common sense the claim that the will — rather than, say, action or material goals — was the only thoroughly good thing. Kant, in any case, thought that his thesis was seen in need merely of an explanation, not a proof:

> But there is something so strange in this idea of the absolute worth of the will alone, in which no account is taken of any use, that, notwithstanding the agreement even of common sense, the suspicion must arise that perhaps only high-flown fancy is its hidden basis, and that we may have misunderstood the purpose of nature in its appointment of reason as the ruler of the will. (FMM, 394-5; emph. added.)

The *Groundwork* develops a theory of the good will that is far from trivial, and the notion of the 'good will' does not turn out to be anything like what "common sense" has in

---

[79](...continued)
of the soul. Kant clearly sides with Plato: Ethics has to do with rational principles, not with the virtues of different parts of the soul.

[80] Baumrin (1982) contrasts procedural and substantive ethics, the former emphasizing fairness, the latter justice. Kant's theory may thus appear procedural, while Aristotle's theory tends to be more interested in a substantive theory of justice.

mind. This is the case even though the *Groundwork* tries to explain the a priori conditions for the possibility of the ordinary notion of duty, just as the first *Critique* tried to explain - and defend against Hume - the ordinary notion of a cause. Kant starts from the ordinary moral consciousness, which (he says) holds — if obliquely — that the only absolutely good thing in the world is the good will:[81]

> We have then, to develop the concept of a will which is to be esteemed as good of itself without regard to anything else. It dwells already in the natural sound understanding and does not need so much to be taught as only to be brought to light. In the estimation of the total worth of our actions it always takes first place and is the condition of everything else. In order to show this, we shall take the concept of duty. It contains that of a good will, though with certain subjective restrictions and hindrances ... . (FMM 397)

Kant's theory of the rational foundations of morality is designed to search for and establish the supreme principle of morality in order to explain a practical synthetic a priori statement, *viz.* that a human agent *qua* agent has to have a maxim of a certain type in order to be moral.[82] Kant proceeds in three steps: The first section of FMM analyzes the ordinary notion of duty, in order to establish that it rests on an objective principle. This awareness of an objective principle creates a conflict 'in ordinary

---

[81] See also *Prolegomena*, 297.
[82] Cf. FMM 447.

consciousness', since a person is also aware of other — "subjective" — aims. In the second section Kant gives therefore a formulation of the objective principle as it pertains to the human will - the Categorical Imperative - and shows how the notion of duty can be derived from principles of pure practical reason.[83] In the third section Kant shows how the Categorical Imperative is possible.[84] Although the *Groundwork* thus gives an answer to the question how a practical synthetic a priori statement - the Categorical Imperative - is possible, and why the principles of pure practical reason are obligatory for humans, we need to consult the *Critique of Practical Reason* in order to understand how the principles of morality — a priori rules of rationality — determine the will.

---

[83] FMM 412.
[84] For reasons I will mention below, I will not be concerned with the Categorical Imperative at all.

Duty and principles

In the first section of FMM Kant does not prove the thesis of the good will as the only moral principle; this would have to be done by analyzing the notion of *will*, which Kant does not set himself out to do in the first section. Instead, Kant explains his thesis of the good will by stating that the good will is will governed by reason (FMM 394/5). Nature must have provided reason — rather than instinct (FMM 395) — to govern the will in order to establish a will that is good qua will (FMM 396). In order to establish the notion of the good will as the only thoroughly good thing, Kant analyzes the notion of duty which he finds in ordinary moral consciousness. The notion of duty contains (with some subjective restrictions) the notion of a good will (FMM 397). Kant establishes three propositions pertaining to the notion of duty in the first section, while sections two and three are concerned with giving the metaphysical foundation for them.

The three propositions of the first section are

> An act has moral worth only if it is "done not from inclination but from duty".[85]
>
> "An action performed from duty does not have its

[85] FMM 397-99; Kant himself does not provide a formulation. At FMM 389 he states that an esteemed action must be "done not from inclination but from duty", which is the formulation I chose.

> moral worth in the purpose which is to be achieved through it but in the maxim by which it is determined." (FMM 399)
>
> "Duty is the necessity of an action executed from respect for law."[86] (FMM 400)

In order to establish [1] Kant analyses the notion of duty, which requires that an action be done from duty, not from inclination. The notion of duty implies that we are being commanded to perform an action, and "an inclination cannot be commanded" (FMM 399). Kant establishes [2] by providing an analysis of the notion of an action that is done from duty. Such an action has to be analyzed as containing an intention (will)[87] and its object. The object of an intended action, however, cannot properly be called morally good since the effects could be brought about by other causes as well:

> [A]ll ... effects ... could be brought about through other causes and would not require the will of a rational being, while the highest and unconditional good can be found only in such a

---

[86] The last proposition regarding duty is a synthesis of the two aspects of moral action as stated in the first two propositions. The first proposition considers the objective aspect — or aim — of an action. The second proposition addresses the subjective aspect of action, namely, that only a certain aspect of an agent's intention may be taken into consideration in determining the morality of an action. The third proposition is a synthesis of these statements. It joins the objective aspect and the subjective aspect (the maxim) through the notion of respect, a feeling that is caused by a rational concept.

[87] The notion of 'duty' analytically contains 'will'. Davidson (1963) argues more generally that we cannot give an explanation of the notion of an action without involving the notion of intention.

will. (FMM 401)

If the effect is not the element of the action that determines its moral character, then neither the actual effect, nor intending that effect are relevant in the case of an action from duty. The analysis of the concept of duty thus establishes that an action that is morally worthy is not worthy because of the object or content of a volition. Since an action in general, and one done from duty in particular requires an intentional component, what remains of an intention once its object has been abstracted is a formal (non-empirical, non-subjective) principle of action. Willing something differs from acting out of impulse or instinct, which are reflexes. Willing an action requires that I present an action to myself as a consequence of a principle of a certain type. But since all effects and aims were established to be relative, it is merely the principle *qua* principle that is morally relevant in an action done from duty. Since the notion of a moral act contains the notion that I am obliged to follow a principle — a duty is a type of command — the principle of an action done from duty is a law, that is, it is universal, necessary and obligatory. Consequently, the will, according to Kant, is nothing but the faculty to act according to the presentation of laws.[88]

---

[88] FMM 412

> Therefore, the pre-eminent good can consist only in the conception of the law itself in itself (which can be present only in a rational being) so far as this conception and not the hoped-for effect is the determining ground of the will. (FMM 401)

Although it is, for example, a duty to preserve one's life, the duty does not consist in the effect (staying alive), but in a principle that such and such maxim is necessary - a form of a command which is a law.[89] After establishing the second proposition that the formal principle of volition, not its object, is morally relevant, Kant establishes [3] by showing that fulfilling one's duty is motivated by rational interest or respect, a feeling that is caused by a rational principle. Since the notion of the principle of the will of an action done from duty contains that of a law — which in turn contains the concepts of necessity and objectivity — an action can only be called good if it arises from an awareness that it is in conformity to that law. A law is universal and necessary and these characteristics of the law make it founded in reason. The motivation to follow the law is thus not externally received, but a spontaneous feeling — directly (not externally) affecting the subject by effecting the feeling of respect:

---

[89] The principle, of course cannot be willed, but the will is determined by it, that is, "by the formal principle of volition" (FMM 400)

> [R]espect is a feeling [which is not] received through any [outer] influence, but is self-wrought by a rational concept. (FMM, 401 note)

An action from duty is thus motivated by respect for the law; respect in turn is the immediate determination of the will by the law.[90] Although a human being has to form maxims (principles of action) in order to act, a maxim derives its moral worth from its conformity to the objective and formal characteristics of the good will, and not from its aims. Since the notion of a duty implies that it is a precept, abstracting the aims from a precept leaves only the form of this command, that it is a law, so that a moral act is one where I form an intention - a maxim - which could become a law. The analysis of the notion of duty thus reveals that the notion of a good will means that the will should have lawfulness as its principle:

> I should never act in such a way that I could not also will that my maxim should be a moral law.[91]

But since humans beings do not have only a good will, but have also non-rational principles that influence the

---

[90] In what follows I will not discuss Kant's theory of rational motivation in any detail.

[91] FMM 402. The term *will* in this statement seems ambiguous, since it requires that I will some law, and it is not quite clear which will is to have this volition. There is a formulation at FMM 436/7 which does not refer to a *will*: "Act according to the maxim which can at the same time make itself a universal law." For an anlysis of what is expressed in the Categorical Imperative, see Footnote 109 below on page 93.

will, ordinary consciousness starts an inner dialectic and wants to know the source of such a duty. For this, conceptual analysis is not sufficient, because we now have to show how the notion of duty can be derived from pure practical principles. This is the task of the second section. The second section of the *Groundwork* is — like the first section — analytic. In the first section of *Groundwork*, Kant took the notion of duty as a fact of ordinary moral consciousness in order to further explain the thesis that only the good will is absolutely good. The result of the analysis of the notion of duty establishes that duty as a paradigmatic moral notion that implies a command also implies that actions are only absolutely good when they are intended according to a principle that could be a universal law.

Rules Determining the Will

The second section of the *Groundwork* investigates the principles of practical reason in order to show how this notion of duty can arise from rational principles and also in order to differentiate maxims that are moral from those that have other principles of volition. The "popular ethical wisdom"[92] cannot provide such a priori principles. The history of ethics indicates that such an investigation of the principles of morality has always been thought necessary in order to establish whether there is a justification for the notion of a duty or moral action. According to Kant, ordinary consciousness undergoes a natural dialectic when deciding which principles of action it should follow and issue. Some of our maxims are principles of action that are formed to fulfill desires — the type of maxims that contradict practical reason's demand to fulfill our duty. For example, "Maximize profit" might be a maxim according to which we should overcharge someone asking for a store-item whose market value she does not know. We are, however, also aware of our duty to refrain from such overcharging. But although we are aware of the notion of duty, this principle of action is merely one of the very many we have. The maxim to maximize profit may compete with the maxim to establish social relations

---

[92] "condescending to popular notions"; FMM, 409.

(because one derives pleasure from them) and both might compete with a principle of action that is in accordance with the Categorical Imperative. We thus need to know the origin of the notion of duty, not only what it analytically entailed by it, in order to have moral requirements firmly established.[93]

Philosophers before Kant have attempted to give an empirical explanation of morality. Kant points out that this attempt is misguided. The notion of duty contains not only the notion of an obligation, i.e. that an action ought to have an intention of a certain kind — which is a type of obligation or command. Like the notion of a cause, the notion of duty also contains the notion of a law. This requires that the principle of any moral action — action from duty — must be universal, necessary, and obligatory. Notions like 'necessity' and 'universality' cannot be explained by an appeal to experience; any attempt to derive moral concepts from experience is therefore misguided. The notion of a moral law has therefore to be derived from the notion of rationality — the only source from which necessity

---

[93] MacIntyre (1988, p. 43-48) accuses Kant's ethics of rigorism — the fact that his morality is a system of rules — and formalism — the charge that there are no substantial implications from those rules. MacIntyre is thus correct in his first assessment of Kant's system. According to Kant, he believes that there are maxims that can be derived from the rules of morality. I will discuss this issue of formalism further below; see also Footnote 110.

and universality can be derived. Since the notion of duty cannot be derived from examples, an attempt to show that the notion of duty is not empty needs to derive it from the very notion of rationality's universal rules.

This requirement, that the notion of duty be derived from the notion of pure practical rationality does, of course, not establish that there is such a universal law. The concept might be 'empty', that is, the notion of duty may be an illusion. The concept, e.g. of a witch, implies that she is evil, but there aren't any witches. The concept of duty nevertheless implies the concept of a will that acts according to a universal law and thus also implies that a will ought to act according to a non-empirical, a priori principle. Ethical concepts and laws have to be derived from pure practical reason, and the second section thus investigates where the notion of duty arises in the realm of pure practical rationality.

> If there is thus no genuine supreme principle of morality which does not rest merely on pure reason independently of all experience, I do not believe it is necessary even to ask whether it is well to exhibit these concepts generally (in abstracto), which, together with the principles belonging to them, are established a priori. (FMM 409)

Contrary to the principles, categories and basic laws of the understanding as described in the *Critique of Pure Reason*, the principles of practical reason must not be restricted to the nature of human reason. The principles of

pure reason have to be restricted to the forms of our intuition, time and space, in order to become empirical knowledge ("Erfahrung"). The principles of practical reason, however, not only need no such restriction, but the first section has shown that any reference to aims or circumstances has to be excluded from considerations regarding objective moral principles, if the notion of duty is not to be empty. Thus, all ethical concepts have to have their origin in pure practical reason and are valid for all rational beings, regardless of their empirical situation.

> [A]ll moral concepts have their seat and origin entirely a priori in reason. ... It is obvious that they cannot be abstracted from any empirical and hence contingent cognitions. ... It is ... of the utmost practical importance to derive the concepts and laws of morals from pure reason ... without making the principles depend upon the particular nature of human reason ... . ... [S]ince moral laws should hold for every rational being as such, the principle must be derived from the universal concept of a rational being generally. (FMM 411f)

The will Kant discusses is thus not empirical will, but pure will.

> The union of causality of freedom with causality as the mechanism of nature, the first being given through the moral law and the latter through natural law, and both related to the same subject, man, is impossible unless man is conceived by pure consciousness as a being in itself in relation to the former, but by empirical [consciousness] to the latter. [94]

---

[94] CPR 6, note. The original reads " ... jenes [Sittengesetz] im reinen, dieses [Naturgesetz] im empirischen Bewußtsein ... " I corrected Beck's 'empirical reason' with 'empirical consciousness'.

That is, while the first section analyzed the notion of duty in order to establish its presuppositions, the second section establishes how the notion of duty requires that it be derived from pure practical reason. This is a metaphysical investigation which needs no examples:

> In this study ... we advance by natural stages from a popular philosophy (which can go no further than it can grope by means of examples) to metaphysics (which is not held back by anything empirical and which, as it must measure out the entire scope of rational knowledge of this kind, reaches even Ideas, where examples fail us). In order to make this advance, we must follow and clearly present the practical faculty of reason from its universal rules of determination to the point where the concept of duty arises from it. (FMM 412)

An action from duty requires that we intend it as well as execute it. Since such an action needs to be issued in the empirical world, the intention must take into consideration that everything in nature happens according to laws, that is, universal and necessary rules.[95] In order to act (and not just to be subject to the causal mechanism of nature) I have to present to myself the laws according to which I act. This principle of action by which I present laws to myself Kant calls "will". But the principles in turn are abstract and need to give rise to particular maxims of the will. Although, for example, I may present to myself the law that I should not harm anyone, I still need to

---

[95] "Nature" does not mean 'world as we know it', but merely "objects under laws". Angels, for example, or any other rational beings may not have our forms of intuition.

derive particular actions from such more universal laws. Since action requires a will, the will is nothing but the principle by which I derive[96] particular actions from more universal laws:

> Since reason is required for the derivation of action from laws, will is nothing else than practical reason. (FMM 412)

Although the will issues in actions according to the presentation of principles, the fact that I have a will does not show which principles determine my will. The notion of a rational being does not exclude that the being has desires or inclinations. If the will derives all principles of action merely from the principles of reason, the will's principles would all be necessary. But since humans are rational creatures which also have inclinations and desires, a human will does not always derive the principles of action only from principles of reason. This does not imply that those actions do not have principles; but the principles are not automatically those of practical reason. The principles of action derived from non-rational principles are what Kant calls subjective principles, and they are contingent. If practical reason completely determined the will, then all the objectively necessary principles would also be

---

[96] The will also is supposed to cause action. But this is an entirely different issue. The point here is that the will has a practical subject matter, but that does not cause an action. The will is responsible for the maxim, and for the execution of the action, but these issues are separate.

subjectively necessary. For such a will, there would be no duties, but merely moral laws. There is thus a difference between laws and moral commands. The concept of law does not imply normativity — all laws of nature and reason are descriptive. But if the will is not completely determined by reason, then we have the notion of a command, whose formula is called an imperative (FMM 413). A totally good will would be under objective laws, but would not have to be commanded. Maxims are needed for all agents, but imperatives, the command that the maxim be in accordance with the form of the will selecting objective principles of practical reason are necessary only for rational beings that are not entirely governed by (practical) reason, but are also directed by desire (have an heteronomously conditioned will). The Categorical Imperative, and this is the nub of the whole *Foundations*, is a synthetic a priori statement, since it prescribes that a will adopt the principles of pure practical rationality:

> [W]e have to investigate purely a priori the possibility of a Categorical Imperative, for we do not have the advantage that experience would give us the reality of this imperative ... [which] alone implies the necessity which we require of a law.
>
> ... This imperative is an a priori, synthetical practical proposition. (FMM 419f)

In a footnote, he explains that the imperative

> connect[s] a priori, and hence necessarily, the

> action with the will without supposing as a condition that there is any inclination [to the action] (though I do so only objectively, i.e., under the idea of a reason which would have complete power over all subjective motives). This is, therefore, a practical proposition which does not analytically derive the willing of an action from some other volition already presupposed (for we do not have such a perfect will); it rather connects directly with the concept of the will of a rational being as something which is not contained within it. (FMM, 420, note.)

The rest of the *Groundwork* is concerned with explaining how such a statement is possible. Although the notion of duty contains the notion of a law the agent needs to present to himself — which is possible only for a rational being[97] — Kant is not concerned with explaining that there is such a thing as a law or rules. Laws of logic, for example, need no transcendental explanation since they are analytic.[98] The notion of duty is found in ordinary moral consciousness, which forms intentions or maxims according to practical reason as well as desire, as Kant states. Accordingly, so Kant, we need a philosophical investigation into the origin and legitimacy of moral notions. Morality requires that, for any moral act whatsoever, one direct one's will according to the principles of pure practical reason, and this demand and the possibility of practical reasons determining the will has to be explained. Since the notion

---

[97] FMM 401.
[98] Analyticity seems not to have been much of a problem for Kant. For a criticism, see Quine (1953), p. 20/21.

of duty was found to require a priori principles of reason, neither the notion of duty, nor that of a moral action are empirical notions. The will Kant refers to is not empirical will, either. Since Hume questions the validity of the notion of a cause, Kant is aware of the fact that he needs to show how such a synthetic a priori statement — the Categorical Imperative — is possible. That is, Kant's project assumes that the rules of morality are synthetic a priori (necessary), and his system it designed to show how this is possible.

Although Kant's theory of morality proceeds by analyzing the notion of obligation, the notion of an obligation is not entailed by pure practical principles. Only the fact that the will is absolutely good if it is determined by reason is implied by pure practical principles. The notion of duty is a requirement for human anthropological morality, or morality for beings that are not perfectly rational. The will is the faculty that forms maxims according to principles, and an analysis of practical reason shows that practical principles themselves are universal, since they are valid for any rational creature. The crucial question for Kant's foundation is the question of the 'Deduktion', how to justify that practical principles become obligatory for non-rational beings. The aim for the *Foundations* — to show how a human will (which is capable of forming

subjective and objective principles of action) can be obligated to perform an action according to practical rationality is addressed in the third section. I will, however, not be concerned with Kant's view that the a priori principles of action are obligatory for humans since a human can form the idea of herself as a free rational being. The question as to what makes the principles of practical reason obligatory — that they are imperatives — has to be distinguished from the question, how the pure practical principles determine the will — how they become maxims that could be laws of nature. For the understanding of moral rules *qua* practical principles — not rules for anthropological morality — Kant's question, how the Categorical Imperative is possible, need not be discussed. This issue is interesting in itself, but relevant only for the 'Deduktion', the justification and explanation of how a being who has other elements than rational desires in his faculty of desire (Begehrungsvermögen) can have the notion of an obligation. This is a transcendental question, while I am concerned only with the systematic question of how a will (not only a human will) is determined by the principles of practical reason. Whatever principles or rules Kant derives from his idea of a pure practical rationality, they

are all valid, a priori, for rational beings.[99] Kant's third section of the FMM is thus an attempt to derive normativity, not to derive that certain practical rules of reason are in place, and that they entail a moral code. The principles of pure practical reason are objective, not normative.

Kant's ethical theory is thus primarily a theory of pure rational motivation, from which an anthropological morality is to be derived. This allows him to define the notion of a pure interest as interest which has a maxim that is universalizable.

> Interest is that by which reason becomes practical, i.e., a cause determining the will. (FMM 459 Note)

---

[99] My reading thus emphasizes the rational, not the imperatival aspect of Kant's theory, which I believe is his main concern. O'Neill (1975, esp. ch 3) and Herman (1993, ch. 10) assume that the principles apply to situations. Kant explicitly mentions that examples are not to be used in metaphysics, that they are a symptom of the "popular philosophy (which goes no further than it can grope by means of examples)." (FMM 412) To show the level of abstraction at which Kant writes, one has to remember that the laws of nature the understanding prescribes for the phenomena, are, for example, expressed in the category of causality which states that every change has a cause. There is a long step from this metaphysical principle to showing whether electrical discharge causes lightning. Similarly, Kant's practical principles and the categories of pure practical reason are abstract in the extreme. I might add that it is the complement to my thesis that philosophers 'blow up' ordinary notions into metaphysical statements that metaphysical statements in turn are 'cooked down' to have ordinary meaning. It is this principle that underlies the 'pragmatic interpretation of Kant'. In this sense I welcome the pragmatic interpretation.

A direct interest in the action is taken by reason only if the universal validity of its maxim is a sufficient determining ground of the will. Only such an interest is pure. Kant thus systematically "purifies" and formalizes motivation.[100]

The *Foundations* are concerned with showing how the notion of a will and the notion of rationality are connected. The issue of moral rules and rule following is alike for rational beings insofar rules are an expression of this rationality.

---

[100] Herman has a different perspective when she writes: "Having introduced ... the idea of the good will as the only possible unconditioned good, and shown that the unconditioned good is possible only if there can be an interest independent motive for action that takes lawfulness itself as a final and sufficient reason, Kant argues in the second chapter that a rational will could be so motivated only by an unconditioned rational principle (law) the Categorical Imperative." (1993, p. 209, note, emph. added). *Pace* Herman, I maintain that Kant's notion of pure, rational interest is primary. A rational will needs no motivation, and the Categorical Imperative is not a rational principle (it's a principle that makes maxims rational, i.e. moral). I think that much of Herman's presentation seems to overlook this point of rational, pure interest of reason, which distinguishes between the practical principles that determine the will, and thus make the will good. It is important to realize that the Categorical Imperative is normative, that is, it is a command to be good, and it gives an instruction for being good — developing a maxim that is in accordance with pure practical principles. Herman, I believe, conflates the two.

Rules and Application of Principles of Morality

I will thus bypass all aspects of the Kantian doctrine that concern the dual aspect of human agents as determined by principles that have their origin in their desires and those of practical reason.[101] This theoretical abstraction from human — empirical or pragmatic — principles is common place since Kant. Chomsky's idea of an ideal speaker — developed in his *Aspects of a Theory of Syntax*[102] — is an example of such a theory. The principles of language Chomsky develops do not take into account the "subjective restrictions and hindrances"[103] of a human speaker, who is, after all, a linguistic "agent". Similarly, an ideal moral agent[104] is a person (not a human) who is a rational agent without what Kant calls heteronomous influence on the will — an agent that has no desires or principles other than the principles of practical reason which determine the will, that is, for whom the subjective maxims and the objective principles are identical. An ideal moral agent needs no imperatives. According to Kant, the

> pure philosophy of morals (metaphysics) can be distinguished from the applied (i.e., applied to human nature), just as pure mathematics and pure logic are distinguished from applied mathematics and applied logic. ... [M]oral principles are not

---

[101] FMM 413.
[102] See Chomsky (1965).
[103] FMM 397.
[104] Kant talks of the godly or holy will (FMM 414).

> founded on the peculiarities of human nature but must stand of themselves a priori, and that from such principles practical rules for every rational nature, and accordingly man, must be derivable. (FMM 410 note.)

That is, we are concerned with the holy will, a pure will that does not contain any relation to desires. The status of the rules (principles, laws, categories, and concepts are all rules for Kant) is thus pure and a priori as well. The principles of morality apply only to those creatures that act, that is, for whom the will is an instrument to present to themselves the law of an action. All events in the realm of experience in which we act are causally determined. This kind of determination depends on the faculty of understanding, which subsumes the manifold of intuition into the unity of consciousness. That is, the only rules that seem to apply to events are the a priori principles of the understanding, so that pure practical principles are either no different from the categories of the understanding — in which case there are no moral principles — or they are empty, since there is no realm to which they would apply.[105] There seems thus no room left for determining the will with anything but causality.

The Critique of Practical Reason discusses this issue, *viz.* how the laws of practical reason determine the will. Regardless of the type of empirical objects of the will, the

---

[105] CPR 68.

will can nevertheless be determined by reason:

> Good or evil always indicates a relation to the will so far as it is determined by the law of reason to make something its object, for the will is never determined directly by the object, and our conception of it. Rather the will is a faculty to make the rule of reason the cause of an action, whereupon an object becomes real.[106]

The application of the categories of the understanding needed a procedure, an application to the forms of intuition. When using the principles of pure reason,

> everything depend[s] upon intuitions to which pure concepts of the understanding could be applied, and such intuitions (though only of objects of the sense), as a priori and hence concerning the connection of the manifold in intuitions, could be given a priori in conformity to the concepts of the understanding, i.e. as schemata. (CPR 68)

A schema, according to the first *Critique*, is a procedure for each category (and concepts in general) by which they are realized, that is, have an immediate relation to what is given by the senses. According to Kant, the connection between categories and appearances is mediated by the transcendental determination of time. But in the case of the moral concepts or principles, no such application can

---

[106] CPR 60; The last sentenced of the Beck translation reads: " ... rather the will is a faculty which can make an object real", leaving the part about rules altogether untranslated: "[W]ie er [the will] denn durch das Object und dessen Vorstellung niemals unmittelbar bestimmt wird, sondern ein Vermögen ist, sich eine Regel der Vernunft zur Bewegursache einer Handlung (dadurch ein Object wirklich werden kann) zu machen."

be given, since they do not apply to appearances at all. Appearances are 'closed' under the notion of causality. It turns out that the opening statement of the first section of the *Foundations* is a direct consequence of the views Kant expressed in the *Critique of Pure Reason*. Since everything in the realm of knowledge — the world of appearances — is determined by the categories of the understanding, the notion of a good action is either chimerical, or 'good' must be understood in reference to some other realm that lies outside the domain of the categories of the understanding *viz*. the categories of practical rationality. The pure will — neither actions, nor the empirical will (which both are appearances) — is determined by the law of reason. The will is a capacity to make a rule of reason a reason for an action.[107] For such a rational creature

> a principle of reason is in itself the determining principle of the will, without regard of the possible objects of the desire (that is, only the lawlike form of the maxim, in which case this principle is a practical law a priori, and pure reason is assumed to be practical per se. The law then determines the will immediately .... (CPR 62)

This moral law determines and gives a deduction of the concepts of good and bad. The law must determine the will immediately and a priori, and according to this will, the object of the will. Moral principles apply to the will, not

---

[107] See CPR 60.

to an object or an action. The formal law prescribes reason nothing but the form of its universal law to be the highest condition of the maxims.

> The concepts of good and evil are the consequence of the determination of the will, and thus presuppose that pure reason is causally active. Therefore, the concepts are not about objects, but are only modi of causality, in so far the determination of this causality is effected by presenting a law by pure reason. (CPR 64)
>
> and:
>
> The determinations of the pure practical reason are possible only to unify the manifold of the desires under the unity of consciousness of practical reason or will, which determines by means of a moral law. (CPR 65)

That is, Kant is here concerned only with the determination of the will, not its practical execution. The practical principles of reason are to determine which maxim I should adopt; they do not cause an action. The practical concepts in respect to freedom (the categories of freedom) acquire meaning, since they create the intentions to which they refer (in the pure will). Since the moral law needs no application to the physical world, it needs no transcendental schema of imagination,[108] but needs only the rules of the understanding. The understanding can supply to an idea of reason the form of a law of nature, which we can term the type of a law of morality:

---

[108] It needs no procedure (*pace* Silber, J.R. 1975).

> But to the law of freedom (which is a causality not sensuously conditioned), and consequently to the concept of the absolutely good, no intuition an hence no schema can be supplied for the purpose of applying it in concreto. Thus the moral law has no other cognitive faculty to mediate its application to objects of nature than the understanding (not the imagination); and the understanding can supply to an idea of reason not a schema of sensibility but a law. This law, as one which can be exhibited in concreto in objects of the senses, is a natural law. But this natural law can, for the purpose of judgment, be used only in its formal aspect, and it may, therefore, be called the type of the moral law. (CPR 69f.)

That is, the law of morality (a necessary rule of practical rationality) is being given a necessary rule of the understanding (a possible law of nature) in order to be applied.[109] The rule of judgment under laws of pure reason is: "Ask yourself whether, if the action which you propose

---

[109] I will not address the question, according to what principle the will has to decide whether the maxim is a moral maxim. Three different interpretations are usually cited. The Logical Contradiction Interpretation states that universalization of the maxim makes the practice inconceivable; the Teleological Interpretation says that the action would not work in a teleological system, and the Practical Contradiction Interpretation states that the purpose would be thwarted if the practice could not be established. (cf. Korsgaard (1985)). Since it seems plausible to assume that Kant view nature as if it were a teleological system, it seems most likely that he would favor the Teleological Interpretation. I think, for the reasons given, that he would not accept the pragmatic contradiction line Korsgaard (1985) advocates. This is, in my view, clear from FMM 436, note: "Teleology considers nature as a realm of ends; morals regards a possible realm of ends as a realm of nature. In the former the realm of ends is a theoretical idea for the explanation of what actually is, in the latter is a practical idea for bringing about that which is not actually real but which can become real ...."

should take place be a law of nature of which you yourself were a part, you could regard it as possible through your will." In Kant's ethical system, moral principles are mediated by the categories of the understanding, which in turn are rules.[110] It goes far beyond the scope of this chapter and this essay to examine the consequences and the precise details of Kant's view regarding laws of nature, since it would require an analysis not only of the first *Critique*, but also of the third *Critique*, in particular the quasi teleological theory of nature developed there. May it be sufficient to summarize Kant as maintaining that laws of morality are abstract and independent of their application, that they are 'applied' by means of a set of further, a priori, formal rules of the understanding. Following moral rules is applying moral principles to (and only to) the categories of the understanding, in so far as they give the form of a natural law. They are not to be schematized, that is, no intuition is suitable for such concepts. The moral rules are then, according to Kant, independent of their application, and independent of the human will, as well as

---

[110] This is the reason it is valid for all rational creatures who may, for example, have very different forms of intuition. O'Neill (1989, p. 160) defends Kant against MacIntyre's charge (MacIntyre 1988, p. 145) of formalism (see footnote 93) by stating that "there can be no algorithm for applying rules or principles in particular situations. (If there were, we would have to apply an intention of rules for applying rules ...)", an argument Kant uses in CPR 172. Principles, however, do not apply to situations, they apply to categories. See also footnote 99.

universal in their application without humans ever applying any of them. Since they are abstract, they have the same status as the Fregean and model theoretic principles of language. With this is mind, I now will turn to rule-utilitarianism, another type of theory of morality as a system of rules.

## Rule-Utilitarianism

J.J.C. Smart has stated that there are two sub-varieties of rule-utilitarianism according to whether one construes 'rule' as 'actual rule' or 'possible rule'. The latter is like Kant's theory

> if it is permissible to interpret Kant's principle 'Act only on that maxim through which you at the same time will that it should become a universal law' as 'Act only on that maxim which you as a humane and benevolent person would like to see established as a universal law.'[111]

The connection between rule-utilitarianism and Kantianism is so close that the only difference seems to be that Kant tries to establish a principle which rule-utilitarianism assumes.[112] The formal properties of both systems are very similar. Hare, for example, remarks on this similarity by saying that

---

[111] Smart and Williams (1973) p. 9.
[112] Cf. Korsgaard (1985), p. 26.

> the formal, logical properties of the moral words, the understanding of which we owe above all to Kant, yield a system of moral reasoning whose conclusions have a content identical with that of a certain kind of utilitarianism. (Hare, 1981, p. 4; emph. added)

The idea of a universal law or rule is at the center of the type of rule-utilitarianism I will briefly outline. Kant never thought of morality but as a system of rules and it is merely their origin and the type of rules that he considers as making a difference.

> If the will seeks the law which is to determine it anywhere else than in the fitness of its maxims to its own universal legislation ... heteronomy always results.
>
> ...
>
> All principles which can be taken in this point of view are either empirical or rational. The former, drawn from the principle of happiness ... are not at all suited to serve as the basis of moral laws. (CPR 441/42)

This connection between Kant and contemporary philosophy of language going back to Frege also holds for some forms of rule-utilitarianism and philosophy of language, precisely because the theory of language is such a paradigm of philosophical analysis in this century.

> I try to base myself ... entirely on the formal properties of the moral concepts as revealed by the logical study of moral language ... (Hare, 1967, p. 25)

In the case of Hare, but not restricted to his case, the formal principles of the moral concepts completely determine

the meaning of the terms, a move familiar from the type of formal semantics I discussed in the first chapter:

> In the case of some word, for example 'all', their logical properties exhaust their meaning. We know everything we could know about the meaning of 'all' if we know what inferences we can validly make from and to propositions containing it; ...
>
> ... [T]he more general words used in moral discourse ... are in this respect like 'all' ... (Hare, 1981, p. 3)

But the formalism and rigorism of rule-utilitarianism is not restricted to those kinds of rule-utilitarianism that accepts the paradigm of formal semantics for moral theory; again the idea of rigorism and formalism is brought to moral philosophy regardless of whether the phenomena can really accommodate such a paradigm. Without any appeal to formal semantics, Harsanyi remarks that

> for [rule-utilitarianism], a moral decision problem is a maximization problem involving maximization of ... social utility. [In contrast to act-utilitarianism, rule-utilitarianism] imposes very different mathematical constraints on this maximization problem. [A] rule-utilitarian moral agent will regard not only his own strategy but also strategies of all other moral agents as variables to be determined during the maximization process so as to maximize social utility. For him this maximization process is subject to two mathematical constraints: one is that the strategies to be chosen for all rule-utilitarianism agents must be identical ...; the other is that the strategies of all non-

> utilitarians are given ... (Harsanyi, 1977, p.57)[113]

Given the formal similarities between the element of Kant's system and some form of rule-utilitarianism, I can thus be very brief with the description of rule-utilitarianism. Rule-utilitarianism, in some of its forms at least, is a system that prescribes a set of ideal rules an agent has to follow in order to do the right thing. Rule-utilitarianism is therefore contrasted with act-uilitarianism. The latter is a theory that holds that a good moral theory be a prescription of what individuals should do as individuals:

> An act is right if and only if it has at least as good consequences under the circumstances as any other act open to the agent. (Regan, 1980, p. 12)

Act-utilitarianism, since it regards only individuals, has a well known problem: in situations of cooperation, the theory seems unable to provide for precepts as to what an agent is to do since another agent's reasons for action are not part of the considerations for an individual's action.[114] This inability of act-utilitarianism to give

---

[113] Thus, the moral agent, if rational, will "always choose that particular social system that would maximize his expected utility, that is, the quantity ... representing the arithmetic mean of all individual utility levels in society." (loc. cit. p. 46)

[114] The point of the problem is that the act-utilitarianism cannot tell which act e.g. A should perform, until we specify how B acts. This is <u>not</u> the so-called Prisoner's

(continued...)

agents precepts in situations of cooperation can be alleviated, if the agents knew that there were certain moral rules that determine which types of action has to be performed. Accordingly, rule-utilitarianism is, loosely formulated, the view that

> the rightness or wrongness of an action is to be judged by the goodness and badness of the consequences of a rule that everyone should perform the action in like circumstances.[115]

Kant would call the rules of such a theory hypothetical imperatives, in that they prescribe something that is possibly prudent for the individual, but not ethical.[116] According to rule-utilitarianism a moral theory should tell us what everyone should do. It ought to be the case that the pattern of behavior that the rules or set of rules prescribe achieve the best possible consequences collectively, not individually. Consider a more elaborate

---

[114](...continued)
dilemma. For the application of this issue of cooperation to Act Utilitarianism, see Gibbard, A. (1965); see also: Sobel (1968) and Barnes (1971). This does not make Act Utilitarianism a version of egoism. Act Utilitarianism simply fails to be able to account for prescriptions that take other person's desires into account, if they are not communicated. Egoism, on the other hand is a system that tells you to ignore others unless they benefit your interests.

[115] Smart and Williams (1973), p. 9.

[116] Utilitarianism is thus a reductive theory of morality, in that it reduces the notion of right to the notion of good (which is most often replaced by 'happiness'). The rigidity of its rules (more and more sophisticated accounts of which are developed in game and decision theories) seems as odds with the rather loose notion of 'happiness' or 'social utility' the rules are to maximize.

formulation of rule-utilitarianism:

> An act is right if and only if it conforms with an ideal set of rules; an ideal set of rules is any set of rules such that if everyone always did, from among the things he could do, what conformed with that set of rules, then at least as much good would be produced as by everyone's always conforming with any other set of rules.[117]

It may be reasonably assumed that Barnes is also talking about an ideal set of rules, and thus of rule-utilitarianism as an ideal theory with an ideal set of rules that adjudicate and evaluate actions. In order for an agent to perform an action that is morally right she has to take into consideration not only her well-being and the consequences of her actions toward it, but she has to take into account the well-being of all agents in the social universe. That is, in order to do the right thing, rule-utilitarianism asks that we take into account an infinite amount of information that needs to be computed before we can call an act right or wrong. This is the view, for example, of Richard Brandt:

> One advantage to abstract rules is increased applicability — perhaps world-wide.[118]

This is an off-hand, but still instructive remark, because it seems to be the intuition philosophers have about rules. Brandt thinks that the rules apply world-wide, while Kant thinks that they are valid for all rational beings.

---

[117] Barnes (1971) p. 57.
[118] Brandt (1979) p. 290.

Platonists may even think that they apply for everything in the universe. In the next chapter, I will describe the rule-following paradox Kripke develops. Kripke's paradox, stating that there is no fact to the matter as to which rules we follow since any action or series of action can be made to agree with an infinitely large number of rules throws some doubt as to whether rules determine their application, as rule utilitarianism as well as Kant's form of deontology assume.

## The Rule-Following Paradox: Kripke's Meaning Skepticism

### Skepticism and Meaning

In his book *Wittgenstein On Rules and Private Language*, appreciated as presenting a string of arguments which "leads quite naturally to a discussion of many of the most significant issues occupying philosophers today"[119] Kripke, using an unusual methodological mix of exegesis and exposition of an argument he himself is reluctant to endorse, presents a skeptical paradox about meaning. Despite this rather unusual philosophical method, it has generated a considerable amount of interest in the philosophical community. I believe this is the case since it generates a semantic paradox from assumptions which are shared by those that subscribe to Model M. Kripke presents the skeptical thesis in terms of meaning as well as in terms of rule-following. The skeptical thesis can thus be extended to the moral theories I presented in Chapter Two, which state that an action or a will to an action is moral just in case it follows certain rules.

Kripke claims that theories of meaning are unable to solve a so-called meaning paradox which states that

> there is no fact about me that distinguishes between my meaning [something] and my meaning nothing at all. (Kripke p. 21.)

---

[119] Boghossian, 1989, p.508.

This is quite obviously a startling thesis, and it would be serious indeed if one were unable to show that and why this thesis is wrong.[120] A critic of a theory questions the validity and soundness of aspects of a theory, thereby not necessarily rejecting all of the theory under investigation. A skeptic, however, typically not only dismisses a single theory, but a whole range of theories, denying that there is a subject matter — a domain — about which anything can be known. This, in itself, seems as strong and absolute a claim as can be found in the targets of skepticism, *viz.*, philosophical theories of e.g. Aristotle and Descartes, or, in this case, contemporary philosophical models - pictures - of meaning. The fact that skepticism makes absolute claims about a particular area - be it epistemology or semantics - is itself an interesting even if only circumstantial piece of evidence that philosophers and skeptics alike take ordinary notions and make them into 'philosophical superlatives'.[121] The Ancient Greek *σκεπτομαι*, from which

---

[120] There is some precedent in the history of philosophy of this paradoxical view that there is no meaning. Cratylus, a follower of Heraclitus, held the view that communication was impossible, since the speaker, hearer and the words exchanged were constantly changing, never the same (according to the Heraclitean principle). Whatever meaning was intended by the word exchanged would be altered by the time they were heard, a view which made it pointless to speak, leaving him with wiggling his finger. Cf. Plato, *Cratylus*, 440e.

[121] The expression is Wittgenstein's (cf. PI 192).

the word *skepticism* is derived, means 'to look about', 'look carefully',[122] and someone dismissing all claims about a subject matter might be well advised to be more skeptical, more careful, about the scope and power of his arguments regarding theories and things s/he dismisses. Skepticism is not a careful investigation, it is a wholesale dismissal. Denying , for example, that there are correct theories of meaning does not warrant the conclusion that there is no meaning, and having a theory that the subject matter does not exist because a paradox can be generated is just that — a theory. Theories of meaning may very well have flaws, be wrong, but we certainly can communicate more than we can by wiggling our fingers. The radical skeptics who end up wiggling their fingers may thus turn out to be just theoreticians of a particular type: they have theories about theories of language. The aim of this chapter is thus to show that Kripke's skeptic has a theory about meaning and rule-following leading to a skepticism about meaning and rule-following, and that the skeptic's theory or picture of meaning and rule-following shares the premises essential to Model M and the moral theories presented in Chapter Two.

Skeptics — even if they usually deny that we can know something about a particular subject, or that the subject in

---

[122] The ordinary meaning of the English term has the connotation of suspending judgment (according to various dictionaries I consulted).

question exists — sometimes present theories about a subject matter as well. Hume, for example, held a form of skepticism which also provided a skeptical solution to his skeptical doubts. Hume (I summarize here Kripke's view on Hume) presents a skepticism regarding our notion of causality and induction. Finding no (causal or inductive) *nexus* between past and future events he presents a skeptical solution that allows him to 'salvage' the notion of causality from skeptical doubt:

> A skeptical solution to a skeptical philosophical problem begins [...] by conceding that the skeptic's negative assertions are unanswerable. Nevertheless, our ordinary practice of belief is justified because — contrary appearances not withstanding — it need not require the justification the skeptic has shown to be untenable. (Kripke, p. 66.)

A skeptic may thus present a constructive thesis, in which case it is not merely a rebuttal of all the available theories, that is, not only a theory about theories. Kripke thinks that Wittgenstein is a skeptic like Hume. Just as Hume offered a skeptical thesis regarding causality for which he then presented a skeptical solution, Wittgenstein - according to Kripke - offered a skeptical solution to a skeptical thesis about meaning. But the need for a skeptical solution depends on the success of the skeptical thesis. In what follows I will present Kripke's argument for his skepticism. In the next chapter I will try to show that such a skepticism rests on premises that are due to a

picture of language and rule-following that is the subject of Wittgenstein's *Philosophical Investigations*. I will therefore not discuss Kripke's so-called skeptical solution to the paradox he offers (on behalf of Wittgenstein). Instead, I will only present Kripke's paradox, and his reasoning for it. Once the paradox and the extension of the paradox to moral philosophy has been presented I will show that there may be a Wittgensteinian dissolution of the assumptions that are essential to Model M and thus to Kripke's meaning skepticism. In the *Philosophical Investigations*, Wittgenstein is concerned with trying to dissuade the philosopher and skeptic alike from looking at rules and meaning in a particular way. Wittgenstein offers no solution (especially no skeptical solution) to the paradox, but is primarily concerned with trying to change the approach philosophers take to language and rule-following. I will thus not present the skeptical solution Kripke offers since I intend to show that Kripke's meaning skepticism is due to a picture of rules and rule-following which I will investigate, and reject, in the last chapter,

There are several strategies available to counter meaning skepticism. One strategy is to show that the premises of the skeptical thesis don't entail the skeptical conclusion. One might, for example, maintain that no facts

about a speaker are needed in order to distinguish between different meanings.[123] Another possible strategy to rebut the skeptic is to show that the skeptic has overlooked a fact which does establish meaning.[124]

There may, however, be yet another strategy — a strategy I will pursue in the next chapter. This strategy tries to show that the skeptic's argument rests on the fact that the concepts involved in generating the skeptical hypothesis are misrepresented. The skeptic — so my suggestion, which follows Wittgenstein's *Philosophical Investigations* — misrepresents certain terms in order to generate the skeptical thesis. This second strategy will not show that there is something that meaning consists in or that there is a fact that guarantees that there are rules or rule-following. On the contrary, it will show that the thesis "There is no meaning" or "There is no rule-following" misrepresents the meaning of 'meaning' and 'rule'. Showing this does not require a refutation of skepticism. Instead, the strategy reminds us that skepticism is a philosophical thesis — and that the terms that are central to Kripke's skepticism — rule and meaning — are misrepresented by

---

[123] A case in point is Schiffer (1987), p. 176f. Notice that Schiffer (1987) does not endorse a theory of meaning; on the contrary, he thinks that there is no viable theory available.

[124] This strategy is pursued by — among others — J.J. Katz, C. McGinn and J. Fodor. I will discuss their attempts to escape the paradox (and their endorsement of Model M) below.

theorists and skeptics alike.

Kripke writes that "in my own way, I will attempt to develop the 'paradox' in question"[125] which Kripke claims Wittgenstein presents in the arguments leading up to Section 201 of the *Philosophical Investigations*.[126] In that section, Wittgenstein makes the following remark:

> 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. (*PI* 201)

Kripke's skeptic presents the paradox by introducing a skeptical interlocutor. The paradox (and skeptical thesis) is then presented in two stages. First the skeptic presents an interim, temporal version of meaning skepticism which is a skeptical attack on the *nexus* between my past and present meaning of 'addition'. This skepticism then becomes skepticism about meaning in general. Several proto-theoretical considerations — and later considerations that amount to philosophical theories of meaning and language all of which are equivalent to Model M — are challenged to rebut the skeptical hypothesis that there is no meaning. Kripke maintains that all theories and proto-

---

[125] Kripke, p. 7.
[126] Strictly speaking, Kripke does not attribute the skeptical position to Wittgenstein.

theoretical considerations are unable to answer the skeptical challenge to produce a fact that guarantees meaning, so that the skeptic can maintain his thesis which concludes that, since there is no fact about me that establishes my meaning anything, there is no meaning at all.

The skeptic challenges us to answer the following question: What is the fact that guarantees that I, in the past, followed this (rather than that) particular rule, or that I , by using a particular term, refer to this rule with a particular term. This connection between denoting, naming, rules, and meaning — a connection that I described in chapter one and called Model M — is introduced by Kripke right at the beginning of his skeptical enterprise, where he states that

> I, like almost all English speakers, use the word 'plus' and the symbol '+' to denote a well-known mathematical function, addition. The function is defined for all pairs of positive integers. By means of my external symbolic representation and my internal, mental representation, I 'grasp' the rule for addition.[127]

---

[127] Kripke, p. 7. This view of meaning, so Kripke, applies to "all meaningful uses of language".

Adding Skepticism

Learning how to add, according to this view, is 'to grasp' a rule that can be applied to positive integers.[128] Once I grasp the rule I can externally represent it, memorize it, have an internal mental representation of it so that I can have an intention regarding the use of this rule or give myself directions as to what to do when I apply the rule.[129] The point of learning the rule and applying the rule is that my intentions - directed by what function I mean — determine a unique answer for indefinitely many new cases. This is the point of a rule: it serves as a type of recipe to achieve certain results. Once I grasped the rule of *addition*, it will allow me to add numbers, and add them correctly. The point that rules get applied to new cases can also be put differently: there would be no need for rules if there were no new cases. Without new cases, the rule could be replaced by a list or table.[130]

---

[128] This is Kripke's definition. It seems to be incorrect, since addition is not only defined for positive integers. But this won't matter.
[129] The various terms ... *grasp, intention, directions, mental representation* etc. ... are all Kripke's; cf. Kripke, p. 5.
[130] Kripke assumes that it is simply unthinkable that there could be a rule for which there could be no new cases. This will be part of the picture of rules Wittgenstein is investigating.

> One point is crucial to my 'grasp' of this rule. Although I myself have computed only finitely many sums in the past, the rule determines my answer for indefinitely many new sums that I have never previously considered. (Kripke, p. 7f.)

That is, for my actual use, a tabulation of pairings would be sufficient. But since I have performed only a limited number of additions Kripke can pick an arbitrary pair of numbers and stipulate that I have never added this pair of numbers before. We are to suppose that I never added the numbers '68' and '57'. Being asked to perform this addition I come up with the number 125. My response

> is correct both in the arithmetical sense that 125 is the sum of 68 and 57, and in the metalinguistic sense that 'plus', as I intended to use that word in the past, denoted the function which, when applied to the numbers I called '68' and '57', yields the value 125.
>
> Now suppose ... [a] skeptic questions my certainty about my answer in what I have just called the 'metalinguistic' sense. Perhaps, he suggests, as I used the term 'plus' in the past, the answer I intended for '68+57' should have been '5'! (Kripke, p. 8.)

It may be helpful to first state what the skeptic at this initial state of the challenge does <u>not</u> dispute or doubt.[131] The skeptic does not dispute the correctness of

---

[131] Strictly speaking, the skeptic does not doubt anything; the skeptic merely presents a hypothesis, which he challenges us to refute. Similarly, one could reconstruct Descartes' skeptic as someone who presents an hypothesis, that all my claims of knowledge could be made under the influence of the Evil Genius. Lacking arguments to refute this hypothesis would then lead to the claim that there is (continued...)

my computation, nor the accuracy of my memory,[132] nor does he dispute the fact that there is a mathematical function *addition*.[133] He does not claim that I 'slipped up' (or might have slipped up) in my calculations — Kripke's skeptic is not a Humean or Cartesian skeptic. The skeptic merely doubts that I can produce a <u>fact</u> that guarantees and justifies[134] me to maintain that when I use the term 'add' now I mean the same thing - denote the same function - as I did when I used that term in the <u>past</u>. The doubt is presented by using a dramatic device — claiming that my past use of the term is compatible with the hypothesis that I, in the past, when using 'add', meant some different function, *quaddition*, defined as yielding 5 for the number pair 68 and 57 and otherwise yielding the values of the computations I had termed *additions*. In the past, so the hypothesis that

---

[131] (...continued)
no fact that allows me to state that I know something, which then would lead to the claim that there is no knowledge.

[132] Kripke, p. 11.

[133] Kripke, p. 12.

[134] The first is what Kripke calls the metaphysical aspect, the second the epistemic aspect of the challenge (Kripke, 11; 21). Since Kripke's skeptic accepts only a (metaphysical) fact to establish meaning, the distinction appears to be doing no real work. If justification in the regular sense of justification would be all right, the skeptic would have to be willing to take my word for it — I meant plus, no doubt about it. Of course, the skeptic will not accept such a justification; but the only reason for his rejection is that such a justification fails to present a fact — which means that the metaphysics of meaning he presupposes determines what kind of justification he is willing to accept. There is thus no real distinction between the justificatory and the metaphysical aspect in Kripke.

he challenges me to refute, I meant 'quadd' when using 'add', where 'quadding' refers to computing the function *quaddition* that yields 5 for the number pair 68 and 57 - for which I ought[135] to get this result if I intend to compute that function - and otherwise yields just those very sums I actually computed in the past for - what I termed (and still term) - *addition*.

Initially, that is, the skeptic does not doubt that terms I use now or that I have used in the past have meaning, that I can grasp a rule; he also does not deny that grasping a rule and/or the meaning of a term justifies me and obliges me to use the term or follow a rule in a particular way (<u>this</u> way). The skeptic does not yet challenge me to show that my terms now have meaning, but he challenges me to show which fact guarantees that the use of the term 'addition' denotes — means — the same thing, refers to the same function as in the past. The skeptic thus doubts the <u>diachronic</u> identity of a term's meaning - the *nexus* between my use of a term in the past, the use I make of it now, and the meaning of the term in the past and now. The skeptic suggests that my answer 'ought to have been' 5 and that he can successfully refer to the function of *quaddition*, which in turn *requires* that I perform *this*

---

[135] The skeptic does not deny that, if I add or 'quadd', I <u>ought</u> to achieve certain results. See below.

computation. This 'local' skepticism will later be expanded to become meaning skepticism, but the skeptic, of course, cannot start with denying that there is meaning, since that position would suffer from some performative or pragmatic incoherence.

The skeptic initially also does not deny that there is a nexus between my present and past use, and the meaning of a term. In order to show that the use of the term 'quaddition' is compatible with my past use of 'addition', he must assume that I can understand the distinction between 'quaddition' and 'addition' from my present vantage point. That is, while the connection between past and present use is in question, the connection between present and past is not (yet) in question.

## Meaning Skepticism

At first it may appear as if the skeptic thinks that rule-following is a concept that is especially relevant to mathematical functions. But the concept of rule-following applies to any term whatever:

> I think that I have learned the term 'table' in such a way that it will apply to indefinitely many future items. So I can apply the term to a new situation, say when I enter the Eiffel Tower for the first time and see a table at the base. Can I answer a septic who supposes that by 'table' in the past I meant *tabair*, where 'tabair' is anything that is a table not found at the base of the Eiffel Tower, or a chair

> found there? Did I think explicitly of the Eiffel Tower when I first 'grasped the concept of' a table, gave myself directions for what I meant by 'table'? And even if I did think of the Tower, cannot any directions I gave myself mentioning it be reinterpreted compatibly with the skeptic's hypothesis? (Kripke, p. 19.)

According to the skeptic, when I grasp the concept *chair, table, morally corrupt*, or *duty*, I grasp some rule that applies to indefinitely many new situations. So, when I am asked whether this over there is a table (and I answer correctly that it is) my response is

> correct — in the metalinguistic sense that ['book'], as I intended to use that word in the past, denoted the [rule] which, when applied to the [situation when I enter the Eiffel Tower and see a chair at the entrance, 'yields' chair].[136]

The first step of the skeptical challenge requires that I show that the past use of a term is use of the same rule or term as the one I made use of today, or that the intentions that I had when I used the term in the past are intentions regarding the same function, thing, or concept that I use now. The skeptic challenges me to produce a <u>fact</u> that <u>justifies</u> me in holding that the use of a term I made in the past is use of the same term I make use of now. The link between uses is thus due to a putative fact that shows

---

[136] Cf. Kripke, p.8 quoted above page 113. The skeptic shares this assumption with Model M and, of course, Kant, who thinks that principles, moral commands, and moral concepts are rules.

that the same rule is referenced in both uses. Thus, having computed the sum, or having pointed to a chair, the skeptic questions my certainty about my answer regarding the use of the term 'add' (or 'chair'). The skeptic suggests that, as I used the term addition in the past, the answer to '68+57' should have been '5'. Obviously, this is crazy, but the skeptic elaborates: As a matter of pure historical fact, I never added the two numbers '68' and '57'. When I am asked to perform the addition, I intend to perform a particular arithmetic procedure (adding) which I previously grasped. But the question is, which rule I previously meant. All use that I made of the term 'add' to refer to a rule in the past is compatible with a response '5' - an intention to perform quaddition - since quaddition yields exactly this result, and addition and quaddition are coextensive for all uses prior to the notorious '68 + 57':

$$x\_y = x+y, \text{ if } x, y < 57$$
$$=5 \text{ otherwise.}$$

The point about the skeptical hypothesis is thus not so much that it is wrong, but that the skeptic claims that there is no a priori, purely logical way to defeat it. Since we can define a function of quaddition, just as we can define a function of addition, a fact about me (including my

mental state)[137] is needed that shows that the skeptic is wrong in assuming that there is nothing about me that guarantees that I denoted *addition*, rather than *quaddition* in the past. This searched-for fact must somehow show that I am justified in claiming that in the past I used 'add' to refer to 'addition', not to 'quaddition'. That is, I need a fact that allows me to state that I am correct in saying that my present use conforms to my past use:

> If (the skeptical hypothesis) is false, there must be some fact about my past usage that can be cited to refute it. (Kripke, p. 9)

What facts justify my claim that my past use was use of a term referring to this function rather than that function?[138] Kripke cites some pre-philosophical, rudimentary justifications and allusions to putative facts about the speaker's mental state, facts that one might cite to justify meaning one function or rule rather than another. This apparently mimics Wittgenstein's own manner of using an

---

[137] Kripke, p. 11; Schiffer (1987), p. 176f writes that "from the inexistence of " Fregean" facts that make it the case that [someone] means addition ... it just does not follow that he does not mean addition ...." But the missing of Fregean facts is not the problem; what is missing are the facts about the speaker. Fregean facts are mind and language independent entities, and they cannot constitute "a fact about a speaker". Introspection, i.e. a purely phenomenological procedure, would thus be fine, if it could produce a fact of the required sort.

[138] One might argue that the skeptic suspends the meaning of the term 'intention', because a very natural response to the skeptic would be that I know what function I meant because I never intended to follow the quaddition rule. See the discussion of intentions below.

interlocutor who need not be aware of more sophisticated philosophical theories, but introduces certain pictures of what s/he might think would answer the skeptical challenge, a germ which may lead to the more "advanced" versions of such pre-theoretic moves.

## A First Defense against Adding Skepticism

There are three different initial attempts to refute the skeptical hypothesis:

*I followed directions*

I gave myself directions to compute 128, not 5, or I gave myself directions to mean table, not *tabair*. The skeptic will not accept such a justification, since *ex hypothesi* I have given myself no such directions for the particular pair of numbers, and I really had no idea that I would come across a table at the base of the Eiffel Tower. The type of direction in question is an explicit command such as "When asked to add 68 and 57, respond with 125"; equivalent stories have to be told about tables, chairs, duties, and other moral concepts. It is the point of a rule that it is to be applied to new cases, and I cannot give myself directions for all the cases in which I apply the rule.[139] This would conflate the notion of a rule with the notion of a list. Since I did not give myself directions, nor could have given myself directions given the meaning of *rule,*[140] this attempt to refute the skeptical hypothesis misses the point.

*This has always been done this way*

When I computed the values for "65 + 57", I did the same thing as always. This attempt to rebut the skeptic fails

---

[139] Again, the metaphor of 'give myself directions' is Kripke's. Since a direction is a rule, it comes close to the idea of giving myself a rule for applying a rule. I will discuss this issue in the next chapter.
[140] Rules have (also) to be applied to new cases.

for similar reasons. Since my use of the term 'plus' is compatible with the functions 'addition' and 'quaddition', and there is no independent fact I can appeal to when I refer to either function, the appeal to sameness fails to establish which function it was that I computed in the past.[141] The notion of 'sameness' requires that I identify the item first — but the identity of the rule meant is precisely at issue.

---

[141] There is an additional problem about the notion of same, which may be interpreted as 'quame', where 'quame' means same under all circumstances other than where functions of 'addition' and 'quaddition' are involved.

*I followed standard procedure*

When I grasped the rule I grasped a certain technique. The technique is a procedure that allows me, for example, to translate every natural number into a string of tick marks; once this is accomplished I join the two groups of tick marks representing the two natural numbers to be added, and count out the whole group.[142] The skeptic does not have to accept such a layman's version of a recursive definition of 'adding', since the technique makes essential use of the notion of *count*. The same reasons that fail to establish that there is a fact of the matter as to which function I applied the term 'add', also apply to the function of 'count'. I 'counted' only finitely many items, and maybe I was really 'quounting', i.e., using a technique that is equivalent to the function *count* up to 124, and thereafter yields 5. This type of proto-theoretical attempt of a reductive definition of addition to counting, deriving a function from a more basic algorithm, having a rule to interpret a rule, fails to refute the skeptical hypothesis. No matter what reductive definition of an algorithmic process I will produce, there will be some algorithm, some rule at the bottom of it all, and for that rule I still do not have a fact of the matter as to what the term applied to

---

[142] This is in essence J.S. Mill's theory. See Mill (1843) p.163.

in the past.[143] This last-ditch rule, then, is still open to the skeptical hypothesis, i.e. the alternate interpretation of the rule the skeptic presented. There is thus no fact that I can appeal to that justifies that my use of the term plus cannot be interpreted to refer to the function 'quaddition'.

[143] Since none of the rules can be identified, such a reduction is not possible.

## Theories of Meaning and Meaning Skepticism

The skeptic will now extend his skepticism to cover meaning and rules in general — regardless of the time of the application. If there is no fact that can be cited that shows that my use of plus meant a particular function yesterday, then none of the terms I meant yesterday can be said to mean any particular function. And, of course, the argument does not depend on the fact that this fact was missing yesterday: the fact is missing, and therefore nothing guarantees that I mean something in particular with the use of a term; the skeptic can thus claim that there is no meaning, since my use of a term fails to designate something in particular. It does not matter what terms we use - when I grasp the meaning of natural kind terms, terms referring to mathematical functions, moral terms, moral rules, predicates of sensations or visual impression and the like, I always, so the suggestion by Kripke, grasp something that "I will apply to indefinitely many future items."[144] Meaning skepticism and rule-following skepticism applies to anything that I have to apply to a new situation.[145]

---

[144] Kripke, p. 19.

[145] Unless I gave myself specific instructions, in which case, of course, the situation is not new. But even if I give myself specific instructions, can't "the skeptic say that these directions, too, are to be interpreted in a non-standard way?" (Kripke, p. 52 note; Kripke here refers to Wittgenstein's *Remarks On the Foundations of Mathematics*, section 1, 3.)

The skeptical challenge can be extended to moral rules as well. Suppose the Kantian — as described in the second chapter — were to stipulate that pure practical reason has rules by which it determines the will. According to Kant, the rules of pure practical reason are not rules that are located in the empirical self.[146] The skeptic can reformulate his challenge for the Kantian. Thinking about a maxim that is in agreement with the rules of pure practical reason requires that I know which rule I apply. The skeptic would, just as in the case of the mathematical function of *addition*, challenge the Kantian to show which fact guarantees which rule it is that is applied, or which fact guarantees that my maxim refers to the same rule by which pure practical reason determines the will. Since the rules are a *type of law* of nature which applies to infinitely many situations one can never establish which rule I am using. The Kantian thus runs into exactly the same skeptical problems in moral philosophy as the Model M theorist encountered for meaning and rule-following. The rule-utilitarian faces similar problems, since s/he has to show what fact guarantees which rules are applied when using a rule that is acceptable for a utilitarian. The skeptical paradox thus applies to moral philosophy just as much as it

---

[146] It is far beyond the scope of this dissertation to address the issue of the relation between the transcendental ego and the empirical one. May it suffice to state that they are different, and thus need a relation.

applies to Model M. There may be a stronger temptation, however, to dismiss the skeptic in the area of moral philosophy. One can imagine that such a dismissal would be based on the argument that the skeptic is merely having 'theoretical' doubts. But this would be too easy. The point of philosophers appealing to rules is that rules are thought to be the anchor of truth and rationality. If it turns out that the rules philosophers appeal to do not by themselves establish such anchorage, the skeptic undermines the whole *raison d'être* of relying on rules as determining an application.

I will now discuss full-blown skepticism as it challenges current theories of meaning; I will primarily discuss the attempt of Dispositionalism to rebut the skeptic. The attempts by dispositionalists to rebut the skeptical hypothesis are similar to those presented by Mentalism[147] and Platonism.[148] All three philosophical positions face two sets of problems since all are attempting to account for what Kripke describes as the infinitary

---

[147] Any type of theory that postulates mental states and/or mental processes as introspectible 'inner' contents that I can find in my mind or that God could find there if he looked into my mind (cf. Kripke, p. 50).

[148] Platonism is any type of theory that denies that rules or any type of abstract object are in any particular mind, or are the common property of all minds. They have an independent, 'objective' existence. It is the nature of such entities to contain all their instances. (See Kripke, p. 53.) Frege, for example, is a Platonist.

character of rules and meaning by presenting a fact. The facts, however, will a) always turn out to be finitary, and b) always fail to account for the normative character of meaning, i.e. they cannot explain why I should or ought to 'go on like this'. The lessons learned from the skeptical arguments against Dispositionalism can thus be applied to Mentalism and Platonism.

Dispositionalism

Wittgenstein has often been thought to have a dispositionalist theory of meaning and mind himself. The exegetical merit of Kripke's book on Wittgenstein may lie in the fact that it discourages such a reading of Wittgenstein. How does Kripke proceed to show that Dispositionalism cannot answer the skeptical challenge? The dispositionalist gives the following analysis of a disposition to answer a question about the application of a function:

> The referent $\varphi$ of 'f' is that unique binary function $\varphi$ such that I am disposed, if queried about 'f(m,n)', where 'm' and 'n' are numerals denoting particular numbers *m* and *n*, to reply 'p' where 'p' is a numeral denoting $\varphi$ (m,n). (Kripke, p. 26.)

The point of this analysis is that the disposition is designed to bridge the gap between past and present use. The dispositionalist states that the difference between meaning the different functions referred to with 'plus' and 'quus' consists in the fact that for each of the different

function I have a disposition to apply the terms in different ways, a disposition that exists at any given time; thus, the past and present use of the term 'addition' mean the same thing. If I were asked what function I meant, I would always respond in a particular way. That is, there is no difference between my responding to the question yesterday and today, since I have the same disposition today as I had yesterday. It is true that, when meaning the different functions, I happen to be in the same mental state, that is, there is no occurrent mental fact[149] that differentiates between 'quus' and 'plus'. But there is a dispositional fact, a fact about my disposition to respond one way rather than another way.

There are three distinct problems for the dispositionalist solution, and I will address these issues in turn.

1.
It misses the skeptical point, that the rule cannot be identified for which there should be a disposition;
2.
It cannot give a justification for meaning and rule-following;
3.
It founders on the problem of the infinitary character of reference.

---

[149] See Kripke, p. 22; the Mentalist will, of course, take recourse to such occurrent mental facts.

*1) Dispositionalism misses the point*

The skeptic is asking for a fact that establishes that I mean one thing rather than another, given that I have used the term 'add' in only a limited number of computations. If I offer a disposition as the required fact, I am not only speaking about present dispositions, but also about previous dispositions - those I had in the past. Dispositions change. The skeptical challenge is thus not at all addressed by giving a dispositionalist answer. Citing a disposition as a putative fact that guarantees that I am 'going on' in a particular way — following rule rather than another — would be a hypothesis about a disposition. It is exactly at issue to show whether the disposition was one to add or to 'quadd'. All my past answers can be construed to have been either disposition. If I were to talk about my present dispositions and maintained that it is a disposition to follow this rather than that rule, then there is no reason to think that this is anything but a stipulative definition, that whatever I do is the disposition XYZ. The dispositionalist still needs to show that the disposition he had yesterday is the same disposition he had today, but since the disposition is only exhibited in a limited number of cases, it cannot be established what function it was a disposition for. The dispositionalist cannot show that his disposition for 68 + 57 would have been to answer '125'.

### *2) Disposition and Justification*

A dispositional analysis of the use of terms is a descriptive account of a system's input/output. This account is insufficient for showing that I have some justification for "going on like this" - something that allows me to say that, if I meant this function, I ought to go like this. This problem may also be couched in terms of explanation and justification: maybe my dispositions will explain why I went on this way, but they do not justify me, show why I ought to respond in a particular way. Dispositions thus explain how I have reacted, and may refer to some fact about me that explains this reaction. But they cannot explain why a certain output is the correct output. In order to have a justification, I need a fact that tells me how I ought to continue, but a dispositional analysis of meaning merely indicates how I actually do continue.

### *3) Finiteness*

The gist of the skeptical hypothesis is this: the finitely many uses of a term do not specify which of the infinitely many functions that yield the values I actually calculated were denoted by my use of the term. The basic objection to a dispositionalist analysis is accordingly that all my dispositions are only exhibited in finitely many cases, and are themselves only a finite state. This finite state fails to identify a function uniquely because of the fact that the function has infinitely many values for infinitely many arguments. The disposition is a finite state, and therefore not sufficient to be a disposition towards infinitely many number pairs and values.[150] In essence, appealing to a finite state to explain reference to infinitely many values is merely repeating the question as an answer:

> How can a finitely long record of uses allow us to establish reference to a unique function with

---

[150] Historically, this is a very influential argument. Descartes, in the *Meditations,* argues that the idea (meaning) of 'wax' cannot be gotten from the imagination, since the imagination is finite, while the wax can have infinitely many shapes. From this argument it is a short step to arguing that the idea of wax is innate.

> I imagine ... [the piece of wax] admits of an infinitude of ... changes, and I nevertheless do not know how to compass the infinitude by my imagination, and consequently this conception which I have of the wax is not brought about by the faculty of imagination. (Descartes (1641) p. 155.)

> infinitely many values? By being a finite state that refers to an infinite range of values.

The dispositionalist might try a comeback by adding "ideal" dispositions, or that, *ceteris paribus*, I would respond, given any arbitrary pair, this way, not that way.[151] The skeptic will repeat that, since it is in question which function I originally meant, there is nothing an ideal disposition will be able to accomplish. An ideal disposition might be given the meaning "if I were to be given an infinite ability, memory, lifespan etc. to follow my intentions....", and here it becomes obvious that the skeptic's challenge has not been answered at all: it is precisely at issue which function I intended to compute that the dispositionalist cannot answer:

> It presupposes a prior notion of my having an intention to mean one function rather than another by '+'. (Kripke, p. 28.)

As a consequence, there is also no non-circular way in which we can exclude errors from dispositions, since the notion of an error presupposes that we have identified the function that is the function meant. In the end, all the arguments against the skeptic founder on this issue. To show this, consider a type of argument that uses a machine as an analogy. The machine analogy is used as a version of the dispositional, or causal theories of reference

---

[151] Cf. Fodor (1990)

(rule-following). Causal theories of meaning are dispositional theories of meaning, which traditionally have trouble explaining reference to mathematical objects because of their infinitary nature. Causal theorists are aware of the relevance of Kripke's skeptic. Kripke argues:

> Even if we were idealized as finite automata, a dispositional theory would yield unacceptable results.
>
> Suppose we idealized even further and considered a Turing machine which has a tape to use which is infinite in both directions. [...] [I]t is well known that there are many functions we can define explicitly that can be computed by no such machine. A crude dispositional theory would attribute to us a non-standard interpretation (or no interpretation at all) for any such function. (Kripke, 1982, p. 36 note 24; also compare p. 26, note 20.)

This is an argument with which current causal theories of meaning have great difficulty. Fodor, for example, simply misreads the argument:

> All we need to know is that, if we did have unbounded memory, then ceteris paribus, we would be able to compute the value of $m + n$ for arbitrary m and n. (Fodor, 1990, p. 95)0

This misses the reach and depth of the argument Kripke presents; Kripke explicitly states that there are mathematical functions that cannot be computed even by a machine with "unbounded memory" (whatever that may mean). This leaves Fodor, and causal reference theory in general, without an argument against the skeptic.

Dummett uses the following argument when he objects to Wittgenstein's philosophy of mathematics:

> A machine can follow this rule, whence does a human being gain a freedom of choice in this matter which a machine does not posses.[152]

'Machine' can be interpreted in two different ways. A machine is a mechanism; and in this case, the argument that dispositions cannot justify my following a rule apply to the machine as well. Given a certain input, the machine is in a particular state, which fails to give a justification for being in this state.[153] If we, on the other hand, understand 'machine' as 'machine program', then the application of the skeptical problem is straightforward. Since the program will contain symbols for functions, for example "+", the symbols can be interpreted in the quus-like manner.[154] The machine analogy is thus merely a version of Dispositionalism, which fails to give an answer to the skeptic. I will thus turn to the two other metaphysical pictures, Platonism and Mentalism, discussing Mentalism first.

---

[152] Dummett (1959) p. 171f (quoted in Kripke (1982), p. 32.
[153] A version of this argument is used to reject an objection that sees meaning guaranteed by some 'mental state', or inner experience (Kripke, p.42.)
[154] The same argument can be used against the "simpler hypothesis" approach (cf. Kripke, p.38.), since we cannot say what the competing hypotheses state, given the skeptic's challenge. The application is straightforward.

Mentalism

Mentalism can - for the purpose at hand - be described as the position that assumes that there are certain introspectible experiences that provide the facts that guarantee meaning. The skeptic poses no limits on introspection, that is, there are no behaviorist limitations.[155] Mentalism amounts to a

> reduction of the notion of meaning and intentions to something else. Why not argue that "meaning addition by 'plus'" denotes an irreducible experience, with its own special *quale*, known directly to each of us by introspection? (Kripke, p. 41.)

We may even consider this state to be the ultimate, *sui generis* state that uniquely means this

> by arguing that meaning addition by 'plus' is a state even more *sui generis* than we have argued before. Perhaps it is simply a primitive state, not to be assimilated to sensations or headaches or any qualitative states, nor to be assimilated to dispositions, but a state of a unique kind of its own. (Kripke, p. 51.)

But this *sui generis je-ne-sais-quoi*[156] state inherits precisely the same problems the dispositional state has: "Such a state would have to be a finite object, contained in

---

[155] See Kripke, p. 14.
[156] "[I]t leaves the nature of this postulated primitive state — the primitive state of 'meaning addition by "plus" — completely mysterious." (Kripke, p. 51.)

our finite mind."[157] The mentalist, no matter what state, process or qualia he thinks is the fact that guarantees sameness of use faces the same arguments that refuted the dispositionalist: the finitely many cases of the previous applications do not give an effective procedure by which we can determine which function, which word was meant.

---

[157] Kripke, p. 52. The same holds for the theory proposed by McGinn (cf. McGinn (1984), esp. p. 161f). McGinn proposes that "there is a legitimate notion of 'introspection' ... [,] the kind of introspection we employ when coming to know what we intend ... " [emph. added]. McGinn misses the argument, since the skeptic argues that what we intend is precisely what a finite mental state cannot show.

Platonism

As last resort, one might consider that the skeptic can be rebutted by appealing to Realism: The mathematical realist thinks that the function of addition is an infinitely large set of ordered triples[158] that contains all its instances, is not in any particular mind, nor common to all minds. But again the same problem arises: Even if we stipulate that there is such a mind and language independent entity like meaning, we still need an idea in the individual's mental fact associated with the sign. The mental presentation, however, is a finite object, and as such the finite object would have to be correlated with the infinite mind and language independent entity. What constitutes grasping one platonic meaning rather than another? We cannot tell, if the mental idea is finite, and the platonic entity contains infinitely many applications. The idea in my mind can be interpreted as determining a quus function, rather than the plus function.

The skeptic concludes that there is no meaning, and that is impossible to follow a rule. He argues this by showing that a particular term *addition* fails to have meaning since it does not denote a particular object or function. It

---

[158] Frege would not agree with this, because functions are unsaturated (a notion I won't attempt to explain). This issue, however, can be ignored. See also Kripke, p. 53.

fails to do so since all uses of a term are compatible with infinitely many different functions, rules or meanings. The meaning of a term is thus thought to be established by a putative denotation of a rule that gets applied to new situations, and the problem, the skeptic maintains, is that we cannot specify which rule a term denotes when we use a term. The skeptic thus denies that rule-following is possible, that there is meaning. The skeptic's argument in a nutshell: Using a term is an application of rules. Infinitely many rules can be applied to the same appearances. Therefore, there is no meaning.

## Responses to Kripke's skeptic

The reaction to Kripke's skeptical argument shows that, despite their forceful criticism, contemporary philosophers accept the presuppositions of the paradox unchallenged. This is not unexpected, given that contemporary philosophy of language accepts the tenets of Model M. The responses to the skeptic by contemporary philosophers like Katz (1990), Schiffer (1987), Fodor (1987, 1990), Boghossian (1989), and McGinn (1984) do not address the essence of the skeptical point, that meaning is the essence of language understood in terms of a relation between two types of objects — an object that is the meaning fact, and one that is the object of reference. This element of the Model M is considered indisputable. I emphasize that it is not the point of this short review of Kripke's skeptic's critics that they may not have a comeback here or there. My point is merely, that there is a certain common model, Model M, of meaning and language that can be attributed to the skeptic as well as his critics. Kripke's critics are merely concerned with the issue, whether the skeptic has overlooked a type of fact that guarantees the referential relation as well as rule-following — something the skeptic doubts can be established. Objects denoted by meanings that are objects

themselves become the paradigm of meaning and reference.[159]

J.J. Katz, in a book-length discussion on meaning and current theories of language writes that on his theory

> the meaning of a word is similar to the word in being, like its morphological structure, part of its grammatical structure, but it is different in being sense structure rather than syntactic structure. Being part of sense structure, meaning will, of course, not be public features of sentences in the way the words in their surface syntax are. Rather, syntactically simple words, even in the deep syntactic structure of sentences, conceal complex meanings. [160]

There are some things in this passage that deserve to be pointed out. Firstly, Katz' theory is not reductive, but is intensionalist. This results in a quasi-circular definition of meaning. Notice that the statement

> The meaning of a word is ... sense structure. ... [S]imple words ... conceal complex meanings.

fails to illuminate the nature of meaning in the way the skeptic demands, although it still subscribes to the view that all words and language in general have a property — meaning. Further below I will show that Katz' program subscribes very much to the thesis of Model M, namely that meaning is a property of words which gets established by a

---

[159] I don't think that any of the rebuttals actually work, but this is not the target of this chapter.
[160] Katz (1990) p. 130; emphasis added.

relation between the term and an object.[161] First, however, I want to point out that Katz endorses Model M and the skeptic's presuppositions, *viz.* that 1) meaning is a universal property of language that can be uniformly explained by a single theory, that 2) meaning has compositional structure and 3) meaning is a relation between two different entities. These are the essential elements or Model M. Katz merely disagrees as to what the basic elements of meaning are (that is, what the basic elements of compositionality are, or where analysis comes to an end), and what the ontological status of meaning is. Recall that Davidson stated that

> [a]n acceptable theory should ... account for the meaning ... of every sentence by analyzing it as composed ... of elements drawn from a finite stock.[162]

For Davidson, the stock is syntactic, while Katz takes into consideration a different kind of "finite stock," different basic entities from Davidson. He claims that

---

[161] Katz gets many things right — but I think he should relativize 'senses' to situations and uses ("language games") and abandon his platonic essentialism. Words do have sense structure and the power of seeing words — data — to contain sub-syntactical structure can be seen in the database industry. If the structure is relativized to a particular language and domain (a given database) then object oriented databases have a considerable advantage over relational and hierarchical database structures, since the object data are given structure very similar to 'markerese'. Once such structure is relativized (to a language game), it's perfectly all right to claim that words have sense structure. See Kim (1990).

[162] Davidson (1970), p.56.

words or syntactic simples have further structure, "sense structure" but does not deny that meaning has compositional structure. All the disagreement is about is where analysis stops, i.e. whether words or some other entity is the basic unit of meaning. Katz's theory also subscribes to some form of Platonism. On Katz' version of Fregeanism, the speaker has to have some access to the rules that form the meaning of a sentence. This, according to Katz, happens by means of an intuition. Intuition is a psychological act, it is supposed to access the abstract rule that is mind and language independent:

> Intuition is, to be sure, a mental act, but its object is not, for that reason, something psychological. Perception is a mental act, but its objects are not mental. ... A proper realist conception of understanding differs from "the empiricist conception," since, for the realist, "apprehension by consciousness" is intuition of abstract objective objects ... .[163]

Katz agrees thus with a further tenet of Model M, namely that the meaning of a term is an object of a certain type that corresponds to a word, an object that is abstract, mind and language independent. The skeptic and Katz thus agree that meaning is a relation between entities, one of which is

---

[163] Katz, J.J. (1990) p. 160. Note, that this fails to address the skeptical point that the object in question has to be grasped ("intuited") by a finite mind. Katz struggles (unsuccessfully, I believe) to show that the paradox does not apply to his theory; compare especially note 34 on page 161 in Katz, *op. cit.*

a rule. This is why intuition does not help Katz to escape the skeptical paradox. His theory also postulates a rule-governed association between objects, a view that is often labeled as 'translational' theory of meaning. A translational theory is rule governed in the sense that is assigns meaning to objects, precisely in the manner of Model M. This issue has been first recognized by Vermaazen[164] and has been summarized by Lewis:

> Semantic markers are *symbols*; items in the vocabulary of an artificial language we may call *Semantic Markerese*. Semantic interpretation by means of them amounts merely to a translation algorithm for the object language to an auxiliary language Markerese.[165]

This theory, then, not only fails to refute the skeptical hypothesis, since the intuition cannot tell me which rule or semantic structure I meant, but Katz' presuppositions of rule-following, compositionality and denotation are identical to the one the skeptic uses in presenting the paradox. Whether the meaning-endowing relation is an interpretation by an item in a model (Davidson) a possible world (D. Lewis) or an object (Russell, Frege, Fodor, and Boghossian) or another symbol (Katz), the skeptic will still insist that we have a relation of a symbol with another item, and that this relation is underdetermined on the meaning side, since the

---

[164] Vermaazen, B (1967)
[165] Lewis (1972) p. 169.

infinitely many objects cannot be referenced by a finite object that is (supposed to be) the meaning.

Paul Boghossian who critiques Kripke's argument in an article concerned with an "assessment of the *philosophical* contributions"[166] of Kripke's book also accepts the same view of language without further ado. After presenting Kripke's argument against reductionists he summarizes:

> All the suggestions that I have seen to the effect that Kripke ignores various viable reduction bases for meaning facts seem to me to rest on misunderstanding. (B, 527)

He then favors an anti-reductionist account of meaning (which he, following Fodor, champions). This anti-reductionism again argues for a finite object with compositional structure that is to be located in a finite mind, preserving the assumptions of Model M since

> an anti-reductionism about mental content is perfectly consistent both with substantive theories about the nature of the propositional *attitudes* ... and so with theories of the *compositional structure* of mental content. (B, 549).

Similarly, Fodor thinks that content (*meaning*) is

> assigned only to the *atomic* expressions, the semantics for molecular symbols being built recursively by the sorts of techniques that are familiar from the construction of truth

---

[166] Boghossian (1989) p. 507; This will be cited as *B*, followed by the page number.

definition.[167]

And again, Colin McGinn, in another book-length discussion of the issue implicitly accepts the notion of meaning the skeptic has by criticizing Kripke:

> [I]t is easy to supply (an account of meaning): to mean addition by '+' at *t* is to associate with '+' the capacity to add at t ...[168]

And this kind of meaning, one might take it, can be supplied for all other words: to mean horse with 'horse' is to associate with 'horse' the capacity to apply 'horse' to horses, etc. This, again, is a relational view, and subject to Kripke's skeptical challenge: How does the subject know that he did not associate '+' with the capacity to quadd at t? The kind of capacity is subject to introspection. A theory about this type of introspection is not available; but, according to McGinn, this does not show that there is no such introspection.[169]

Kripke and his critics thus share certain presuppositions about language and rule-following. Kripke's critics use the same presuppositions as the skeptic about

---

[167] Fodor (1990), p. 58.
[168] McGinn (1984), p.174.
[169] See McGinn, *op. cit.*, p. 161. I think that this theory is very seriously flawed. I re-emphasize that it is not the point of this chapter to show that Kripke's skeptic's critics may not have a comeback here or there. My point is merely, that there is a certain common model, Model M, of meaning and language that can be attributed to the skeptic as well as the critics.

language and rule-following in order to show that Kripke has not considered all of the possible theories of meaning and rule-following. They rely on the picture of meaning as *denotation*, or reference to an object, a view that is essential to Model M. They also rely on abstract rules to apply to infinitely many objects and situations; rules determine the application without recourse to an actual application. This latter view about rules is also shared by the moral philosophers I discussed in the second chapter. Kripke's argument shows that, given the picture of rule-following and meaning, there is no fact about a speaker that would guarantee the uniqueness of rule-application — rules Model M and rule-oriented moral philosophy rely on. If the actual use of rules is compared with the view of rules as abstract application, then, it turns out, following a rule can always be interpreted as a following of infinitely many different rules. In the next chapter I will try to show that the rule-following paradox is a consequence of looking at rules and language in a particular way. It may be that Wittgenstein's later writings offer a method of avoiding the conclusion by investigating the shared presuppositions of Model M and rule-orineted moral philosophy.

## A Family of Cases: Wittgenstein's Reminders Regarding Rules

In the first chapter I characterized Model M's view that language has an essential property — meaning. This essential property is understood in terms of a relation between a word or other linguistic elements and an object. The linguistic elements are combined by recursively specifiable rules to generate the meaning of a sentence. In the second chapter I described two very influential positions in moral philosophy — Kant's rule-deontology and rule-utilitarianism. Both types of theory picture morality as a system of rules that determine moral evaluation. In the third chapter I chose Kripke's presentation of the rule-following paradox in order to show that picturing a system of rules to be the essence of meaning and morality may be jeopardized by the rule-following paradox which calls into question the view that rules are the most fundamental and apodeictically certain elements of language and morality.

In this chapter I will try to investigate whether there is a response to this picture of rules and rule-following in moral philosophy and the philosophy of language which allows us to bypass or dissolve the paradox. I will try to give what I want to call a deflationary account of the role of rules and rule-following. If this can be done, rules will be the diverse and quotidian things they are — things about

which one cannot say anything in general.[170] I will attempt to use a strategy similar to the one Kripke uses in his book on Wittgenstein by presenting philosophical considerations I find persuasive and which strike me as reasonably close to those that can be attributed to Wittgenstein. Like Kripke, however, I make limited exegetical claims — I do not claim that the considerations in this chapter are Wittgenstein's; the considerations merely 'strike me' when reading Wittgenstein. There are some who disagree with Kripke's interpretation of — or inspiration by — Wittgenstein. Similarly, my method of approaching the rule-following paradox may not convince everybody, or may be found to be different from Wittgenstein's. I can only hope to present considerations that will strike a chord in those that find something in Wittgenstein.

## Philosophy: A Method against Model M

Wittgenstein's interpreters often see his *Tractatus* as advocating that thought has certain necessary structures and the *Investigations* advocating that we invent those necessities. David Pears, for example, writes that philosophers

> want to know whether our thoughts necessarily

---

[170] Wittgenstein remarks that it would be pointless to give a philosophically significant definition of *rule* (cf. Wittgenstein (1979) pp. 153-4).

> have the structure that they actually have, because it is imposed on them from outside by the world, or whether necessity is always internal to a particular way of thinking—a particular language-game—which we have adopted without having been forced to do so. In the *Tractatus* he favored the first of these two suggestions, and later, when he found the second one more convincing, it seemed to him that his earlier mistake had been typical of metaphysical thinking.[171]

I don't think that Wittgenstein holds the view that we adopt language games; instead, he holds that there is no essence to language and rules. To find some methodological bearings I will present what I consider a Wittgensteinian approach to the elements of Model M. The method will be exhibited by presenting considerations regarding meaning, reference, and rule-following. One of the elements of Wittgenstein's investigation of Model M and rule-following is the connection between rules and games; I will thus also discuss the notion of a *game*.

Philosophers often argue for Model M by claiming that not all of the occurrences of the term *meaning* are semantic in nature and are therefore irrelevant from a semantic point of view.[172] In utterances like 'he didn't mean to harm you', 'he didn't mean to be late', etc. the term 'meaning' does not have any semantic implications.[173] Nevertheless, *meaning* is held to have a semantic element which has to be

---

[171] Pears (1988) p. 206
[172] Compare Devitt and Sterelny (1987) p.15.
[173] In the two examples, the term 'meaning' means 'intend'.

abstracted from some of the non-semantic uses the term has. This purified 'meaning' is the essence of language which philosophers try to 'isolate'. The paradigm of this property are names and their bearers,[174] and the view that a combination of names generates the meaning of a sentence. In an investigation that is critical of such theoretical and philosophical moves, a possible 'countermove' might try to show that a theory of meaning that somehow is the result of such a theoretical process of abstraction ignores some part of the notion of 'meaning' which are semantically relevant. This, of course, is a general description of almost all theoretical criticism that does not attempt some form of 'paradigm' shift. Showing that a theory is inadequate by showing that certain data cannot be explained is one of the tasks theoreticians of all persuasions pursue. A different type of critical approach to Model M — the one I will take — will try to show that Model M takes terms out of context or presents linguistic data in a non-standard way. This very move — so my claim — creates the <u>need</u> for a theory of meaning in the first place by creating an apparently homogeneous phenomenon. If an explanation is needed only due to this philosophical maneuver, then a critical approach to the maneuver leaves things as they are — multifarious and heterogeneous, and in no need of a philosophical theory. If

---

[174] I.e., the elements of language that are combined by means of rules.

meaning and reference are not the essence of language, then there is no single thread that connects all our linguistic activities; our diverse activities make up language much like different fibers make up a thread.[175]

Wittgenstein prepares us to shift our paradigm by using the notion of a game to exhibit activities that are in some sense similar to Model M. This notion does not commit Wittgenstein to the view of language as a calculus of strict and uniform rules which Model M sees as the essence of language, and Kant and rule-utilitarianism see at the center of morality. Wittgenstein's reflections are thus merely reminders for the purpose of exposing how the notion of a rule-governed activity can trail off in all sorts of directions. When we think of language and morality as a calculus, we may think of chess as the example of a rigid and surveyable game. But chess is not the only game.

> The work of the philosopher consists in assembling reminders for particular purposes. (*PI* 127)

The reflections of the *Philosophical Investigations* do not attempt to show that there is something deep or that the reminders - cumulatively, as it were - exhaust the depth of the subject matter. There is no subject matter, at least not in the way philosophers conceive of it. There is thus

---

[175] The metaphor is Wittgenstein's; cf. *PI* 67.

no result, nothing that one accomplishes in the traditional sense when one performs such philosophical investigations. In particular, no philosophical thesis will result from such investigations. The remarks and observations can be trivial and should not be considered a treatise.

> The real discovery is the one that makes me capable of stopping doing philosophy when I want to. ... [W]e now demonstrate a method, by examples; and the series of examples can be broken off. - Problems are solved (difficulties eliminated), not a *single* problem. (*PI* 133)

What we have to resist is the temptation to be dragged onto the terminological ground philosophy prepares: whenever we discuss a term, we should resist thinking of some conceptual depth which we have to fathom with philosophical theories. We have to investigate from the outside, assembling reminders of the bigger structures that hold the words in place, rather than following the (merely) promissory notes of clarity and distinctness.[176] If we remind ourselves of the bigger structures, we can see the distortions of philosophical conceptualizations more easily.

---

[176] Compare Tractatus 5.5563 and Inv., § 97, for the different notions of clarity.

## Language

How does Wittgenstein show that language has no essence? Since he investigates Model M, he must investigate the view that *meaning* is *reference* or *naming* and the view that language employs an abstract system of recursively specifyable rules that mechanically determine their application. Consequently, I will first investigate the naming relation Model M sees as the essence of meaning, and thereafter the notion of rule-following. Philosophers, I have argued, argue for Model M with statements that are at first perfectly correct as they stand. Augustine most likely did not intend the remarks quoted by Wittgenstein in *PI*, Section One, to be a theory of language at all.[177] But such an innocent observation may also be used for other purposes. Davidson's statement that the meaning of a sentence has to be derived from the meaning of the word is clearly philosophically tendentious, although we may not notice it as such. Both the innocent observation, as well as the initial move or 'motivation' for a proto-theory of language are based on the view that every word signifies something. The statement, however, that a word signifies something does not mean what Model M makes it out to mean:

> When we say: "Every word in language signifies something" we have so far said *nothing whatever*;

---

[177] Cf. Goldfarb (1983).

> unless we have explained exactly what kind of distinction we wish to make.[178]

Wittgenstein's claim is justified by the fact that this is how we use 'signifies'. The term is primarily used to make distinctions between different terms, not between, say, tables and chairs on the one hand, and marks and noises on the other — the former failing to signify things, while the latter has a philosophically puzzling property (meaning) which makes marks and noises signify something. The idea that recursively specifiable rules allow a decompositional analysis of complex meanings is similarly misguided.

> [I]t may come to look as if there were something like a final analysis of our forms of language, and so a *single* completely resolved form of every expression. ... This finds expression in questions as to the essence of language. (*PI* 91)

This view that there are fixed rules that determine semantic and moral evaluations is common to Model M and the Model of Morality shared by Kant and rule-utilitarianism.[179] Exhibiting the full context of 'meaning', a term which Model M elevates to mark the essence of language, is intended to show that the need for a certain type of theory does not arise. Wittgenstein does thus not focus his attention on refuting theories or skeptics, but he wants to investigate

---

[178] *Inv.*, § 13. Emphasis added. Philosophers take ordinary notions and make them into 'superlatives' (*Inv.* § 97). Cf. also Goldfarb (1983) for an elaboration on this idea of the relation between the move from ordinary concepts to philosophical theories.

[179] Compare *PI* 92.

whether a need for theory or explanation would still arise if we consider the full context of the terms in question. It is in this context that we engage in the activities of justifying and explaining.

The philosophically significant terms are *naming, meaning, truth, understanding, knowing, believing,* and *following rules*. Investigating the philosophical preconceptions and theories for any one of these terms faces two tasks. A detailed study of the term's use is designed to show that such use need not give rise to the philosophical questions; exhibiting the local context of a term is done to achieve a detailed understanding of the term in order to avoid misunderstandings.

> In order to see more clearly, here as in countless similar cases, we must focus on the details of what goes on; must look at them *from close to*. (*PI* 51)

Once the details give us a picture of the use of the term, we need to show how the term in question is related to other terms which are equally essential to Model M. The latter task often amounts to general pointers into the direction of what one might call an overview of a term under consideration, when such a connection is suggested in the particular passage. We need not only show that a term is used in a variety of contexts but that a variety of concepts are interwoven with the term in question. This net of

relations among terms shows the wide variety of contexts of the term, its interconnection with other 'landmark' terms that are related to it.

> A main source of our failure to understand is that we do not *command a clear view* of the use of our words. (*PI* 122)

## Method

This method of investigation is actually rather simple; it investigates philosophical problems by trying to present the *grammar* of the terms that have been given a role in Model M. The *grammar* of a word is the role it has in various circumstances — in detailed nuances and general connections to other terms. Being able to use a term in a variety of circumstances is what understanding a term amounts to.[180] Philosophers, on the other hand - so I suggest after reading Wittgenstein - look at only a limited use of a term, thereby limiting the grammar of the terms. This restriction makes the concepts strange, misunderstood; once we no longer understand them, we try to explain them with a philosophical theory. It is a failure to understand by limiting the context and scope of a term that gives rise to philosophical theories. The more philosophy we do, the

---

[180] The 'grammar' of a term is thus very similar to what Katz calls 'sense structure'. It is relativized to a language game and not "deep structure", an abstract, platonic entity which is independent of our knowledge. Especially in the 60's and 70's, commentators were greatly concerned with what seemed terminological subtleties of Wittgenstein's work: the meaning of terms like 'criterion' 'grammar' etc. These terms were taken to be essential to his philosophy. I think that this is misunderstanding Wittgenstein. The *Philosophical Investigations* are intent on preventing our making concepts into super-concepts — a task to be accomplished by exhibiting the normal context of terms. A good dictionary (or a good William Safire-type analysis) should be sufficient to tell us what Wittgenstein (as everybody else) means with the words he uses. If this is not correct, his *Investigations* would be trying to espouse a theory or philosophy, which he denies.

more philosophy seems required in order to explain the more and more mis-understood terms. Once we look at the grammar of the terms, we will understand them again, and no longer see the need for a philosophical explanation. This, I believe, is a thesis that can be attributed to Wittgenstein, but it is not a 'philosophical' thesis - understanding a word is to be able to use it (correctly) in a variety of contexts. The only way one can argue for such a thesis is by showing how the needs for explaining terms disappear once we investigate certain philosophical moves.

I take the first dozen sections of the *Philosophical Investigations* as an example of this method. They detail the use and the different types of meanings a term has depending on the context and area of use. In the early passages of the *Philosophical Investigations*, Wittgenstein is pursuing two strategies: he wants to point out that language is not only or essentially a reference tool and wants to introduce his method which consists of giving the grammar of a term so that we start losing the need for a particular type of explanation. We can start investigating the idea that language has essential properties by investigating the idea that language has an essential property, namely reference. This property relates language by means of names to the world. We may start investigating this view by investigating the circumstances under which we

use names. This does not amount to a nominalist point of view, nor does Wittgenstein want to legislate use. Wittgenstein does not prohibit us from saying that "the rule determines the application" as long as one does not read too much into such notions and avoids misunderstanding them.[181] Wittgenstein has a 'central' notion in his philosophy of language - he operates with the notion of family resemblance holding between different applications of a term. This notion, on the other hand, resists a philosophical definition since its point is that there is no essence to this resemblance. The notion of family resemblance can be used to resist an attempt to think that we need to give an 'explanation' to rule-following, or a solution to the rule-following paradox. There is no general explanation for rule-following over and above the explanations we give for the rules when we teach their use.[182] We 'recognize' or see resemblances, but there is nothing those resemblances have in common beyond the fact that some resemble others. Resemblance is not transitivity. The reason for grouping things the way we do cannot go 'behind' all groupings once

---

[181] cf. *PI* 48

[182] This fact will exclude certain rules from being *bona fide* rules. The Categorical Imperative is thus not a bona fide rule. We may think of it as a rule (like the Golden Rule) since its form is so similar to rules we otherwise employ. But it is hard to conceive of criteria that would allow us to tell that a practice is recognizably a practice that follows this rule. The same consideration can be raised with respect to the Golden Rule, although I don't want to argue this point here.

and for all because reasoning and understanding employs the rules and reasoning we use when those rules and groupings are explained. The argument for such a 'theory' of resemblance has to strike you — I think that the term used by Kripke is very adequate — so that you look at it in a particular way.[183] There is very little one can appeal to other than the appeal a paradigm has when it is exhibited.

The early passages of the *Philosophical Investigations* are intend to establish the fact that a feature of language which has a particular purpose in a particular type of situation cannot be construed as a feature that describes a function or property all of language must have. There are, in Wittgenstein's terminology, no metalogical concepts,[184] - concepts that tell us what meaning, or understanding, or referring <u>is</u>; there is nothing that *meaning* means other than what it is used for in various sorts of explanations.

---

[183] Cf. PI 144.
[184] Cf. Section 8 in *PG*.

## Names

A survey of the criticism of one element of the Model M Wittgenstein presents in the early sections of the *Philosophical Investigations* may serve as an example of this method. The element in question, is, of course, the idea that language establishes — needs to establish — a relation to the world using words as names of objects. Model M states that the meaning of a word is an object or function of a certain type and that a sentence is a combination of such meaningful words. If the thumb nail sketch of Wittgenstein's method and general outlook from above is correct, one has to expect that investigating this view will contrast Model M's view of 'meaning' and 'name' with the *grammar* of those terms. This method requires commenting on various strands of the meaning of 'meaning' and 'name'— an attempt to point out a lot of different misunderstandings that arise from seeing language as a system of names.

One of the issues that start philosophers thinking about language is the term *reference*. One might easily combine this with Brentano's thesis, that it is the mark of the mental that it refers, while everything else just <u>is</u>. This is part of the view that sees language as a tool for denotation, something that links the mind and the world. Reference thus needs to be explained, since it seems to be <u>the</u> essential feature of language and mind. There are, of

course, cases in which words are names of objects, just as the passage quoted from Augustine suggests. But this is merely an example of one of the many uses of language — it is a feature that characterizes our use of language - not a metaphysical explanation of language. In contrast to this characterization of language and language use, Model M is not interested in finding out what kind of uses names have. Instead, Model M is a theory of how language establishes a connection to something other than language - facts or the world. This creates a distinction between language and all other things. Names are to bridge this schism between language and something metaphysically different. Reference is thus analyzed as a relation. The use of names e.g. for tagged items is seen as a relation between two elements and considered as the model of the metaphysical relation between language and the world. But to equate the notion of meaning with reference and continue by equating the notion of reference with that of a name simply ignores a wide variety of different uses these terms have. We, for example, use inscriptions in order to designate that something has such and such a name. We use inscriptions also to designate the owner of something, or its price, or any number of things. The notion of reference, Wittgenstein tries to show, is not a notion that can explain, from scratch, how we understand or master language, but has a variety of uses primarily designed to avoid misunderstandings once language is already

in place.

> Augustine describes the learning of human language as if the child came into a strange country and did not understand the language of the country; that is, as if it already had a language, only not this one. (*PI* 32)

This is a point that is difficult to establish if it were conceived as a thesis. Instead, what we should expect are therapeutic pointers that investigate what kinds of uses names have, and when questions about meaning arise. We have to be lured away from insisting on the question, how language relates to the world or objects, lured away from the model that suggest that language as a whole stands in some sort of vocative relation to the world, making language a system of names and denotations of things.

Names are a motley crew. Consider the fact that we use names and inscriptions to designate objects. We use name tags not only to denote a person, but also to present her and introduce her. The use and function of the tag is different before the tag has been picked up at the conference table, and after it has been attached to the fabric. Similarly, we use inscriptions on paintings, or book titles. The use of names and titles, however, varies widely. While a book title usually is an inscription that is on the cover of the book, the book is usually only a copy of a print run of, say, 1250 copies. The name or title of the book does not refer to the copy, but the book, which is

a totally different thing. A name tag, on the other hand, is designed to identify just one particular person, and even if there are 3 different 'Joe's around, that does not make them copies of a 'Joe'. The name 'Hamlet' is used without having a bearer, and 'Germany' is a name of a country that is constantly changing. It is thus quite unclear what kind of names the words are supposed to be the Model M has in mind.[185] The statement that words name objects, and that this is in general the case of all language and its relation to facts gives no explanation just what type of names are supposed to be employed in the model of reference and explanation. The ordinary use of names does not invite the idea that names are a relation between a mark and an object - while the metaphysical thesis that there is a relation between language and the world that is established by a name is a completely different thesis from saying that Ludwig was Wittgenstein's first name. It seems more likely that the idea becomes much less plausible once the various ways in which names are actually used are presented.

Suppose someone were to insist that we can always reduce all features of language to the naming of an object. In addition to the above considerations — that there are various ways we use names — we may point out that, for

---

[185] Commenting that that they are logically proper names does not make things any clearer.

example, referring to something by saying 'tiger' and thus using what appears to be a name of an object or a natural kind, often has the function of differentiating between words once a certain language skill is already in place. This permits us, for example, to ignore vast differences among the referents involved. Asking, for example, what 'two' means, we can point to two books, or two tables; asking what 'pipe' means, we can show someone a pipe, or the picture of a pipe. In cases like this we use the term 'means' correctly, but quite obviously, the term 'means' is used in a variety of different ways in those interactions. The grammar or use of terms like 'means', 'refers', etc. indicates that their usefulness and need arises only once the different logic or different grammar of other words has already been understood. 'Meaning' and 'name' is thus a secondary practice - is serves to distinguish within the practice and among the uses of words, and cannot give it a foundation. It is then that terms like 'means' and 'refers' get to be used, in order to differentiate between different words and their function. They thus do not 'anchor' language to reality, since they are used to explain words, help avoid misunderstandings; they are not concepts that explain the essence of language. If it is characteristic of 'meaning' that it is primarily a tool to help avoid or clear up misunderstandings, then the philosophical question "what is the nature of 'meaning' and 'reference'?" would have to

be transformed to "what is the nature of misunderstanding words and what is the nature of its remedy?" — a question which is intelligible if we do not interpret 'nature' as being a question for essence and metaphysics in general.

The above sketch points, I believe, into the general direction of the method Wittgenstein uses in order to avoid philosophical theories. What one needs to do with philosophical questions is to ask whether they would arise if the notion that is under investigation were not used in a non-standard way. We have a very clear understanding of how we use names, and what they are used for. They are not a relation between language and the world. The very idea that there is a 'relation' that is established is the one Wittgenstein would like us to re-consider.

## Games

Starting out with the quotation from Augustine, Wittgenstein notices that its underlying thesis is a pedestrian version of some very sophisticated theory of language. Wittgenstein's first reminder is, that denotation and combinations of names is not all that there is to what we call language. 'Five red apples', for example, is part of the English language and does not conform to the picture of language Augustine gives. But although our language is vastly more variegated than the Augustinian picture suggests, a picture which thus does not give the essence of language, we can imagine circumstances in which we use language to name objects. Naming under those circumstances makes no sense without having some purpose; we don't go for a walk in the woods and name the leaves that flutter by, or the rain drops that glisten on the window pane.

When we introduce children to the meaning of words we introduce them to a limited context - a limited use - of a term. For certain types of communication, like the type of communication the builders use in the early section of the *Investigations*, such introduction may be compared to being introduced to a person by name. Meeting someone introduced as 'Hilary' will not give you an idea who that person is; similarly, a child being introduced to 'tree', 'dog', or 'slab' can be introduced to the meaning of the term by

having someone point out an example of the kind in question.[186] Such an introduction to a term's meaning might be appropriately phrased "this is called a dog". The child has thus a very limited conception of what a dog is, even though the indefinite article might give the child a clue that 'dog' is grammatically, i.e. as a matter of its different linguistic function, different from 'Lassie' or 'Fido'. Introducing a term (and the rudimentary form of its meaning) like this does not give the child the full meaning of 'dog', i.e. that it is a member of the canine species etc. The child will not know what the *grammar* of the term is; it will, for example, have probably no conception that "the dog hopes that the weather will be fine tomorrow" is, in this sense of 'grammar', not what we say about dogs — except in stories or cartoons. The child may simply not see the difference between its toy dog and the animal, since it may not know that the term 'dog' implies 'animal'.

We thus present only a limited context to the child. We design certain simple circumstances in which we use the term — use it correctly — which are transparent to the child. The meaning of the term is, of course, not exhausted by the use we introduced. But, on the other hand, we could easily conceive of the meaning of a term to be 'complete', and use

---

[186] In case it is the kind of word that designates some concrete object.

the term only as introduced to the child, with a very limited and circumscribed meaning. American English has morphological markers to indicate terms which have such narrow meaning. In conversation with children, a dog is a 'dogie', a horse is a 'horsie' etc., indicating that the meaning of the term is different from the term 'dog' although semantically rich enough to ensure that the child knows that 'dogie' does not mean 'horsie' etc. That is, the term has sufficiently many 'language game moves' that overlap with the 'adult' word. They resemble each other. Nevertheless, when children talk about dogs, cats and the like, the meaning of the term can be seen as complete, but narrow, tailored to the knowledge capabilities of the child.

In this sense, the term is accurately used when it overlaps sufficiently with the 'adult' term. This leaves one with at least two options. One can now either say that we have given the child a fragment of the language or word meaning. In that case we assume that language as a whole and words in particular are characterized throughout by a certain property which the child has not fully understood: there is the meaning, and the child knows only a part of it. The fragment of such a language to which we introduced the child is part of a uniform entity *language* which has certain properties - *meaning* - where meaning is the denotation of an object by a word. But there is, at this point in the

argument, no evidence for the assumption that language has an essence. There is no good evidence that there is such a thing as 'the' meaning of a term, which is either gotten or not. There is thus a second methodological strategy available. If we don't assume that the bits and pieces by which we can train people to acquire linguistic behavior amount to being introduced to a thing with an essence, we may instead assume that they accomplish the task of learning certain skills that are somehow related to each other, but are not part of an essence, a universal property.

Instead of trying to understand language as being a body with an essence, one might try to analyze language according to certain features concepts have. This is why it is useful to compare the concept of *language* to that of game.[187]

There are apparently no necessary and sufficient conditions for something to be a game. A game has no central properties, no essence. There are similarities among games, but there is no essence they have in common. One may try to say that they are for entertainment, but war games are not entertaining. One might say that they are played at least with two players, but there are games for one player only.[188] This does not force us to admit,

---

[187] The German "Spiel" means both "play" and "game". This should be kept in mind when we read Wittgenstein. "Play" seems much less oriented towards fixed rules than "game".
[188] Given the current number of video games played, the saddening tendency is towards the game for one person.

however, that a game is an indefinable thing, a view Moore held about moral terms.[189] There are certain features of some games (not all games, but a good many of them) that allow us to tell when a person has learned a game, a way in which the child can get it right or wrong, when we teach the child a game, or the term 'horse' or 'chair'. The fact that there is something that the child can get right or wrong indicates that there is something that is taught, that is, for which there is success or failure in the process of imitation and training. We train people in various skills, and there is thus a certain measure of what might count as a success. But teaching someone to dance, for example, does not require that there are rules. It is teaching them to express themselves through bodily movement (ballroom dancing is a different story). In the vague sense in which there is success, there is, in the same manner, a way in which we can say that the skill, the activity or the function that the child is to imitate is somehow governed by rules, if only for the reason that it is an activity that is recognizably the same when done again, *viz.* when it is imitated by the child. If what we are able to teach are certain regularities, then being able to do the same, pick up a skill, is to be able to see those regularities. The notions of regularity and sameness are thus related.

---

[189] Moore (1903), esp. section 8.

> The use of the word "rule" and the use of the word "same" are interwoven. (*PI* 225)

The fact that terms are related means that when we introduce one term, we will make reference to the other as well:

> Then am I defining ... "rule" by means of "regularity"? - [I]f a person has not yet the concept[s], I shall teach him to use the words by means of *examples* and by *practice*. — And when I do this I do not communicate less to him than I know myself. (*PI* 208)

But whether we do say someone is following a rule, or whether she is doing the same now as she did before depends on a variety of criteria, or interests:

> If from one day to the next you promise: "To-morrow I will come and see you"- are you saying the same thing every day, or every day something different? (*PI* 236)

I suppose Wittgenstein means to say that this <u>is</u> an open question, a question for which there cannot be an unqualified "yes" or "no". If we wanted to define whether someone were to follow a rule, or said the same thing, it would require that we give a rule for saying when we follow rules, or when "same" is same. This way of looking at identity as the rock bottom of recursive rules (a=a) is rejected as well:

> But isn't *the same* at least the same? ... We seem to have some infallible paradigm of identity in the identity of a thing with itself. ... Then are two things the same when they are like one thing is? And how am I to apply what the one thing

shows me to the case of two things? (*PI* 215)

If it is thus not up to a priori, abstract entities that our practices have the regularity they have,[190] it seems to depend on the games what the rules are that we follow and what rule-following is. The rules that govern the game may be rules only in the way in which 'go south' is a direction, and 'be careful' advice. This is an important feature for the concept of game, yet we cannot claim that all games have elements that need to be taught, or that all games allow for a strategy, have a point, have rules, or have to be played with more than one person. Thus language may be usefully compared to the notion of a game. Language is not a game, but language contains different types of games. Just as there are different games, games of different type, there are different areas of language. While chemistry and mathematics resemble games like chess, other language games are much more complicated and less open to codification, especially because we make up new rules and dump others as we go along (especially in political and moral contexts). And such parts of language, sections of language, Wittgenstein calls language games. Moral, social, and political discourse, poetry or literature in general often do not conform to games that have fixed and rigid structures which are surveyable, or have certain, clearly visible

---

[190] The regularity admits of infinite varieties of types and degree.

points and aims like the building people in the second section of the *Philosophical Investigations*. Those types of discourse are nevertheless part of our language. What, for example, one might ask, are the rules for playing Frisbee, or playing with a dog, a child, or playing with a ball all by yourself? Just as the notion of painting allows for a de Kooning and a Rembrandt, the notion of game is flexible enough to allow for things like Chess or Go, on the one hand, and playing with a ball or playing with a hand to see its reflection and shadow on the surface of a swimming pool. There is thus no essence to the notion of a game. Similarly, certain parts of language seem much more suitable to seeing language as a calculus of names. Instead of denying that there is anything to this view, Wittgenstein does the next best thing and compares it to some of our games that have certain rules. But in order to shift the focus from a very simple game that Model M presents, he takes the analogy of a game and lets the considerations drift further and further away from the ideal of language as a rigid mechanism of rules. Instead, the simple language games that would conform to something like the calculus of rules would not only make for an extremely simplistic life, but even in those games we are expected to sometimes make the rules up as we go along.

When we thus change the paradigm for seeing language

(and how would one want to argue for this other than by exhibiting the strength of the new paradigm), we may loosen the grip of trying to fit language into the Procrustean bed of compositional semantics and Model M. In as much as there are, for a good many games, regularities, rules, and interaction between a trainer and a trainee, the analogy of a game seems not to beg too many questions. It also is designed to abandon the notion that language games are the link between language and the world, sustain the representative relationship between language and reality.[191] The most rudimentary of such games are learned without explanation, but still require some training; this suggests that certain responses and groupings are natural to us, and that, in the case of language games, the language game of explanation builds on this natural response. The natural response is something that would be excluded from explanations.

Games, like the rudimentary instructions for the use of a term we give to a child, allow for expansion: We may start with a square board of 16 squares that alternate black and white, with two sets of figures, call the ones 'pawn', and a different one 'king', and declare the aim of the game to 'check mate' the *king*. Of course, 'king' and 'pawn' are not really what they are in 'real' chess, but they are

---

[191] Hintikka (1976) maintains this thesis.

sufficiently similar. On the other hand, we can very well imagine that this game is a game to itself, just like 'Tic Tac Toe' which may or may not be seen as a beginner's version of the Chinese game 'Go'. We similarly imagine the way children are introduced to the meaning of words by being introduced to terms in a limited context.

## Rule-Following

The first two sections of this chapter were intended to persuade the reader to get to a point where answers to analytically oriented questions about language are not the first reaction to a discussion about language. Once one suspects a philosophical tendency of assimilating differences and neglecting details and whole areas of the fabric of language, the idea of language as a thing with certain essential properties is no longer the default assumption. This allows Wittgenstein to introduce a picture of language that is different. Certainly language has regularities, and there is no doubt that language has certain rules we follow when we speak language or communicate in general. Thus, the question of how we are to understand what following those rules amounts to seems at first reasonable. But we may simply be misled by the fact that we have one word for very many different things. The question what rules have in common, and what explains the fact that we follow them is not obviously misguided like the question what all the people by the name of *Donald* have in common. Still, the question what rule-following consists in and what the essence of rules are questions that Wittgenstein ultimately rejects. The point I would like to lay out, much, I hope, following Wittgenstein, is that there is not really one thing that rule-following amounts to, and that the question, what explains rule-following does not

have an answer that points to an essence. this in turn implies that there cannot be one rule that is underlying all our moral or linguistic practices. This way of putting the issue also makes it unnecessary to show what Wittgenstein's 'solution' to the problem of following a rule is. In a different context, Wittgenstein writes:

> If I am inclined to suppose that a mouse has come into being by spontaneous generation out of gray rags and dust, I shall do well to examine those rags very closely to see how a mouse may have hidden in them, how it may have got there and so on. (*PI* 53)

Our mistake is, perhaps, to think that there is a mouse. The investigation of the rags and dust may reveal just that: rags and dust. If we don't find the mouse, it may be because there is none. The investigation into the issue what rule-following is targets the idea of language as a certain type of game: a game that Model M describes. Stating that games can be useful for the understanding of language must, however, not be misunderstood: language itself is not a language game.[192] Language is a family of language games.[193] There are different forms in which we employ games, just as there are different kinds of things that we call numbers, different kind of actions we call good, or different kind of circumstances under which we are inclined to say that something is meaningful. The notion of

---

[192] Cf. Floyd (1991) p.147.
[193] Cf. Rhees (1959) p. 171

a game, especially when used to refer to a simple one in which we find strong regularities at work, may trigger the question what rule-following consists in and what justifies or explains following a rule. With the considerations of the above two sections about method and language in mind, investigating the rule-following paradox may shed some light on how Wittgenstein resolves or dissolves this philosophical puzzle, and it may enable us to escape the philosophical question as to what rule-following consists in, and what the essence of rules is.[194]

There is no doubt that language and morality make essential use of rules, but those rules are not 'primary' in the sense that they could be understood without the practices and use from which they arise. This view does not commit us to say that they are 'invented', either. We invent games, and invent computers. And we discover that AIDS is caused by a virus. These are grammatical remarks; this is simply how those terms are used. It stands to reason that the philosopher who wants to debate whether necessity is invented is simply asking a nonsensical question, since the words simply don't have the meaning the

---

[194] The reason I spend so little time discussing the rule-following paradox when discussing rule-following is part and parcel of my deflationary reading of "Wittgenstein's philosophy" of rule—following.

philosopher thinks they have.[195]

It is this interplay between practices and rules that philosophers want to ignore in their theories. Idealism seems to think that rules are invented, while Platonism and Mentalism seem to think that rules are there without the practices. It is hard to imagine that only one person once played a game, which would indeed be possible if the Platonist were right. But it seems equally impossible to imagine, that all our rules and laws are mere inventions (that we could decide tomorrow that gravity is just a boring game). Wittgenstein claims that he wants to make the nonsense of philosophy more patent. If Platonism and Mentalism are described like this (and that isn't all that inaccurate) they become the nonsense Wittgenstein claims they are: Gravity is just an invention of ours! Language existed before the Big Bang, etc.

Following rules is thus not an application of an objective mechanisms, but is equivalent to the understanding exhibited in grasping the practices that form the rules. While Kant diagnosed the inability to apply rules as

---

[195] Similarly, to use an example Chomsky introduced, we could ask whether green ideas sleep furiously. How can one convince someone that this is a nonsensical question? The only answer that is available is the one that maintains that ideas cannot be green. Similarly, AIDS is not 'invented'. This is not a metaphysical rebuttal.

stupidity,[196] the idea, in turn, that rules automatically (mechanically) determine their application[197] may be adequate for mechanical beings - beings that lack intelligence. I will try to show that the notion of rule and rule application are intimately interwoven with the notions of understanding and agreement. Rules and rule-following is — just as understanding — fundamental to our language games:

> Following according to the rule is FUNDAMENTAL to our language game.[198]

Rules and rule-following have no essence. Understanding the meaning of a sentence is not applying some Russellian or Fregean calculus to 'noise or marks on paper'. "To do the right thing" is not a mechanical maximization of utility governed by ideal rules, nor intending something that could be a universal law, if only because the requirements for moral evaluation are diverse. The rule-following paradox is thus just one of the many considerations Wittgenstein presents in order to investigate the *grammar* of the notion of a rule. While investigating the grammar, he investigates many aspects about the notion of a rule, and the rules he

---

[196] CPR, B 171.
[197] Kant held the view that the principles of reason apply to the rules of the understanding.
[198] *Remarks*, p. 330 (vi, 28). Wittgenstein should have said "language games", not "language game". Our language is not a language game. (Cf. J. Floyd's essay mentioned in footnote 192.)

presents have very many different characters. It is thus, I believe, a mistake to think of the rule-following paradox as a pivotal element in the considerations regarding rule-following. I concentrate on the rule-following paradox mainly in order to show its relative (in)significance for a picture of language. If rule-following is seen as the essence of rationality, then rationality is a much wider notion than employing a calculus. The notion of *rationality* is in fact interwoven with the notion of *rule*, but not in the way the algorithmic models of language and morality have seen it by reducing *rule* and *rationality* to notions that fit into the space taken by an algorithm.

Wittgenstein covers the paradox in more or less 5 sections, sections 198 through 202; and he seems to have various things to say about the paradox which offer a couple of straight dissolutions of the paradox. Here is one of the various formulations one might use to present the paradox:

> 'Orthodox' statement: A rule determines an action.

We formulate rules since we think that they determine an action. If they were not, it would be hard to explain what it is that they are supposed to do. But one seems to be able to maintain that there are various ways in which rules can be interpreted. "From a logical point of view" there is always the possibility to give the rule an interpretation

that allows a different result from the one that is the one believed to have been determined by the rule. Thus we should agree to the following statement:

> Statement of logical fact: Every action can be interpreted in such a way that it conforms to a rule.

This contradicts the thesis that a rule determines an action, since the notion of determining implies that one gets such and such results from the application of the rule. The fact that a rule can be interpreted so that no such necessity can be found implies the following statement:

> Paradoxical statement: A rule does not determine an action, since <u>any</u> action can be made to conform with it.

What is required, of course, is a response to the assertion that, from a logical point of view, any action can be made to agree with the rule, since this is tantamount to saying that a rule does not determine an action. One is therefore left with two different assertions:

*A rule determines an action* vs. *A rule can be interpreted any which way.*

The paradoxical statement has, of course, the disadvantage that its name <u>means</u> that no-one believes it. How does anyone, then, accept it as a statement? The paradox seems rather a symptom of a misunderstanding, and Wittgenstein diagnoses it as a misunderstanding of what

agreement and interpretation means. It may seem that the paradox would be best addressed by arguing that it is a *petitio tollendi*, that is, it begs the question for something that one wants to dismiss.

> 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. (*PI* 201)

Wittgenstein rejects this solution, and I will try to explain why this is the case. Instead of outlining the various strands of Wittgenstein's attempt to try to give the grammar of the notion of "rule" and "determining" I will focus on the notion of interpretation, since this notion as much as any is part of the paradox.[199]

Section 198 puts the rule-following paradox in a nutshell:

> But how can a rule shew me what I have to do at this point? Whatever I do is, on some interpretation, in accord with the rule.

The interlocutor seems to suggest two things at the same time: on the one hand it seems to be suggested, that a rule needs an interpretation, since the statement states that any rule can be interpreted, thus making no difference between an interpretation of a rule and its — as it were — regular or normal meaning. On the other hand, the statement also

---

[199] For a discussion of Wittgenstein's investigation of 'determines' see the article in footnote 192.

distinguishes between a rule and its interpretation, that is, it assumes that the rule is somehow independent of the application, as if it had an independent, logical existence, which needs to be understood, interpreted. It may thus be useful to approach the issue of rule-following from the perspective of the interpretation, rather than by discussing the issue of what a rule is, or in what sense we can say that a rule determines an application.

What do we mean when we say that we interpret something? We employ this phrase most commonly when the meaning of a term or rule is not clear or not originally meant to apply to a certain situation. We sometimes don't know the meaning of something in a particular situation, but there are cases for which we do have indisputable instructions, so that a 'straight' meaning is available. Suppose we said an augur interprets a bird's flight pattern; someone saying this seems to assume that there is something that has a certain point and regularity to it, and has some meaning. If one were less generous one would say "a so-called seer "interprets"...". Interpretation thus is different from inspiration. Similarly, suppose one were given a sheet on which color dots were randomly distributed, and were asked to interpret it. This would be incomprehensible; we may freely associate, but if the dots are distributed at random, one should not use the term *interpretation*, but rather

*inspiration*. We then may be said to 'read' into the picture, or we associate something with the picture, but we do not interpret the picture. The point about Freud's association techniques is that the target lacks a feature that can be interpreted, so that any 'interpretation' must be purely subjective, that is, tell us something about the person. If we know that there is nothing there, interpretation makes no sense at all. Although, and I insert this in order to prevent misunderstanding, we have not made clear what pattern, and what 'regularity' mean. If there is no pattern at all then we can either say that anything we say is acceptable, or, alternatively, we can say that the notion of interpretation is not applicable.

Interpretation thus requires that there is something for which interpretation gives further detail. It depends on there being something that can be interpreted. If this were not the case, then interpretation would be independent of what it interprets, independent in the sense of 'autonomous'. But the grammar of 'interpreting' indicates that it is relative: we don't say in the same sense of 'interpretation of' 'rule of'. We often say that there is a rule of law, or a rule of inference, but this 'of' is a genitive that indicates the area for which the rule is relevant. The phrase 'interpretation of' indicates that it is not independent of the thing for which it is an

interpretation. Since it is as such an item that has a dependent existence, it cannot be the one and only one thing that determines what the thing it interprets is.

> [A]ny interpretation still hangs in the air along with what it interprets, and cannot [be the one that props it up]. Interpretations by themselves do not determine meaning.[200]

This leaves us with the question, what kind of relation there is between a rule and the action that is the application of the rule. Instead of attacking this issue, I will again try to approach it from a different perspective: when do we say that a rule is not applied, and when do we say that the rule is not properly applied? Suppose you notice that your cab driver runs every red light. When the signal shows green, she stops. The cabby explains that she does not want to get killed by someone running a red light. It is certainly up to interpretation what has happened in this case; and it does not seem to matter which option we pick, as long as we don't get confused about what happened. In the case of the cab driver, we can say that the traffic rule (law) that prescribes that you have to stop at a red light has simply been replaced by one that directs you to stop at a green light. This is not an interpretation. There are, of course, cases in which the case is not so

---

[200] *PI* 198 [translation changed]. The original reads: "sie kann ihm nicht als Stütze dienen." Anscombe translates "cannot give it any support", which I find rather misleading.

clear cut. For example, "Walk/Don't Walk" signals in Manhattan are ignored by a good number of the pedestrians. Those rules are no longer uniformly followed. That is, the way we use the phrase "a rule determines our action" is when we want to indicate that we follow rules, and do not ignore them.

Notice that in this scenario we used the phrase 'a rule determines an action' without ever raising the question as to what 'metaphysical' facts are responsible for this determining. We simply made a distinction between a certain type of population that follows rules, that have been brought up to such and such behavior, and others, that are not:

> We use the expression: "the steps are determined by ...". *How* is it used? ... [We may for example say] "These people are so trained that they all take the same step at the same point when they receive the order [Walk]. *(PI 189)*

The question in what sense a rule determines an action has nothing to do with the (philosophical superlative) question what metaphysical facts there are that establish that we follow rules. That is, following a rule, in the sense that I intend to follow it, requires that I be directed by the rule, that I orient myself after the rule. It is not a question what metaphysical fact determines that I follow rules. The question of following rules comes up in situations where not following rules is at issue, for

example when someone suggests you do something spontaneous. We may say that someone is spontaneous, or that she is a free spirit, in contrast to someone whose life is too regular, or who has too many rules and regulations in their life; for them it may be a good idea try not only to allow for a little latitude and leeway, since not everything in one's life should be planned and regulated. Rule-following is thus a notion that often is contrasted with spontaneity. That is, in the context of rule-following, we seem to ask for the difference between something that is following certain rules as opposed to something that does not follow rules, or follows different rules. The metaphysical question, by contrast, seems to ask: what is it that makes us follow rules vs. what is it that makes things behave randomly, or any which way. But the point about randomness is that it is beyond understanding, which means that the notion of understanding and explanation are 'cousins' of the notion of following rules. They are on the same "intellectual level", and we cannot give an explanation of what makes us follow rules which goes beyond the rules that we actually follow. This, of course — to extend the metaphor Wittgenstein uses — allows for family members to be asked for support. Where we understand, we employ rules; and the varieties of understanding in our life are the varieties of rules. Each rule-following is explained by us understanding the practice the rules establish. Rules are

thus not 'a priori', independent of our understanding. Interpretation is part of understanding. If we are able to interpret everything any which way, understanding would be impossible.

In section 201, Wittgenstein tries to show that one way out of the skeptical paradox is not feasible:

> The answer [to the paradox] was: 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. (*PI* 201)

The rule x+1 for argument 1 'rules' that the next number is 2; according to the view (rejected as misunderstanding) in 201, writing down '2' would thus be as much in agreement with the rule as writing down anything else, and thus (and this would probably find Wittgenstein's agreement) rejecting the linkage between rule and agreement.

That this is a misunderstanding is due to the fact that we still operate with the notion that there is such a thing as an interpretation that is independent of the rule. To make this more explicit, the passage should read:

> Everything can be interpreted. If this is the case, it can be interpreted to accord with a rule, and interpreted to conflict with it. And so there would be neither accord nor conflict here.

Wittgenstein thinks that this is a misunderstanding. Kripke seems to accept such an argument:

> The skeptic ... puts the challenge in terms of a skeptical hypothesis about a change in my usage. Perhaps when I used the term 'plus' in the *past*, I always meant quus.
>
> Of course, ultimately, if the skeptic is right, the concepts of meaning and of intending one function rather than another will make no sense. For [if it is correct] that no fact about my past history ... establishes that I meant plus rather than quus ... there can be none in the *present* either.[201]

Kripke uses thus the same reason for introducing his dramatic device that Wittgenstein uses in order to show which way out of the paradox is misguided.

> Kripke: meaning makes no sense if plus could be quus.
> The rejected solution at 201: Agreement makes no sense if plus could be quus.

Since Wittgenstein rejects the solution to the paradox he sketches at section 201, he would also reject the Kripkean scruples about introducing the skeptic. Wittgenstein rejects the solution and the scruples because of the notion of 'interpretation'. If we allow the skeptical hypothesis of *quus* and *plus*, we rely on the fact that there are rules such as *quaddition* and *addition*. Kripke's skeptic accordingly distinguishes between "plus" and "quus". In fact, Kripke makes the distinction all along which means that those rules are understood without "interpretation", and the distinction relies on the fact

---

[201] Kripke, p.13

that we understand the rules. And this, to invoke Wittgenstein, is simply a trivial reminder for the particular purpose of establishing the grammar of understanding a rule.

The paradox, then, turns out to be just one of the many ways in which our talk about rule-following can be made to lead to exaggerated, metaphysical claims. The propositions that any action can be made to accord with a rule or the proposition that a rule determines the action in advance show that rules are understood against a background of practices that are in place for which we do not need an interpretation. Much like the discussion of ostensive definition in the *Philosophical Investigations* which is designed to show that ostensive definitions rely on the fact that the position of the terms in a language game is prepared, we understand the difference between rules and between different interpretations of rules only against the background of customs, Gepflogenheiten.[202] Once those are in place, we can ask for interpretations; but affirming that, of course, does not mean that rule-following has a new

---

[202] *Gepflogenheit*, in German, derives from *pflegen*, to take care of something. This is related to *Pflicht* (duty/obligation). (This connection was pointed out to me by S. Schwarzenbach). It is also connected to Middle English 'pliht', now 'plight'. There is thus a 'grammatical' connection (I use this term now very liberally) between necessity and custom. But this is not a metaphysical statement.

metaphysical underpinning, this time by the name of custom. It is custom that is needed in order to make sense of questions like which rule was followed, not which fact establishes that rules are followed. If there were not a stable net of interactions, and rules, we would not be able to ask questions about the proper understanding, or alternate understandings, of a rule.

Suppose someone were to say that the word 'plane' has no meaning since "She is drawing planes" can be given different interpretations. This only shows that the word is vague, not meaningless. When we thus talk about the interpretation of rules, we can have very many different issues in mind, and the rule-following considerations should be understood as showing that none of them allow us to derive a metaphysical picture of what rule-following consists in, or what rules are. If we deny that rules have the important as well as the logical status Model M attaches to them, then we might be less inclined to think of language and morality as a calculus of rules, although we need not deny that there are rules, and that there is no problem in following them in general. Of course, in specific cases — difficult legal discussions — there may be difficulties of interpretation. Morality and language are areas in which we follow rules, but it is also not a system of rules, nor are the rules independent of us, and our understanding. But we cannot

suppose that rules or rule-following have an essence.[203] The question, thus, what rule-following consists in and what rules are, seems a question that is misplaced once we consider the enormous variety, the multiplicity of types of actions that we consider part of regularities, and the way we, constantly, invent, reform and abandon rules.

Just as there are ambiguities in language which we resolve by asking the person who uttered an ambiguous statement, we can ask someone who presents a series that could be the expression of several different rules, which rule he was thinking of. But this question, as to how (for example, in the absence of the person who generated the series) to find out which rule generated a series is very different from the statement, that we could always interpret a rule in such a way that any action could be made to agree with the rule. The reason for this is simply that the terms agreement and rule form somewhat of a closed curve in space:[204]

> The word "agreement" and the word "rule" are *related* to one another, they are cousins. If I teach anyone the use of the one word, he learns

---

[203] I think the notion of 'rule' is, like the notion of 'cause' very heterogeneous. There is simply not one thing that corresponds to the notion of a cause. And , of course, the Aristotelian '$\alpha\iota\tau\iota\alpha$' as standing for 'cause' 'principle' and 'logical principle' is a heterogeneous, multiply connected big family concept (Großfamilienähnlichkeit).

[204] The metaphor is used (in a different context) in Quine (1953), p. 30.

the use of the other with it.

> The use of the word "rule" and the use of the word "same" are interwoven. (As are the use of "proposition" and the use of "true".)[205]

That is, when we talk about interpretation, we often assume that the rule is not expressed (we have a series that could be the series generated by a variety of rules), or is too general to tell us what to do in particular instances (drive carefully!, Don't read too much into it! etc.). Once the rule is expressed (and one of the criteria is, that it neither be too general, nor ambiguous), we can follow it. There are, of course, an unlimited amount of shades and gradations between the rules for chess, for golf and those for writing poetry (there are, after all, seminars for this). There are also ambiguous rules, or rules that are too broad. It is then that we have to apply interpretations. The rules of chess, for example, do not tell you what to do if the chess figures were to change their color every hour. Would that mean that these rules are open to interpretation as well?

---

[205] Inv., §§ 224 and 225

## Conclusion

The underlying idea of this dissertation has been that some philosophical theories regarding language and morality arise because of the way in which certain terms, in particular the notions of rules and their application, are systematically misunderstood. I have attempted to show how the various misunderstandings can be avoided. I consider this a therapeutic solution to a philosophical problem. A therapy does not refute, but tries to change the perspective of those who are in the grip of a certain idea. There is thus no theory at the end of such a therapeutic treatment. The approach is anti-Socratic, and, by extension, anti-metaphysical in that it tries to show that there are no essences, that the concepts we form do not correspond to objects or things.

This dissertation will, no doubt, not settle all the issues addressed here, or all questions a reader may have. I will thus try to end the dissertation by addressing questions that may still be open, in particular those that relate to the relevance of rule-following problem to moral philosophy. Kripke's paradox, I tried to show, is a paradox about rules and rule following. The paradox was poignant since lack of a fact that determines how a rule is to be followed would apparently force the philosopher to conclude

that I can follow a rule any which way I want. This is because the philosopher thinks that rule following *needs* a an explanation. Philosophers, since Plato, are essentially in the business of providing that fact, idea, essence etc. that justifies, explains and 'generates' the sameness, essence, identity of what appears to be diverse.

The point of the rule following paradox is important to a theory of meaning, since such a theory claims that the underlying essence of meaning is such a rule and rule following. If there is no such fact, then there would be no such meaning as the philosopher claims to explain. This is relevant for moral philosophy as well. Suppose there is a moral rule "Return property to the true owner." Of course the rule is there to help you act morally; that's the whole point of the rule. It is a simple rule: I see you lose $ 100, I pick it up, and return it to you. Because,

> I, like almost all English speakers, will be able to grasp the rule "Return ... ". Although I myself have followed this rule only finitely many times in the past, the rule determines my actions in indefinitely many new cases that I have never previously considered.[206]

Now suppose I encounter a bizarre skeptic. This skeptic questions my application to this new situation. Perhaps, he suggests, as I used the rule in the past, I should give the money to Patricia, since she needs it more than you do.

---

[206] For the phrasing of this section, see Kripke, p. 7f.

After all, I did not give myself explicitly instructions to return the money to you when I thought of the rule "Return F...."

But of course the idea is that, in this new instance, I should apply the very same function or rule that I applied so many times in the past. But who is to say what function that was? In the past I gave myself only a finite number of examples instantiating this rule. All involved other persons, not you, or other situations. So perhaps in the past I used 'Return.." to denote a rule which I will call "Nruter...." It is defined to return properties to those who lost it, unless you can give it to someone who is needy. Who is to say that I did not mean that rule, rather than the other. And here the skeptic goes again.

The skeptical point is thus not only that there is no fact, but the skeptic agrees with the philosopher that a) such a backing of rules is necessary for rule following and b) that the fact would be comprehensive since rule following is some essentially identical activity. *This is where the skeptic and the philosopher agree, and Wittgenstein disagrees.* The rule following skepticism applies, as Kripke makes clear, to all forms of rules, including, but not limited to, arithmetic, baseball, chess, and the rules of

forming a line at a bus shelter.[207] Kripke challenges us to present a fact that determines the application of a rule, and, since he maintains that no such fact can be found, he concludes that there is no such thing as following a rule: "Each new application we make is a leap in the dark" (Kripke, p. 55). And this was precisely what rules were designed to prevent.

Regardless of its general scope, and its application to any rules and regulations whatever, the impact of such a skeptical view regarding rules and their application is that it strikes at the heart of theories which are essentially rule based. Such philosophical theories rely on rules precisely because of their putative rational, objective and deterministic character. If philosophers who hold theories of this type cannot answer the challenge for a fact that establishes that the application of rules is not a 'leap in the dark', this should have a tremendous impact on their theories. If the application of rules of logic or syntax are not deterministic, then those theories are in serious trouble. If it turns out that the rules of tennis are not deterministic (I'm sure no-one ever thought they were) nothing of consequence should follow. Telling the person in the street that there might be all sorts of interpretations

---

[207] Compare Kripke's remark on page 58: "it can be applied to any rule or word."

of "No standing here" may produce little if any discomfort. Telling a function theoretically oriented linguist that the rules of her calculus still make any application a "leap into the dark" will strike at the heart of the very purpose to give a theory of language. A moral theory that claims that "Don't lie" and "don't commit adultery" are moral rules that you must follow to be moral, but, [it's very distressing] they cannot ever tell you what to do, might not have much appeal, theoretically or otherwise. There is no point to moral rules if they don't tell you what to do, you never know which one you follow, or if they don't help you decide what to do.

Philosophy of language, Kantian ethics, or rule-utilitarianism, each without rules, are simply pointless. A Kantian moral law is not "do whatever comes to mind".[208] May be anything is unthinkable without the fact that we can follow rules, but philosophical theories that are essentially rule based had better come up with an answer to Kripke's skepticism, since they hold that because rules determine their application they can make the claims they are making. One need not worry too much about those activities that never relied on this fact. May be no theory is ever troubled by skepticism. But this is an issue I

---

[208] I don't think a Heideggerian, to mention just one possibility, will be much affectecd by rule skepticism.

cannot discuss here.

Kantian moral theory and rule utilitarianism are therefore a prime target for Kripke's skepticism. Skepticism in general is not primarily concerned with refuting ordinary beliefs just as an aim in itself. There would be no point of raising skeptical doubts to refute such ordinary beliefs. Trying to show that skepticism also refutes and "endangers" our quotidian beliefs would not be an interesting philosophical point. Skepticism has a different target. It targets philosophical theories that try to give an explanation of knowledge, goodness, or the beautiful.

Philosophers disagree about what kind of explanations philosophy attempts to provide. If one agrees that, for example, aesthetics is the philosophical theory of the beautiful, metaphysics the theory of being as being, and semantics the philosophy of meaning, then philosophical theories attempt to establish the principles of their subject matter. Not all philosophers will agree with all of the above, but I think it cannot be said that this view <u>has to be false</u>. Similarly, it seems plausible to hold that ethics or moral theory is a theory of the right or the

good.[209] Ethics or moral philosophy is thus a theory of moral principles, or what makes something right viz., an explanation of certain properties, and of how to act morally. Perhaps not everybody sees morality this way, but I think it is plausible to characterize it this way.

Briefly then, Kant held that an act is morally good if the will is determined by the principles of pure practical reason.[210] I think it is also at least possible to think that some rule utilitarians hold that a person's action have a certain moral property when the person follows the rules specified by the utilitarian moral system. The rules are of a particular type, since rule utilitarianism is a consequentialist theory, but I believe the above characterization of rule utilitarianism is at least a defensible one. The rules of utilitarianism must be identifiable.

If one thus assumes that moral theory tries to prescribe an action, then rule utilitarianism explains the moral value of the action as resulting from following a particular rule (say, rule x). It is because I follow a certain rule that

---

[209] I am aware that those term designate also different parts of ethics, moral theory and a theory of justice. But I may be permitted to make the general point while ignoring this distinction.

[210] This is different from what makes a human action good. A human action is never really good in Kant's view. Only the good will is good in the unqualified sense.

the action is moral. This type of moral theory is the target skepticism, because skepticism claims that this moral theory fails since rule following is impossible. Rule following is specific and essential to this theory, in a way that is analogous to how rule following is essential and specific to modern semantics. When the skeptic targets rule following and, as a consequence, theories of meaning (as Kripke's skeptic does), he targets the philosophical explanation that the meaning of a term consists in the rule we follow when we apply a term. The result of the skeptical challenge is that the meaning theorist who thought that meaning consists in following a rule is left with nothing whatsoever, since rules are no more 'determined' than the meaning which was to be determined by rules.

Similarly, if a rule utilitarian can be, not must be, construed as holding that the moral value of an action consists in the fact that we follow rules of a certain type, rule skepticism shows that following rules does nothing to establish a moral value, since I might follow any rule, of follow a rule in any which way I please. This skeptical argument takes away the whole point about rules in moral philosophy. Unless we are prepared to hold that rule utilitarianism has no philosophical point in having rules as part of a moral theory, rule skepticism has a *prima facie* relevance for rule utilitarianism. This does not deny that

"meaning skepticism" will also be applicable to rule utilitarianism, since meaning skepticism would deny the stability of concepts. But epistemic skepticism also applies to rule utilitarianism because I could always doubt that there are actions; other mind skepticism, since I doubt that there are other sentient creatures, etc. But clearly, skepticism must have specific points towards specific issues in philosophy.

It is perfectly consistent with rule skepticism to deny that concepts are rules; holding that meaning is stable is thus compatible with rule skepticism. Not everyone thinks that concepts are rules, and there is no necessary connection between the concepts 'rule' and 'meaning'. It is thus possible to consider rule skepticism gua rule skepticism, and not gua meaning skepticism. A wide range of moral theories do not rely explicitly on rules. The application of rule skepticism to theories other than semantics is pertinent if those theories rely essentially on rule following. Denying that concepts are rules does not deny that morality is a system of rules. Rule skepticism does therefore not directly challenge meaning. But rule skepticism directly challenges rule utilitarianism. If we considered rule skepticism decisive, we could then still opt for, e.g., act utilitarianism which does not require that we follow rules, at least not in the way rule utilitarianism as

described and characterized does.

Does the rule skeptic also apply to all sorts of rules that we employ in everyday life? Of course it does. Kripke explicitly states this. But of course, no person in the street ever thought of giving a philosophical explanation of what morality or meaning consist in (that to speak a language, or to be moral, essentially involves rule following). The meaning of a term, for an ordinary person, will be given in terms of synonyms, translations etc., since no ordinary person usually formulates a theory of meaning. Similarly, good or right acts are usually understood to be acts of a certain type, and there may be no reference to rules or universalization. Most people, in my experience, think that moral behavior is rather personal, and not based on rules.

Of course, skepticism also could be used to show that "the person in the street" has no knowledge, or that the application of rules is "a leap in the dark"; skepticism thus applies *mutatis mutandis* to the ordinary person's beliefs. But making skepticism target the views of an ordinary person is quite obviously to give the skeptic the wrong audience. It is the putative philosophical justification and stipulation that rule following is involved in those activities that the skeptic attacks. Ordinary beliefs, and ordinary rule following, is doubted

*pari passu*, but is not the skeptic's target. I thus tried to show how skepticism challenges those theories for which it matters.

This view is corroborated by the response to Kripke's skepticism. Of course the skeptical challenge can be applied to baseball rules. Up until yesterday you were out by the third strike, and you thought that you struck out. But it is your lucky day! The umpire followed the rule "shmout" after the third strike, which means you are our after the third strike, unless the pitcher is a southpaw from Boston. But was there a general outcry that baseball rules were simply no longer applicable because a Princeton philosopher argued that rules no longer determine their application? There was not. Baseball rules are just not of philosophical interest. Moral rules are, and mathematics is as well. There was a slew of articles written by philosophers of language and mind that tried to refute the skeptic. The reason for this is obviously that philosophy of language has, especially since Frege, relied on functions in order to express its theories. But this means those philosophers think that the functions are safe and objective, and they have tried to construe meaning in terms of those functions, in order to have a theory of meaning that is finitely computable. The same holds for morality, where rules, according to Kant, are necessary, universal,

and a priori. In the same manner, the type of rule utilitarianism I characterized can be understood to rely on rules to determine the moral value of an action. If application of a rule were a leap in the dark, there would be no point to those theories.

Accordingly, theories of language and morality that rely heavily on rules seem to be the proper targets of rule skepticism. Kripke's argument apparently allows us to conclude that there is neither meaning, nor morality, as long as semantic and ethical theories make essential reference to rules. Kripke's meaning skepticism follows from the skepticism about rules if and only if one adds the assumption essential to Model M, that concepts are rules. Model M makes those assumptions for the reasons that I stated in Chapter One. There is a very well understood theory (function theory), and it seems to be a nice model to help explain language. Language is thus treated as if it were a calculus. The functions contemporary philosophers are talking about are very special sort of rules. But there is a far more natural application of the rule following skepticism, which Kripke presented in the banal terms of *addition*, to rule based moral theory. Those who promote the view that morality is a system of rules will be the immediate target of the skeptic's conclusion about rules.

I thus have tried to show in the first chapter that

modern semantics defines rules primarily in mathematical or causal/mechanical terms. The application of the rules is held to be unproblematic precisely because those rules and functions are thought to absolutely and universally determine their application. In the second chapter I have tried to argue for an interpretation of Kant's practical philosophy that shows that Kant also relies on rules as a priori, necessary and therefore certain grounding of morality. Kant thinks that the objectivity and universality of morality can be derived from an a priori and necessary source, pure practical reason, which is the source of principles. This is precisely the "fact" that Kripke challenges philosophers to present for rule following. It can be seen as the "mental fact" which Kripke discusses explicitly.[211] Similarly, I argued that there are some versions of rule utilitarianism that rely on rules because of their objectivity and certainty. The theories assume there is no problem with rule following, just as pre-Humean philosophers thought there is no problem to claim that the causal relation was a necessary one. That does not mean that trying to give a function based semantic or rule based moral theory does not get its philosophical impetus from the

---

[211] Kripke discusses a Kantian version of the rule following problem on page 62: "How is language possible". It is easy to see that this can be translated into: How is morality possible? In both cases, Kant would resort to a priori principles.

fact that rules are supposed to be the source of objectivity.

Most certainly, rule utilitarianism has other essential elements, foremost the consequences its rules are to promote. Its rules are rules to achieve certain ends (happiness, social welfare, etc.). One might note that the dissertation only touches on certain aspects of rule utilitarianism, and the relevance of rule skepticism for rule utilitarianism may be a border-line case. But that does not mean that there is no case to be made. Rule utilitarianism relies heavily on rules, and rules are thought to take capriciousness and individual deliberation out of the equation. They are thought to guarantee objectivity because rules determine how an act has to be done if I act according to a certain rule. Although rule utilitarians rely on rules to a varying degree, some versions of it rely on rules rather heavily. Those versions are, I find, obvious targets for rule skepticism.

Especially in the first chapter, I have talked about the *mechanical* application of rules. In general, the application of algorithms is considered "mechanical" in the sense that function theory relies on the fact that there could be a mechanical procedure for computing the value of a formula. Even though a "Turing Machine" is merely a theoretical device, it could be constructed, and computers

are the next best thing. The "mechanical" picture of rules relies on computability theory or the theory of algorithms; algorithms are finite, and purely combinatorial procedures by which the values of a function are obtained.[212] Modern semantics, again, relies on such function theoretic views. One might object that no such function theoretic views are to be found in Kant or in rule utilitarianism. Maintaining that Kant held rules to be recursive functions is a clear anachronism. But Kant clearly knew of Leibnitz' universal language, and his calculus. In the first *Critique*, for example, Kant explicitly states that "concepts rest on functions" (CPR B 93). The idea that concepts (including moral concepts) are functions is not the only one that points to the fact that Kant thought about rules and laws as determining their use. After all, Kant's theory tries to show how the necessary and a priori rules of rationality *explain* the necessity of causal interaction and moral duty.[213] He thus writes:

---

[212] This explains the extensive discussion of finiteness in Davidson, Kripke, and others. For the notion itself, see e.g. Davis (1967) and Pollock (1990) p. 102.

[213] For Kant, the types of rules, laws and principles I am concerned with are all necessary, a priori, and universal. Notice especially the reference to the terminology Kant uses, which is a function theoretic term ("exponent").

> That there should be principles at all is entirely due to the pure understanding. Not only is it the faculty of rules in respect of what happens, but is itself the source of principles according to which everything that
>
> (continued...)

> I assume that there really are pure moral laws which determine completely a priori ... what is ... to be done ... ; and that these laws command in an *absolute* manner (not merely hypothetically, on the supposition of other empirical ends), and are therefore in every respect necessary. (CPR B 835; first emph. add.)

The above quote from the first *Critique* shows that Kant thought of moral principles as a priori procedures.[214] This makes Kripke's paradox especially relevant for Kant's theory, at least in my view. Does the Kripke's paradox then also apply to baseball and its rules? Kripke's paradox shows that there is a problem with rules if they are considered to guarantee an application. I have no doubt that, in baseball, this question never even matters. But it matters to Kant, and this I tried to show in the second chapter. I do believe that it would be important for Kant to answer the rule skeptic. I don't think that Mr. Steinbrenner would care to address the rule-following paradox. The application of the rule skepticism to baseball

---

[213](...continued)

> can be presented to us as an object must conform to rules. ... The laws of nature ... all ... stand under higher principles of understanding. ... These principles alone supply the concept which contains the condition, and **as it were the exponent**, of a rule in general. (CPR, B 197/8; all emph. add.)

[214] It may be the case that Kant's terminology does not favor the view that rules "mechanically" determine their application. But this is a question of terminology. Kant thinks, as the quote shows, that the rules absolutely determine their application. Kripke and Wittgenstein discuss this as a metaphor for a machine-like, mechanical application. I think this is an apt comparison. (Cf. Kripke, pp. 32-37, and *PI* 93/4.)

would be pointless. I don't think it is pointless to apply it to rule based moral philosophy. A lot of contemporary philosophers think it matters to philosophy of language. Wittgenstein thought so, too.

Rule utilitarianism evaluates the moral character of an action by first asking under which rule does it fall under. It is different from Kant's view in that consequences are taken into consideration. The rules have empirical content, and this is the main difference between utilitarianism and Kant's view. Does Kripke's paradox apply to rule utilitarianism? I think there can be no doubt. Does the skeptical conclusion that we cannot follow rules show that rule utilitarianism has lost one of its essential explanatory and justificatory elements for moral evaluation? I think the answer is a clear yes. There may be some forms of rule utilitarianism that depend less on rules. Rule utilitarians fond of mathematics, logic, and economic theory will try their best to find a formal, axiomatic system resembling systems to be found in economics and semantics. These system might even be interpreted 'realistically', viz. they not only prescribe rules to be followed but define the moral property "good" or "right" as being that of a system in a particular state, which has to be brought about by those rules being followed. Others, may be of a more "conceptual" inclination, may never work those formal

elements out, and speak of rules as "guides", or rule or thumb, or generalizations. But not all rule utilitarians see their rules as generalizations.

The former theories are then more vulnerable to the skeptical doubt that rules do not determine their application. I think this makes the application of rule skepticism to rule utilitarianism more interesting than its application to baseball, especially in the light of the role rules have in this type of moral philosophy. They serve to give a rational, objective standard for moral action. Not everyone thinks objectivity can be achieved in the moral area. Viewing morality as a system of rules that determine which action is to be followed, and how a rule is to be applied is not universally accepted. In the end it is a question of philosophical relevance, whether one highlights the fact that and how the rule skepticism is relevant for moral philosophy.

Rule utilitarianism is a theory that explicitly relies on rules to be followed to determine the moral value of an action. It is very different from, say, Aristotle's theory, who, at least according to some interpretation of Aristotle, has <u>virtues</u>, not rules, at the center of his ethics. I don't deny that rule skepticism could be applied to Aristotle as well. But I think that this would in deed be pointless:

> [R]ight reason is not some rule, or set of rules, known to the practically wise but is practical wisdom itself. Aristotle holds that an excellent character has to be guided by experience combined with deliberative and executive skills, and it must be aided by a capacity to judge how to act properly in various situations, a capacity which has to be developed in a good education.[215]

Aristotle is concerned with *right reason*, not *right rules*, and in this sense, Aristotle's ethics does not become an obvious target of Kripke's paradox.

This point can also be developed by looking at Kripke's use of Hume's skepticism. As developed by Kripke, Hume's skepticism is directed at what we presume to be the contents of our notion of e.g. causation, not the belief in causation itself.[216]

Hume, according to this view, attempted to show that the notion of 'cause' does not warrant the 'superlative fact' of a necessary connection in the things themselves. "Cause" has to be given a different explanation since, according to Hume, there is no necessity *in objecto*. This may, of course, make very little difference to a physicist who formulates laws that describe the phenomena. Physicist are simply not in the business of explaining causality. There is no such thing as a law of causality in physics. So, of

---

[215] Urmson (1988) p. 86.
[216] Kripke (1984), p. 60 ff

course, Humean skepticism leaves physicists cold.

Sometimes, however, metaphysical pictures make a physicist reject a physical theory, precisely because they are not "deterministic".[217] And I believe that moral philosophers are like those physicists who think that metaphysical pictures decide which theories to adopt. Moral philosophers, by the nature of their business, have to rely on notions like 'person', 'good', etc., notions traditionally termed 'metaphysical'. A rule utilitarian theory should at least explain why setting a cat on fire, or wanton murder are immoral. Moral theories presumably try to explain at least some of our moral judgments, and those, some Realists may think, are due to a fact, not a feeling. A realist rule utilitarian, who thinks that the rules actually establish the property of right and wrong, should be especially concerned about the rule skepticism.

Thus, the rule utilitarian could be pressed to give an answer to the question "Now that we have a rule for the right action, what is it that determines the application of the rule?" This question simply will come up because the environment in which the rule utilitarian operates is the environment of philosophical justification, and in this environment the reliance on rules is to rely on that very

---

[217] Most famously, Einstein rejected Heisenberg's theory because he thought that 'God does not play dice'.

fact. Precisely this was the motivation of Leibnitz' idea of a universal language and Frege's *Begriffsschrift*. Leibnitz had the idea of a universal language to communicate ideas better than in natural languages, and Frege thought, similarly, that language could not be used to express thoughts without ambiguity. Frege thus invented a scientific language to avoid such problems. Modern semantics has dropped this normative element of Leibnitz and Frege, and language now is said to be a calculus.[218] This analogy between rule utilitarianism and theories of meaning, I think, is not too far-fetched. And much like Frege's idea of a calculus of meaning, the calculus of morality is seriously threatened by the skeptical challenge. The calculus is designed to determine which action is to be done. If Frege[219] and his followers are identified as the target of rule skepticism, extending this to rule utilitarianism does not strike me as being strange at all.

If rule skepticism is right, the rule utilitarian is then "no better off" than virtue ethics, often criticized for the lack of specific instructions how to act in certain situations. Rule utilitarianism, I believe, makes rules

---

[218] I summarize here what I described in the first chapter of the dissertation.
[219] As a reminder, I would like to emphasize that I am not claiming that this is historically accurate. It is the way Frege is received today. I have shown and argued this is the first chapter.

central since the application is taken for granted because of the metaphysical picture Wittgenstein thinks underlies our ordinary way of talking about rules.

Kripke thus states that Wittgenstein agrees with Hume that

> there is no such 'superlative fact' (§192) ... [that] determines in advance what I should do to accord with this [rule]. But he claims (in §§183-93), the appearance that our ordinary concept of [rule] demands such a fact is based on a philosophical misconstrual - albeit a natural one- of such ordinary expressions as ... 'the steps are determined by the formula'. (Kripke, p. 65f)

It seems plausible to me that this might have an impact on philosophers who have construed rules and systems of rules because "the steps are already taken." Relativism about moral goals is one thing, but if one were to show that the rules do not guarantee the application, a rule utilitarian seems to have lost a good deal of appeal. This view that rules are "backed up" by a metaphysical fact is, in my view, simply an old trend in philosophy which is not always explicit. Rules, principles and formal systems are more likely to be "familiar without qualification" (Aristotle, NE 1095b1). I don't want to suggest that all philosophers, or all rule utilitarians, subscribe to this idea. But skepticism is a threat to the explanatory impetus of philosophical systems, and - there should be no doubt - rule utilitarians are in the business of giving just such an

explanation. One of the reasons that rules, procedures, and formalization have had such an impact in philosophy is because of the idea that formal elements are just one step short of a further, more ultimate justification, an idea to be found in Plato's *Republic* (510b-c). This ultimate justification for rules was what the skeptic calls for, and he concluded that there was none. Wittgenstein rejects the view that such a justification need to be given, especially since he holds that the notion of a rule, like any other notion, is a family resemblance concept. It is Wittgenstein who tries to not give a theory of why there is or is no such foundation, but who shows us where we go wrong when we are asking for one.

## References


Aristotle *Nicomachean Ethics* Tr. W.D. Ross revised by J.O. Urmson in *The Complete Works of Aristotle; The Revised Oxford Translation,* vol. 2 J. Barnes (ed.) Princeton: Princeton University Press

Austin, J.L. (1961) *How to Do Things with Words* Oxford: Oxford University Press

Baker, G.P., and Hacker, P.M.S. (1984) *Language, Sense and Nonsense* Oxford: Basil Blackwell

Barnes, J. (1971) "Utilitarianism" in *Ethics* **82**

Baumrin, B. (1976) "Autonomy in Rawls and Kant" in *Midwest Studies in Philosophy,* vol. 1, p. 55-57

———, (1982) "Two Concepts of Justice" *Midwest Studies in Philosophy* vol. 7, pp. 63-72

Boghossian, P. (1989) "The Rule Following Considerations" *Mind* vol. 98, pp. 507-49

Boole, G. (1854) *The Laws of Thought* (reprinted 1958, New York: Dover)

Brandt, R. (1979) *A Theory of the Good and the Right* Oxford: Clarendon Press

Canfield, J.V. ed. (1986) *The Philosophy of Wittgenstein Vol. 6: Meaning* New York & London: Garland Publishing, Inc.

Cavell, M. (1986) "Metaphor, Dreamwork and Irrationality" in Le Pore (1986) pp. 495-507

Chomsky, N.(1980) *Rules and Representations* New York: Columbia University Press

Davidson, D. (1963) "Action, Reason, and Causes" reprinted in Davidson (1982), p. 3-20

———, (1965) "Theories of Meaning and Learnable Languages" reprinted in Davidson (1984) pp. 3-15

———, (1977) "Reality without Reference" reprinted in Davidson (1984) p. 215-25

———, ( 1978) "Intending" reprinted in Davidson (1982), pp. 83-102

———, (1982) *Essays on Action and Events* Oxford: Clarendon Press

———, (1984) *Inquiries into Truth and Interpretation* Oxford: Clarendon Press

Davidson, D. and Harman, G., eds. (1972) *Semantics of Natural Languages* Dordrecht, Holland: D. Reidel Pub.

Comp.

Davis, M. (1967) "Recursive Function Theory" in Edwards, ed. (1967), v. 7, 89-95

Descartes, R. (1641) *Meditations on First Philosophy* (1641), in Haldane, E.S. and Ross, G.R.T., eds. (1978) *The Philosophical Works of Descartes* Vol 1 (Cambridge, Cambridge University Press). p. 155

Devitt, M. and Sterelny, K. (1990) *Language and Reality An Introduction to the Philosophy of Language*. Cambridge, Mass.: The MIT Press

Dretsky, F. (1984) *Knowledge and the Flow of Information* Cambridge, Mass.: The MIT Press

Dummett, M. (1967) "Frege" in Edwards (1967), vol 3. pp. 225-37

———, (1981) *The Interpretation of Frege's Philosophy* Cambridge, Mass.: Harvard University Press

———, (1973) *Frege: Philosophy of Language* (New York: Harper & Row)

———, (1978) "The Philosophical Basis of Intuitionistic Logic" in Dummett (1978) p. 215-247

———, (1978) *Truth and Other Enigmas* Cambridge, Mass.: Harvard University Press

Dworkin, R. (1977) *Taking Rights Seriously* Cambridge, Mass.: Harvard University Press

Edward, P. ed. (1967) *Encyclopedia of Philosophy* (New York: Macmillan Publ. Co.) 8 volumes

Evans, G. and McDowell, J., eds., (1976) *Truth and Meaning* Oxford: Oxford University Press

Floyd, J. (1991) "Wittgenstein on 2.2.2....: The Opening of Remarks on the Foundations of Mathematics" *Synthese* **87**: pp. 143-180

Fodor, J. A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* Cambridge, Mass.: The MIT Press

———, (1990) *A Theory of Content* Cambridge, Mass.: The MIT Press

Foster, J. (1976) "Meaning and Truth Theory" in Evans and McDowell, eds. (1976)

Frege, G. (1919) "Notes for Ludwig Darmstaedter" in Frege (1979)

———, (1979) *Posthumous Writings* Hermes, H., et al. eds., Chicago: University of Chicago Press

———, (1892) "On Sense and Reference" reprinted in: Moore, ed., (1993) pp 23-42

———, (1892) "On Concept and Object" reprinted in Geach and Black eds., (1980)

French, P.A. et al. eds. (1981) *The Foundations Of Analytic Philosophy* Midwest Studies in Philosophy, vol. 6 Minneapolis: University of Minnesota Press

Funke, G. (1974) *Akten des 4. Internationalen Kant-Kongresses* Berlin, New York: Walter De Gruyter

Geach, P. and Black, M., eds., (1980) *Translations from the Philosophical Writings of Gottlob Frege* Totowa: Rowman and Littlefield

Gibbard, A. (1965) "Rule Utilitarianism: Merely an Illusory Alternative?" in *Australasian Journal of Philosophy*, 43, 211-220

Goldfarb, W. (1983) "I Want You to Bring Me a Slab" *Synthese* **56** (1983) 265-82

Goodmann, N. (1955) *Fact, Fiction, and* Forecast Cambridge, Mass.: Harvard University Press

Grice, H. (1957) "Meaning" *Philosophical Review*, 66, pp. 377-388

Hare, R.M. (1976) "Ethical Theory and Utilitarianism" Reprinted in Sen and Williams (1982), pp. 1-38

———, (1981) *Moral Thinking Its Levels, Method, Point* Oxford: Clarendon Press

Harsanyi, J.C. (1977) "Morality and the Theory of Rational Behavior" Reprinted in Sen and Williams (1982), p. 39-62

Henrich, D. (1975) "Die Deduktion des Sittengesetzes. Über die Gründe der Dunkelheit des letzten Abschnittes von Kants 'Grundlegung zur Metaphysik der Sitten'" in Schwan, A. (1975) *Denken im Schatten des Nihilismus* Darmstadt

Herman, B. (1993) *The Practice of Moral Judgment* Cambridge, MA: Harvard University Press

Hintikka, J. (1976) "Language Games", in Canfield (1986) vol. 6: 231-252

Hunter, G. (1971) *Metalogic. An Introduction to the Metatheory of Standard First Order Logic* Berkeley and Los Angeles: University of California Press

Kant, I. *Prolegomena zu einer jeden künftigen Metaphysik (1783)* trans. P. Carus and revised by L.W. Beck as *Prolegomena to any Future Metaphysics* Indianapolis:

Bobbs-Merrill Library of Liberal Arts, 1950

———, *Kritik der reinen Vernunft* (2. ed. 1787) trans. Norman Kemp Smith as *Immanuel Kant's Critique of Pure Reason* New York: Macmillan, St. Martin's Press, 1965

———, *Grundlegung der Metaphysik der Sitten* (1785), tr. Lewis White Beck as *Foundations of Metaphysics of Morals* Indianapolis: Bobbs-Merrill Library of Liberal Arts, 1959

———, *Kritik der praktischen Vernunft* (1788) tr. Lewis White Beck as *Critique of Practical Reason* Indianapolis: Bobbs-Merrill Library of Liberal Arts, 1956

———, *Kritik der Urteilskraft* (1790) tr. J.H. Bernard as *Critique of Judgment* New York: Haffner Liibrary of Classics, 1951

Katz, J.J. (1990) *The Metaphysics of Meaning* Cambridge, Mass.: The MIT Press

Kim, W. (1990) *Introduction to Object-Oriented Databases* Cambridge, Mass.: The MIT Press

Korsgaard, C. (1985) "Kant's Formula of Universal Law" in *Pacific Philosophical Quartely* **66** pp. 24-47

Kripke, S. A. (1982) *Wittgenstein on Rules and Private Language. An Elementary Exposition* Cambridge, Mass: Harvard Universtiy Press

Leibnitz, G.W., (1951) "Dialogue on the Connection between Things and Words" in *Selections* Wiener, P.P., ed., New York: Scribners

LePore, E. ed. (1986) *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson* Oxford, New York: Basil Blackwell

Lewis, D. (1972) "General Semantics" in: Davidson and Harman, eds. (1972) pp. 169-182

MacIntyre, A. (1981) *After Virtue: A Study in Moral Theory* London: Duckworth

McDowell, J. (1993) "On the Sense and Reference of a Proper *Name"* in Moore (1993) pp. 111-136

McGinn, C. (1984) *Wittgenstein on Meaning* New York: Basil Blackwell

Mill, J.S. (1843) *A System Of Logic. Ratiocinative and Inductive* ch. 6, section 2 in: Nagel, E., ed. (1950) John Stuart Mill Philosophy of Scientific Method New York: Hafner Publishing Co., p. 163

Millikan, R. (1986) *Language, Thought and Other Biological Categories* Cambridge, Mass.: MIT Press

Moore, A.W., ed., (1993) *Meaning and Reference* Oxford: Oxford University Press

Moore, G.E. (1903) *Principia Ethica* Cambridge: Cambridge Universtity Press

Nagel, T. (1970) *The Possibility of Altruism* Princeton: Princeton University Press

O'Neill, O. (1989) *Constructions of Reason* Cambridge: Cambridge University Press

Plato *Cratylus* in: Hamilton, E. and Cairns, H., eds (1980) *Plato The Collected Dialogs* Princeton: Princeton University Press

———, *Republic* Transl. P. Shorey in op. cit.

Platts, M. (1979) *The Ways of Meaning: An Introduction to the Philosophy of Language* London: Routledge and Kegan Paul

Popkin, R. (1967) Skepticism in Edwards, P. (1967), p. 449f.

Pollock, J.L. (1990) *Technical Methods in Philosophy* Boulder, Colorado: Westview Press

Putnam, H. (1978) *Meaning and the Moral Sciences* Boston: Routledge & Keegan Paul

Quine, W.V.(1953) *From a Logical Point of View* Cambridge, Mass.: Harvard University Press

———, (1960) *Word and Object* Cambridge, Mass.: The MIT Press

Rawls, J. (1955) "Two Concepts of Rules", Philosophical Review 64

———, (1980) "Kantian Constructivism in Moral Philosophy: The Dewey Lectures 1980" in *Journal of Philosophy* 77, 9 (Sept. 1980), 512-72

———, (1985) "Justice as Fairness: Political, not Metaphysical" in *Philosophy and Public Affairs* 14, no.3 (Summer 1985): 223-51

Regan, D. (1980) *Utilitarianism and Cooperation* Oxford: Clarendon Press

Resnik, M.D. (1981) "Frege and Analytic Philosophy" in: French et al., eds. (1981) pp. 83-103

Rhees, R. (1959) "Wittgenstein's Builders" in *Aristotelian Society Proceedings*, 1959-60, 171-186

Ricketts, T. G. (1986) "Generality, Meaning, and Sense in Frege" *Pacific Philosophical Quarterly* **67** 172-195

Salmon, N. (1986) *Frege's Puzzle* Cambridge, Mass.: The MIT Press

Schwarzenbach, S. (1991) "Rawls, Hegel, and Communitarianism" in *Political Theory Vol. 19, No.* 4, pp. 539-571

Schiffer, S. (1987) *Remnants of Meaning* Cambridge, Mass.: The MIT Press.

Sen, A. and Williams, B. (1982) *Utilitarianism and beyond* Cambridge, UK: Cambridge University Press

Silber, J.R. (1974) "Verfahrensformalismus in Kants Ethik" in Funke (1974)

Smart, J.J.C. and William, B. (1973) *Utilitarianism For and Against* Cambridge: Cambridge University Press

Sobel, J.H. (1968) "Rule Utilitarianism" *Australasian Journal of Philosophy* 46

Tarski, A. (1944) "The Semantic Conception of Truth" in *Philosophy and Phenomenological Research*, 4, pp. 341-75

Urmson, J.O. (1988) *Aristotle's Ethics* New York: Basil Blackwell

Vermaazen, B. (1967) Review of Katz, J.J. and Postal, P., *An Integrated Theory of Linguistic Descriptions.* and Katz, J.J. *Philosophy of Language, Synthese* **17.**

Wheeler, S.C. (1986) "Indeterminacy of French Interpretation: Derrida and Davidson" in LePore (1986), pp.477-495

Wittgenstein, L. (1953) *Philosophical Investigations* New York: Macmillan Publishing Co.

———, (1974) Philosophical Grammar (ed. R. Rees, tr. A. Kenny), New York: Basil Blackwell

———, (1978) *Remarks on the Foundations of Mathematics* G.v. Wright et al. eds, revised edtion Cambridge, Mass.: The MIT Press

———, (1979) *Wittgenstein's Lectures: Cambridge, 1932-1935* Ambrose, A. (ed.), Chicago: The University of Chicago Press

———, (1961) *Tractatus Logico Philosophicus*, tr. D.F. Pears and B. McGuinness, New York: Routledge