University of Alberta

Rule-Following, Meaning, and Thinking about Thought

by

Julian Fowke

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

Master of Arts

Philosophy

Fall 2009
Edmonton, Alberta

Examining Committee

Bruce Hunter, Philosophy

Adam Morton, Philosophy

Bernard Linsky, Philosophy

Anne-Michelle Tessier, Linguistics

August 26, 2009

Abstract

What is it that determines what our thoughts mean, and how do we *know* what they mean? Obviously a thought must mean something in order for us to know what it means, and frequently we *do* know the content of our thoughts. But the converse does not hold, for we can have thoughts to whose contents we are entirely obtuse. For example, people frequently do things for reasons contrary to those they might sincerely profess. But if meaning and thought are not epistemically given, how did we develop awareness of them at all? How do we think about thought—both our own and that of others? This essay is an exploration of these issues to the end of understanding how it is that we come to be able to represent our purposes, intentions, and meanings.

Contents

Introductory

§ I: Introduction

Thinking about thought is not merely a philosopher's pastime. There is a sense of 'thinking about thought' according to which it is something that people do frequently in the course of their daily lives. Giving a reason for having taken a particular course of action involves having a thought about the prior thought which prompted the action. That is not to say that every verbal performance is preceded by an internal, mental act of reflection. Oftentimes, when one is speaking candidly, the vocalization and the thought it expresses are indistinct—the thought and its verbal expression are the same. But even when one does not silently deliberate before giving answers to such questions as "why did you do *this* rather than *that*?" or "what did you *mean* in saying that?", the act of giving a reason for one's past actions (even if the act in question occurred only a moment ago) can be said to be a thinking—a thought episode—about one's thought.

The purpose of this project is to explore the question of how we think about thoughts—both our own and those of others. In more perspicuous terms, the question concerns how a certain class of intentional*[1] acts (namely, our thoughts) can be *about* other intentional* acts. How do we represent intentionality to ourselves? A primary goal of this essay will be to show that the ability to think does not guarantee the ability to think about thought. Philosophers following in the footsteps of Descartes and Brentano have often ignored or denied certain subtleties of the question of thought-about-thought, assuming that simple

[1] Since 'intentional' can be equivocally applied to both acts of intending as well as phenomena that exhibit intentionality, I will use 'intentional*' to signify references to the latter.

possession of intentionality is the sole requirement for awareness of intentional* states or episodes—for the ability to cognize about them. I aim to argue that the Cartesian picture is wrong, and that an answer to the question of what it is to think does not also suffice as an answer to the question of what it is to think about thought. Although both questions fall under the purview of the philosophy of mind, the former is of a more metaphysical design, while the other is more epistemological in nature. As such, I will refer to them respectively as the metaphysical and epistemological problems of thought.

To a first approximation, the metaphysical question can be put in the following way: what features of a person make it such that she or he can think? In what does the *ability* to have thoughts consist? (Of course, this would include the ability to have thoughts that are about other thoughts.) In asking this question, I am assuming that it cannot be answered solely by a neuro-physiological description. I am presupposing that a complete metaphysics of thought involves something more than a description of the brain-processes through which it is realised. Just what that something more might be will be the primary focus of the second chapter of this essay. The epistemological problem, also roughly put, is as follows: how are we aware of thought? How do we come to be able to think about thought? Kant's notion of 'condition of possibility' is an appropriate term for giving shape to what I am calling the epistemological problem of thought, for while these questions are *in some way* epistemological, they do not specifically concern how we can have evidence or be justified in holding beliefs about our thoughts (though these issues are in the near proximity); rather, they concern the

development of concepts that make thinking about thought possible. At this point, all of these questions are still quite vague and ambiguous. Indeed, it might appear that the latter questions are merely restatements of the former, and that I am confused in claiming that there is any distinction at all between the questions of how we think and how we think about thought.

To dispel the apparent ambiguity, it may be useful to look at another, similarly ambiguous question. Consider, for example, the question 'how do you walk?' The average, competent walker would be perfectly right to give an answer along something of the following lines: "Well, in order to walk, I just put one foot in front of the other".[2] The near-analyticity of such answers makes them of little interest as far as explanations go, but they are (not insignificantly) representative of the self-understanding of one who possesses a fairly complex ability. Of course, an equally correct answer would involve a more detailed biomechanical explanation that makes reference to, among many other things, rocking back and forth between the feet, pushing with the toe to maintain speed, combined interruption in rocking and ankle twist in order to turn, and shortening and extending the knees to prolong the 'forward fall'.

What I am calling the epistemological question of thought roughly corresponds to the way in which the average walker interprets the question 'how do you walk?' It concerns the self-understanding of thinkers who possess the ability to think about thought *qua* thought. The concomitant metaphysical

[2] It is significant that not all competent walkers would be capable of giving even this sort of answer to the question. Many toddlers, for instance, clearly know how to walk but are incapable of explaining how they do so. They may not even have the faintest clue as to what walking is. That one can possess an ability without being aware of it or knowing how he or she succeeds at it will be a central point of emphasis throughout this project.

question, then, corresponds to the biomechanical interpretation of the question 'how do you walk?' It requires a theory that explains how we are able to represent things by thought.

For those who remain unconvinced that the questions I have highlighted pertain to distinct categories, I can only ask you to bear with me. The primary goal of this chapter and the next is to show not simply *how* the questions pertaining to the metaphysics and epistemology of thought are respectively different, but to show that they *must* be different. In sum, this will amount to the claim that we are not *aware* of what we think (or even *that* we think) simply in virtue of the fact that we *do* think. Thought is not epistemically given. But if awareness of thought is not guaranteed simply by the having of it, then, on the positive assumption that we *are* conscious of our thoughts—at least to some degree—how is this possible? How do we come to be able to conceive of and know thought? This question will be the topic of the third chapter.

As social, rational beings, the practice of giving reasons for *ourselves*—for the actions we perform and the utterances we make—features daily in our lives. It is my hope that the discussion and arguments presented here will shed new light on, among other things, the oft-observed fact that while knowledge of our thoughts is not impossible or even uncommon, people frequently do things for reasons other than those they themselves might give, even when being sincere.

§ II: Framework, Further Assumptions, and Argument Structure

Doing philosophy effectively requires finding or creating (most often it is a bit of

both) a framework whose vocabulary is at once sufficiently general and sufficiently precise for the task at hand. To the end of reining-in the vagaries of the preceding section, I propose to frame my discussion of thought-about-thought in the terms of the Kripke-Wittgenstein paradox. In doing so, I am committing myself to the view that thought and the linguistic performances in which it (sometimes) culminates are both forms of rule-governed behaviour. Indeed, in the third chapter I will argue that there is an essential connection between intentionality and normativity. This in turn commits me to a denial of both the meaning scepticism and the so-called 'sceptical solution' Kripke extrapolates from Wittgenstein's *Philosophical Investigations*.[3, 4]

The conclusion of the sceptical paradox is that there is no fact of the matter as to what we mean by our words or what we intend when we purport to follow a rule; in short, our intentional* episodes or acts (i.e., thoughts) have no determinate content—they do not *mean* anything. The sceptic's search for meaning comes up empty-handed. Since intentionality is an essential ingredient in what makes a thought a thought, the very possibility of thought is among the meaning sceptic's would-be victims. The sceptical conclusion appears to render the idea of a thought utterly vacuous, for what could thoughts be if not bearers of meaning? The implications for the possibility of thought-about-thought are

---

[3] L. Wittgenstein, *Philosophical Investigations* [hereafter *PI*], trans. G.E. Anscombe, (Oxford: Blackwell, 2001).

[4] A common caveat: in the introduction to his monograph, Kripke writes, "…the present paper should be thought of as expounding neither 'Wittgenstein's' argument nor 'Kripke's': rather Wittgenstein's argument as it struck Kripke, as it presented a problem for him." (S. Kripke, *Wittgenstein on Rules and Private Language* [hereafter *WRPL*], (Oxford: Basil Blackwell, 1982), 5.) Kripke did not want his readers to think that the sceptical paradox and solution were either exegetically faithful or representative of his own views. Since I often put the arguments in Kripke's voice simply for ease of exposition, please keep his caveat in mind.

obvious: if thoughts bear no meaning, then the term or concept 'thought' does not refer to anything; and if that is the case, the question of thought-about-thought does not even arise for there is no thing for the second-order thought to be a thought of.

Though I will devote a considerable amount of space to expositing and discussing the sceptical paradox and even briefly outline the solution Kripke ascribes to Wittgenstein, my general approach will be to assume that the scepticism about meaning requires dissolution rather than concession. The specifics of how a general discussion of rule-following pertains to language, thought, and thought-about-thought will become clearer throughout the course of the project. The ultimate purpose of framing my project around the Kripke-Wittgenstein paradox is to both clarify and motivate the claim of the preceding section, namely, that what I referred to as the metaphysics and epistemology of thought *must* be distinct. Indeed, I will argue that a running together of the two is, at bottom, the true source of the paradox.

After outlining the sceptical paradox, I will spend the remainder of the next chapter exploring some specific requirements that any purported dissolution of the paradox must meet. The second and third chapters will be focused, respectively, on the metaphysics and epistemology of thought and thought-about-thought.

Chapter I: Meaning Scepticism

§ I: The Kripke-Wittgenstein Paradox

There are at least two ways in which one might non-accidentally follow a rule. First, one might non-accidentally follow a rule without an explicit or conscious intention of doing so. For example, apparently one thing all competent swimmers do in order to stay afloat while swimming is to retain more than the usual amount of air in their lungs each time they exhale. Unless I happen to stay afloat in some exceptional way, it is likely that I follow the rule 'do not fully exhale' while swimming. While it *is* my intention to stay afloat when I swim, I am fairly certain that it has never been my explicit intention to follow the rule 'do not fully exhale while swimming'. At best, we might say that I *implicitly* intend to follow the rule, since following it is necessary to the realisation of my explicit intention to stay afloat. That said, I follow the implicit rule all the same, and my doing so is not accidental because my success in following it serves as a part of the explanation for how it is that I succeed in realising my explicit intention to stay afloat. While such forms of unintended rule-following behaviour may or may not accord with our usual ideas about what it is to 'follow a rule', that they are not instances of mere accidental conformity to a rule—that they can be done 'correctly' or 'incorrectly'—makes it such that there is a sense in which they are properly characterised as rule-governed.[5]

[5] Articulating the nature of the relationship between this form of non-accidental rule-following and what is perhaps the more intuitive form that involves explicit intentions to follow rules will be central to understanding the relationship between the metaphysics and epistemology of thought-about-thought.

The second form of non-accidental rule-following is, of course, that which involves explicit intentions to follow rules. Unlike my following of the rule 'do not fully exhale while swimming', an explicit intention to perform an action is one that an actor explicitly or *consciously* represents—the intention is explicitly represented *as* an intention. When I form and abide by the intention to drive on the right-hand side of the road or to apply the rule for addition to a set of two or more numbers, I am following a rule in this second sense, and it is this second way of following a rule that is the subject of scrutiny in Kripke's reconstruction of Wittgenstein's 'sceptical' argument from *PI*.

In what do such consciously represented acts of intending consist? What detail or feature is it about me that might *uniquely* determine the *content* of my explicit intentions at a given time? Take the rule-following involved in everyday conversation, for instance: it is not at all uncommon for a listener to ask of a speaker, when the former is uncertain of how to interpret a statement made by the latter, what the latter *meant* (or intended) by what she or he said. The typical response will be of the form, "What I *meant* to say was…". But even in cases where such a response satisfies the listener, it is clearly just another interpretation of the speaker's linguistic intentions and itself just one more statement open to interpretation. Since *any* statement has the potential to be interpreted in an infinite number of ways, there is no obvious reason why those preceded by 'I meant' should possess the kind of privileged authority such that they might uniquely determine a speaker's linguistic intentions.[6]

[6] And yet, such statements do possess *some* kind of special authority; it is simply not the sort that might uniquely determine the content of one's thoughts or assertions. The epistemic authority of

Perhaps when I intend to follow a rule or mean something by a word, my intention is determined by an abstract representation of the rule. For instance, when I intend to apply the rule for addition to two particular numbers, maybe my intention is determined by a representation: *x* plus *y* = *x*+*y*. But then we could always wonder how we are to interpret the abstract representation. Yet the presupposition that there is a determinate content with respect to our intentions—linguistic and non—is pervasive. The everyday currency of statements such as "What I (or he or she) *meant* to say was…" would seem to suggest that the assumption that there are *facts* about what we mean or intend is widespread. Contrary to this, however, Kripke's Wittgenstein elaborates on the abovementioned considerations and famously concludes that there cannot be *any* fact of the matter as to what anyone means when they intend to follow a rule or use an expression. There is no fact about me that justifies my use of a term at any given time.

The argument which yields this sceptical conclusion about acts of meaning or intending is relatively simple and straightforward. Since at any time there are an infinite number of rules that correspond to my present and past usage of a term but are incompatible with one another, behavioural facts about me cannot uniquely determine which rule I currently intend to follow. Goodman's infamous predicate provides a helpful illustration of this point. Suppose 'grue' applies to all green things before now and to all blue things from here on after. Based strictly on the evidence provided by my past usage of the word 'green', it is indeterminate

---

such claims is a significant feature of the role played by attributions of self-knowledge, a closely related issue which is dealt with in more detail in the penultimate and final chapters.

whether what I *meant* in past applications was 'green' or 'grue'. Even though it might be wildly implausible to suppose that I meant 'grue' by my past tokenings of the term 'green', the rule for 'grue' is entirely consistent with my past usage of 'green'. So much the worse for behavioural evidence, we might say, no one is a behaviourist in that way anymore. Surely what is *before the mind* is what correctly and uniquely determines the contents of our intentions. The problem with this move should be obvious: as was my overt behaviour, any introspectible facts about my occurent mental states are open to an infinite number of incompatible interpretations. Suppose that when I learnt 'green' in the past, I had a mental image of green and instructed myself to call something green *only* if it was the same colour as the image. But such instructions are of no help, for now instead of imagining an absurd interpretation of 'green', we can simply do the same for 'colour'. It seems to follow that there simply is no fact of the matter as to what I mean. And furthermore, the problem is completely general: in Wittgenstein's words, "no course of action could be determined by a rule, because every course of action can be made to accord with the rule".[7] For any action, a rule can be interpreted in such a way that the action accords with it. The sceptical conclusion does not just render meaningless the notions of 'rule', 'correct' and 'incorrect', and even 'action' (in the sense of a deliberate undertaking), but so too the notion of *meaning* something by a word. And if one cannot mean something by a linguistic expression, then surely the expressions themselves do not mean anything either.

[7] Wittgenstein, *PI*, §201.

In short, the sceptical argument purports to constitute a *reductio ad absurdum* of any and all forms of semantic realism. Though the precise form of the meaning-constituting facts whose existence is being denied will become clearer through the course of the ensuing exposition of Kripke's argument, as a rough first approximation, we can think of the missing facts as of the sort tied to classical conditions for correspondence truth. On the correspondence view, each declarative sentence corresponds to a possible fact in the world by way of some kind of isomorphism. If the fact obtains, the sentence is true; if the fact does not obtain, the sentence is false. But if Kripke's Wittgenstein is right and there are no facts to which meaning claims might correspond—if there is no determinate content to language or thought—then it is certainly impossible for there to be any kind of correspondence relation between sentences and the world. The rules of isomorphism which would determine the truth or falsity of a statement are nowhere to be found. If meaning is truth and truth is correspondence between language or thought and the world, then there must be determinate facts concerning what language and thought are about—there must be something in the world that makes meaning statements true, and presumably meaning derives from the linguistic intentions of language users. The sceptical paradox attempts to show that, on the assumptions of the correspondence framework, the linguistic intentions which are supposed to ground meaning are chimerical. In denying that there is any fact about me which might justify claims such as "I mean (or meant) such-and-such", the goal of the sceptical paradox is to reveal the ultimate incoherence of the correspondence theory of truth.

To make the source of the problem more intuitive, Kripke begins his exposition with a simple mathematical example that begins with a seemingly unproblematic outline of what it is to learn and follow a rule:

> By means of my external symbolic representation and my internal mental representation, I 'grasp' the rule for addition… 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. This is the whole point of the notion that in learning to add I grasp a rule: my past intentions regarding addition determine a unique answer for indefinitely many new cases in the future.[8]

Given that 1) addition is defined over all pairs of positive integers and applies to infinite pairs of arguments, and 2) that I have only performed a finite number of addition computations, it is guaranteed that there is a sum which I have never computed before whose arguments exceed all arguments from my previous computations. Let us assume, following Kripke, that '68+57' fits this criterion. We are assuming, for the sake of argument, that I have never added numbers as large as 57. I now compute 68+57 and arrive at the answer of 125.

Now we are to imagine that I meet a sceptic who demands a justification for the answer I have just given. How do you know, he asks, that the answer you really ought to have given was not '5'? More specifically, how do you know that the rule you *actually* intended to follow when it was your intention to compute sums in the past was the rule for addition, and not *quaddition*, defined as

$$x \text{ quus } y \qquad \begin{array}{l} = x+y, \text{ if } x, y < 57 \\ = 5, \text{ if } x \geq 57 \text{ or } y \geq 57? \end{array}$$

The scepticism on offer here, Kripke explains, is not meant to pose a problem about arithmetical or mathematical knowledge specifically; rather, the problem is

[8] Kripke, *WRPL*, 7-8.

meta-linguistic. If the sceptic was questioning my knowledge of arithmetic, then I could respond to his doubt by giving a mathematical proof which begins with the relevant axioms and culminates in showing that the sum of 68 and 57 is 125. But the sceptic is not questioning how I can be sure that 125 is the correct answer to 68+57, he is questioning whether addition is in fact the rule I followed in the past when intending to compute sums. (Restricting the scepticism to my past performances is merely an argumentative device for explicating the paradox. If the sceptic's doubt is unanswerable, then not only will there be no fact of the matter as to whether I was adding or quadding in the past, but neither will there be any fact about my present intention.)

Having given myself a set of instructions or rules of thumb to guide my application of the rule for addition fails to answer the rule-following sceptic since these instructions—which are rules as well—would likewise be open to both standard and quus-like interpretations. Suppose, for instance, that when I wish to add *x* and *y*, I give myself primitive directions of the following form:

> Take a huge bunch of marbles. First count out *x* marbles in one heap. Then count out *y* marbles in another. Put the two heaps together and count out the number of marbles in the union thus formed. The result is *x*+*y*.[9]

Such a response to the sceptic's doubt, Kripke points out, is doomed to failure. For instead of questioning what I meant in the past by 'addition', the sceptic now questions what I meant by 'count'. How do you know, he asks, that in the past you meant 'count' and not '*quount*', where 'quounting' is defined as… The sceptical worry is merely transferred rather than displaced altogether. We are left

[9] Ibid., 15.

with rules for interpreting rules with no principled end in sight to the regress of the meta-meta-…-language hierarchy.

Having never before added numbers as large as 57, it follows that my past performances are compatible with both the rule for addition and that for quaddition; they alone are insufficient to determine whether I was adding or quadding in the past. Introspectible facts—facts about my occurent mental states—do not uniquely determine my past intentions either, since any metalinguistic instructions I might have used to guide myself are themselves open to myriad incompatible interpretations. Thus, it seems that there is *no fact about me* which determines whether I intended to follow the rule for addition or quaddition in my previous computations. And if there is no fact about me which determines my past intentions with respect to addition, then surely there can be no fact which uniquely determines the content of my present intentions either. What's more, Kripke concludes, the sceptical problem is perfectly general: if the sceptic's argument with respect to my past intentions regarding addition goes through, then the conclusion generalizes completely to the effect that there is no fact—*now or ever*—as to what *anyone* means by *anything* they might say or inscribe or as to what they *intend* when they purport to follow a rule. Contrary to Kripke's outline quoted two pages prior, the sceptic concludes that there can be no such thing as grasping and following rules: whenever anyone purports to follow a rule or mean something by a word or phrase, there simply is no fact about them that determines the content of their intention; whenever anyone purports to follow a rule, they do so without justification and *blindly*.

Although much of the sceptical argument is couched in epistemic terms, Kripke urges that the scepticism on offer is not epistemological, or at least not *simply* epistemological. It is a scepticism that undercuts the very possibility of meaning and intentions. To form an intention is, among other things, to set a standard against which one's actual behaviour can be measured. Intentions have normative force. And, at the level on which we develop explicitly represented intentions such as when I intend to compute a sum, Wittgenstein's considerations show that there simply is nothing that might justify the claim that my intention is one thing rather than another. In giving a response to the sceptic, there is no determinate content to which I could point, so-to-speak, and say, "*that* is the rule I am intending to follow."

§ II: The Inadequacy of Dispositionalist Responses

According to semantic dispositionalists, the fallacy in the sceptical argument lies in its reliance upon introspective considerations concerning occurent mental states. My intention to follow the rule for addition ought to be analysed in terms of my having the *disposition*, when asked for the sum of 68 and 57, to respond with '125'. If my intention was to follow the rule for quaddition, then I would be disposed to respond with '5'. Thus, the dispositionalist holds, the proper locus of intentional* or meaning-determining facts is not situated within our occurent mental states; nor is it to be read-off the previous actions that were motivated by the intention to follow a particular rule. Rather, the rule one intends to follow at any time is uniquely determined simply by what one is disposed to do—*ceterus*

paribus—when intending to follow the rule. Dispositionalist accounts locate the fact determining that I mean 'plus' rather than 'quus' in my disposition to reply to '68+57' with '125' rather than '5'.

Kripke canvasses several objections to dispositionalism: for one, dispositions are finite, while rules such as that for addition have an infinite number of application circumstances. Even though the rule for addition is defined over all pairs of integers, my disposition to add is surely indeterminate when the numbers involved are so large that I cannot grasp them. Second, people are often systematically disposed to *make mistakes*—for instance, they can be regularly disposed to forget to carry when adding long lists of numbers. What is more, people are sometimes disposed to accept correction from others when their original application of a rule was correct. But even if dispositionalists could fill-in the *ceterus paribus* clause so as to provide non-circular responses that dealt with the problems of finitude and dispositions to err, they would still seem to simply miss the point of the sceptical argument. The fact denied by the sceptical argument must have *normative* force, and it is far from clear how dispositional facts could meet this criterion in anything but circular fashion.

Dispositionalist responses to Kripke's sceptic fail to provide a standard by which it may be judged whether what I have *in fact* done *corresponds* with what I *intended* to do. Any fact proffered in response to the sceptic must show why, if it is my intention to compute the sum of 68 and 57, I *should* answer with '125' and not '5' *regardless* of whether my dispositions coincide with the rule for addition or quaddition. "Whether my dispositions are 'right' or not," writes Kripke, the

meaning-sceptic is denying that "there is anything that mandates what they *ought* to be."[10] It is not a fact about what I *would* do (even if circumstances were completely ideal and my dispositions coincided perfectly with the rule for addition); rather, it is a fact about what I *should* do in light of my intentions. No sophistications of dispositionalist accounts which might handle problems concerning my finitude or dispositions to err can provide the normative element of the fact demanded by the sceptic. In short, dispositionalism equates competence with performance, and thus leaves the rule-following paradox untouched. Recall Wittgenstein:

> no course of action could be determined by a rule, because every course of action can be made to accord with the rule… [I]f *any* action 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.[11]

The 'rule' Wittgenstein is speaking of here, is, of course, the content of one's intention—the rule one *means* to follow. If what one intends to do is simply what one is *disposed* to do, it would not make sense to say of those who are disposed to forget to carry when computing large sums that they are *mistaken*. Nor would attributions of correctness make sense, either. Dispositions to act are just as normatively impotent as the actions to which they give rise; on a dispositionalist account of meaning and intention, it would be impossible to speak of any particular application of a linguistic expression or the rule for addition as 'correct' or 'incorrect'.

[10] Ibid., 57.
[11] Wittgenstein, *PI*, §201.

As with nearly all facets of his argument, much ink has been spilled in arguing that the shrift Kripke gives to semantic dispositionalism is far too short.[12] Unlike Kripke, I'm under no illusion that the arguments levelled above against dispositionalism are decisive (as though any philosophical arguments ever *are* to anyone other than the individuals that advance them); however, I believe they do show that a great deal of legwork is required of semantic dispositionalists if they are to explain how dispositional facts can possess normative force such that my computation of the sum of 68 and 57 *should* come out 125 and *not* 5. With that, I propose to leave dispositionalism aside for now and move on to present a brief sketch of Kripke's distinction between 'straight' and 'sceptical' solutions to the paradox.

§ III: Two Forms of Solution

After rejecting dispositionalist responses to the paradox, Kripke goes on to distinguish between two forms a response to the sceptical conclusion might take: a *straight* solution would comprise a rebuttal of the sceptic's conclusion, showing by way of a subtle or otherwise elusive argument that, upon further reflection, the sceptical conclusion is unwarranted and that there is indeed a fact of the matter as to what someone means or intends when they purport to follow a rule. Such a fact would have to provide the 'correctness conditions' by which it could be judged whether a person's actions were in accord with his or her intentions. A *sceptical* solution, on the other hand, would begin by *accepting* both the observations of the

[12] See, for instance, Graeme Forbes, "Skepticism and Semantic Knowledge," in *Rule-following and Meaning*, ed. Alexander Miller and Crispin Wright (Montreal: McGill-Queen's UP, 2002), 16-27.

sceptical argument and the sceptical conclusion, but it would not simply stop there. Upon acknowledging that the sceptic is unanswerable, a sceptical solution would go on to explain why we are not obliged to answer the sceptic's demands for an ultimate (or foundational) meta-linguistic justification—why the particular type of justification demanded by the sceptic's queries is unnecessary. While countenancing the sceptic's argument and conclusion, a sceptical solution would locate the value of our common practice of ascribing meaning and intention to each other's linguistic performances (and intentional* behaviour generally) somewhere other than where the sceptic's argument presupposes it to be. Proponents of a sceptical solution would need to show, among other things, how ascriptions of meaning and intentions can be meaningful and worthwhile without having to correspond to 'facts in the world'.

§ IV: Further Requirements for any Would-be Solution

There is a sense in which understanding a paradox is the same as solving it. Solving the sceptical paradox about meaning requires more than simply providing pertinent examples in which meaning is clearly exhibited; it requires understanding why the sceptical argument seems right even if we think (or know) its conclusion clearly to be false. Thus, to solve the Kripke-Wittgenstein paradox, we must understand why it *is* a paradox. We can begin this task by noting that the argument is paradoxical because its conclusion controverts the pervasive philosophical (and, I believe, common-sensical) doctrine that our linguistic utterances and inscriptions have meaning because *we mean* something by them.

If there are no facts as to what we mean or intend, then *a fortiori* there can be no facts concerning the meaning of our linguistic and non-linguistic productions. From a theoretical point of view, the paradox seems to deny a fundamental presupposition behind most theories of meaning and language (except, of course, for postmodernism). But what is perhaps even more striking is that the air of paradox is not merely felt at the level of philosophical reflection. The sceptical conclusion denies the meaningfulness of *all* linguistic productions. So it is not merely philosophical sensibilities that meaning scepticism is at odds with: that people generally feel there to be a fact of the matter as to what rule they are intending to follow or what they mean to say is evidenced by the currency that statements such as "what I (or he or she) *meant* [to say] was…" have in everyday speech. And it seems reasonable to think that a presupposition behind *giving a reason* for one's actions is that there are *facts* about one's intentions. Even now as I type, I feel quite certain that choosing my words carefully is one way to ensure that what I write will correspond to my linguistic intentions at this moment. These considerations point towards a criterion that a solution to the paradox—straight or sceptical—must satisfy, namely, that it somehow account for the beliefs that are so strongly contrary to the sceptical conclusion.

Consider, by way of analogy, Zeno's paradox about motion: clearly, physically demonstrating that motion is possible by taking a step does not properly dissolve Zeno's paradox. A proper dissolution must show where Zeno's reasoning went wrong so that it led to a conclusion that is obviously false. Likewise, a straight solution to the rule-following paradox cannot consist in

simply demonstrating the fact which the sceptic denies, nor would a sceptical solution be any solution at all if it merely agreed with the sceptical conclusion and left it at that.

A straight solution must not only show what fact it is about me that constitutes my meaning 'plus' and not 'quus'; it must also relate the proffered fact to the way in which I already knew (or at least thought) that I was following the rule for addition. Kripke writes,

> …there is a condition that any putative candidate for such a fact must satisfy. It must, in some sense, show how I am justified in giving the answer '125' to '68+57'. The 'directions'…that determine what I should do in each instance, must somehow be 'contained' in any candidate for the fact as to what I meant. Otherwise, the sceptic has not been answered when he holds that my present response is arbitrary.[13]

The 'meaning fact' must somehow explain the consciousness of content and intentions that instigated the sceptical considerations in the first place. For a proffered straight solution to fail in meeting this requirement would be for it to be partial at best. It might determine the fact about me which uniquely determines the content of my intentions, but such a solution would still leave me following the rule blindly. It would make all non-accidental rule-following of a sort with 'do not fully exhale while swimming'. Such a solution would seem to deny the very possibility of agency, for it would leave us able to conform to rules, but entirely *unaware* of the rules we follow and even that we follow them. (Which in turn would render the very possibility of formulating ideas of rules and what it means to follow them problematic.) The possibility of doing something *deliberately* would remain on Kripke's chopping block. This signals an epistemic

[13] Kripke, *WRPL*, 11.

criterion that a straight solution must satisfy; namely, it must show how facts about meaning are available to consciousness. It must show how we can be *aware* of what we mean and how we can be aware *that* we have thoughts and intentions. In the next chapter, I will examine Ruth Millikan's attempt at a straight solution and argue that it fails to meet this criterion.

Perhaps ironically, the need to accommodate the intuitions that are so contrary to the sceptical conclusion does not place such stringent demands on a sceptical solution. The proponent of a sceptical solution can simply say something to the effect of "Look, the sceptical argument clearly shows that we were simply wrong to think that *this* (i.e., facts about what we mean) is why we do *that* (i.e., make claims about what we mean or intend). But since *that* is something we regularly *do*, and *doing* it has *value* and *utility* in our lives, the moral of the story is that philosophical theories about truth and meaning are just so much hot air. Facts and theories don't justify, people do."

§ V: Wittgenstein's Sceptical Solution

Kripke claims that Wittgenstein himself gave a sceptical solution according to which the sceptic's argument is not to be rebutted on its own terms. But, the sceptical solution maintains, this does not subvert the purpose or value of our everyday ascriptions of meaning and intention to one another. It merely shows that our ordinary idea of meaning or intending—of rule-following generally—cannot be given a philosophical ground by way of correspondence to intentional* 'facts'. The value of meaning ascriptions (and linguistic performances generally)

is simply to be found in the roles they play—in the work they do—in the language game of our community:

> Now if we suppose that facts or truth conditions, are of the essence of meaningful assertion, it will follow from the sceptical conclusion that assertions that anyone ever means anything are meaningless. On the other hand, if we apply to these assertions the tests [generally, 'don't think, look!'] suggested in the *Philosophical Investigations*, no such conclusion follows. All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives. No supposition that 'facts correspond' to those assertions is needed.[14]

So even though the sceptic is correct in asserting that I possess no ultimate meta-linguistic justification for answering '125' rather than '5' to 68+57, his assumption that such a justification is necessary to the very possibility of meaningful discourse (and intentional behaviour generally) is unwarranted. Even though I may—in the sceptic's terms—be following the rule for addition blindly, my simple inclination to answer '125' is all the justification I need in order to do so. Furthermore, what determines the propriety of my answer—what provides the *normative* standard by which my actual performance is judged—is simply a matter of whether my answer is in agreement with my community's practices, that is, whether my inclinations are in agreement with those of my fellow community members. So even though there is no intention-determining fact about me which I can point to in order to answer the sceptic's doubts, according to the sceptical solution, the notion of rule-following on which the sceptical argument is based is simply wrong-headed. With specific regard to language, the sceptic falsely

[14] Ibid., 77-8.

assumes that meaning-determining linguistic rules must be correspondence truth rules.

At this point, we might begin to wonder why the sceptical solution ought to fare any better than dispositionalist responses to the paradox, for it seems to have simply transferred the norm-providing dispositions from the individual onto the community. What is correct or incorrect is merely what my fellow community members are generally disposed to agree or disagree with. Indeed, can we not re-formulate the sceptic's queries so as to be directed against the entire community? Surely there is a number so large that no member of the community has added it to another. For the time being, I will simply echo Kripke's response to this objection: although the sceptical solution is in some sense meant to resolve the sceptical paradox, it is not intended as a replacement for the correspondence truth conditions presupposed by the sceptical argument. The sceptical solution rejects correspondence truth altogether by replacing truth conditions with assertability conditions:

> One must bear firmly in mind that Wittgenstein has no theory of truth conditions—necessary and sufficient conditions—for the correctness of one response rather than another to a new addition problem. Rather he simply points out that each of us *automatically* calculates new addition problems (without feeling the need to check with the community whether our procedure is proper); that the community feels entitled to correct a deviant calculation; that in practice such deviation is rare, and so on. Wittgenstein thinks that these observations about sufficient conditions for justified assertion are enough to illuminate the role and utility in our lives of assertion about meaning and determination of new answers. What follows from these assertability conditions is *not* that the answer everyone gives to an addition problem is, by definition, the correct one, but rather the platitude that, if everyone agrees upon a certain answer, then no one will feel justified in calling the answer wrong.[15]

[15] Ibid., 111-2.

One thing we might question here is whether the observations rehearsed by Kripke are in fact correct. Think, for example, of a person standing on a busy street corner hollering absurdities.[16] Even when what is being said is comprehensible (imagine the person is a Nostradamus-type character prophesying the end of the world), we typically *do not* engage such people in debate. We more likely avert our gaze and silently curse the walk signal for not changing more quickly. And there are numerous, more mundane examples in which it seems we regularly refrain from correcting misjudgements or misuses of terms. Though rare by comparison to the amount of correct usage, it is nevertheless not unusual in the course of an ordinary conversation for someone to misuse a term in some way. And it seems that we rarely correct such mistakes. It seems that typically, if we think we know what someone *meant* to say when he or she has misspoken, we *refrain* from offering correction and carry on the conversation without interruption.

Although these considerations could lead into a lengthy digression, I raise them simply to point out the non-trivial issue of Wittgenstein's conception of communities and the language-games they play. Language-games are not to be thought of as identical or even analogous to natural languages. It may be true that speaking the same language as someone else entails that I am able to play at least one language-game with them, but that is all. Even though the Wittgensteinian notions of language-games, community, etc. are notoriously nebulous and vague, it would seem clear that they are not to be thought of as static or as possessing rigid borders. Moreover, language-games are not immutable, both in terms of

[16] Thanks to Bruce Hunter for drawing my attention to examples of this sort.

membership and in terms of the rules and 'pieces' of the games themselves. Case-studies from the history of science may help to illuminate this point: new terms are introduced to and often subsequently dismissed (think of phlogiston, for example) from the language-games scientists play; speculation and further research often lead to a term gaining new meaning (compare what 'atom' meant for the ancients to what it means for us now). All this suggests that Nostradamus examples do not necessarily contravene the observations highlighted in the preceding quotation. In Wittgensteinian terms, refusal or failure to engage with the person on the street-corner or to correct a friend who consistently misuses or misapplies a term is, simply, to not play a particular language-game—to not be a member of that particular community.

(However, *if* Kripke's Wittgenstein took such a line in response to the objection raised, it would might start looking as though he was not remaining true to his promise of simply 'telling it like it is'. As we find ourselves pressed to add flesh to the skeletal description of the Wittgensteinian community, such an entity increasingly appears to be an idealized abstraction; what we were promised was a brute empirical description, but the notion of community required to fit the sceptical solution appears as thought it might be significantly abstracted from reality.)

On the one hand, the sceptical solution seems to provide a wholly unpalatable account of linguistic and conceptual norms: Wittgenstein's answer to the question of *why* we speak as though there is a 'fact of the matter' concerning our intentions is, basically, that it is simply 'how we do things 'round here'—

there is just no more to it than that. On the other hand, the sceptical solution possesses some *prima facie*, intuitive force: upon realizing that there is nothing before my mind which ultimately justifies my answer to a problem of arithmetic, nor my use of an expression in everyday speech, it becomes apparent that at bottom, the only thing I could possibly be conscious of that impels me to give *this* answer or use *that* expression is that it 'feels right' to me at the time. But, as Kripke's Wittgenstein has shown, such feelings cannot ultimately *ground* the normativity of meaning. In practice, the only normative 'facts' with which I might determine whether I have followed a rule rightly or wrongly are facts about the conventions or practices of my community—whether my compatriots would say or do the same as I.

§ VI: Conclusion

While I believe the sceptical solution is importantly relevant to a proper dissolution of the Kripke-Wittgenstein paradox, the commitments outlined at the beginning of this chapter should make it clear that I do not think it is the conclusion of the matter. Just how the sceptical conclusion figures into what I take to be a complete answer to Wittgenstein's hypothetical sceptic will not be made clear until the final chapter. In the meantime, it will prove useful to consider the straight solution proffered by Ruth Millikan. Exposition and examination of her answer to Kripke's Wittgenstein will be the primary focus of chapter II.

Chapter II: A Biological Solution

§ I: Introduction

At the outset of the previous chapter I said that the goal of this project was to proffer an answer to the questions of how we think and how we think about thought. Before discussing Ruth Millikan's response to the Kripke-Wittgenstein paradox, it will be worthwhile to take stock of the implications the arguments of the previous chapter have for the metaphysics and epistemology of thought. While it would be true to say that, in denying the factuality of meaning and intentions, the sceptical paradox and solution simply deny the meaningfulness of metaphysical and epistemological questions concerning thought, leaving it at that would scarcely be illuminating.

If the analogy I drew between thinking and walking is correct—that is, if thinking is an ability that one can have and know that she or he has it *without* knowing the theory that correctly specifies in what such an ability *consists*—then the possibility that we are able to know *that* and *what* we think is consistent with both the considerations raised in the sceptical argument, and even (possibly) with the observations comprising the sceptical solution.[17] That is, such a conception of thought and thought-about-thought is not ruled out *a priori* by Kripke's arguments to the effect that facts about what one means or intends cannot be found merely in what is open to consciousness, nor is it necessarily incompatible with the idea that our notions of semantic normativity stem from the public

[17] These possibilities are not, of course, consistent with the *conclusion* of the sceptical argument. How we can countenance the premises of the sceptical argument and even the observations of the sceptical solution while denying meaning scepticism will become clear in chapter III.

practices of our linguistic community. But without further argument, all of this would simply remain promissory note-ish conjecture.

The central concern of the present chapter is to exposit and examine Millikan's solution to the paradox. I want to suggest that it has the form that an answer to the metaphysical problem of thought must take. I think her account is plausible, but I do not think it fully addresses the central problem of the sceptical paradox. Thus, my concern will not be to critically assess it in a head-on fashion. My critical discussion will be focussed on the omissions that I believe render her biological solution inadequate in the sense that even if it is a correct response to the metaphysical problem posed by the paradox, it leaves the epistemological problem untouched. Detailing the merits and pitfalls of her account as a proffered (complete) straight solution to the Kripke-Wittgenstein paradox will provide greater clarity and cogency to one of the central theses of this essay, namely, that the metaphysics and epistemology of thought must be distinct.

§ II: Locating Semantic Facts in Evolved Purposes

Millikan's approach to solving the Kripke-Wittgenstein puzzle involves locating basic intentions or root purposes in behavioural forms which have arisen and proliferated through processes of adaptation and natural selection. There are myriad examples from biology depicting organisms (and even parts of organisms) as following certain specifiable rules, and the explanation for the occurrence of such forms of rule-governed phenomena is that they have proven advantageous to the survival and procreation of the species in which they emerge and persist. For

Millikan, a heritable trait or pattern of behaviour comes to have a purpose or proper function in virtue of being selected for and passed down through subsequent generations. Accordingly, she argues that not only is there a fact of the matter as to what we mean by our linguistic performances, but that our capacity to follow determinate (or realist or correspondence) semantic rules is as much a product of natural selection as is the heart's capacity to pump blood or the eyelid's to blink and moisten the cornea. That it is the eyelid's biological purpose to keep the cornea moist and not, say, to periodically point toward one's knees, is because conformity to the rule 'regularly spread tears over the surface of the eye' is what historically (i.e., evolutionarily) accounts for the continued proliferation of the eyelid and eye-blink reflex.[18]

In keeping with her notion of 'biological purpose', Millikan begins her straight solution to the rule-following paradox by stipulating a new sense of the verb 'to purpose': in contrast with 'intend' which strongly suggests an explicitly represented purpose, she proposes to use 'purpose' as a broader term which is meant to include both expressed and unexpressed purposing. (Recall the portion of the argument for the paradox that gave rise to the meta-linguistic regress: my following this rule rather than that cannot be explained or justified by the fact that I have given myself instructions for following the rule, for then I would need further instructions for following those instructions which are *themselves* rules. All of these rules are to be taken as explicitly represented; intending to follow them would

[18] I realise that this may read as a Procrustean distortion of the notion of a rule, but it is in keeping with Millikan's terminology in her response to the paradox.

itself be open to consciousness.) She proposes that unexpressed, root purposing is the basic form of purposing which uniquely determines the content of explicitly represented intentions to follow a rule. "To understand what it is to have an explicit purpose that one represents to oneself we must first understand what it is to have a purpose the content of which is *not* represented. Basic or root purposes must be *unexpressed* purposes."[19] Once root purposing has been conceived as the basic form of purposing which is not explicitly represented, Millikan's task becomes that of articulating just what such unexpressed purposing might consist in.

With her new sense of 'purpose' in hand, Millikan characterizes three possible ways in which an individual could conform to a rule, of which the latter two correspond to those distinguished at the outset of the exposition of the sceptical paradox in chapter I. First, an individual's behaviour may accidentally coincide with a rule, as my past addition computations have accidentally coincided with the rule for quaddition; second, an individual may follow a rule by explicitly intending to do so, as when I compute sums with the expressed purpose of following the rule for addition (recall that this type of rule-following is the sort that gave rise to the paradox in the first place); third, an individual may purposefully (in the stipulated sense of 'purpose') conform to an unexpressed rule. Recall the examples of swimming and walking from the previous chapter: in Millikan's terms, I purpose to conform to the rules 'do not fully exhale while swimming' and 'push with the toe to maintain speed while walking'—even

[19] Ruth Garrett Millikan, "Truth Rules, Hoverflies and the Kripke-Wittgenstein Paradox" [hereafter TRHKWP], *The Philosophical Review* 99, no. 3 (1990): 329.

though they themselves are not consciously expressed intentions in my mind—when it is my intention to swim or walk. It is this third way of following a rule, Millikan claims, that must ultimately "account for the *normative* element that is involved when one means to follow a rule, [that must] account for there being a *standard* from which the facts, or one's dispositions, can diverge".[20] Her strategy is to ground the normativity involved in rule-following behaviours that involve explicit intentions—which are, of course, the topic of concern in the sceptical argument—in the various unexpressed, purposive, rule-following behavioural forms that have arisen as a result of natural selection.

Millikan proceeds to describe several examples which are instances of purposive rule-following behaviour in the third way of conforming to a rule. The first involves a pattern found in the behaviour of male hoverflies: when an object of a certain angular size and traveling at a certain angular velocity stimulates a male hoverfly's retina, it reacts by turning and accelerating so as to intercept the object. It does not matter what the particular object happens to be—if its angular size and velocity fit the criteria of the 'proximal hoverfly rule', the male hoverfly will turn and accelerate to intercept it. Her reason for labelling this rule 'proximal' is to emphasize that it does not concern how male hoverflies should respond to distal objects in their environment; rather, the rule specifies only how males ought to respond to a proximal stimulus—i.e., moving spots across their retinas. Of course the reason evolution has 'programmed' males of the species to engage in such patterns of behaviour is that *female* hoverflies are fairly uniform in size and speed, and males accelerate at a fairly uniform rate. Thus, the fact that

[20] Ibid., 329.

male hoverflies *purpose* to follow the proximal hoverfly rule is explained by the 'distal hoverfly rule': if you see a female, catch it.

That the distal rule explains why males are wired to follow the proximal rule is what makes it the case that male hoverflies *purpose* to follow the latter. They purpose to follow the proximal rule because, often enough, doing so has resulted in successful conformity to the distal rule. Their root biological purpose in following the proximal rule is to follow the distal rule. Moreover, their success in following the proximal rule—whether they have in fact darted-off after a tossed pea, a shoe, a bird, or a female hoverfly—is non-accidental for the same reason. It is non-accidental because competence in conforming to the proximal rule has led to conformity to the distal rule, and this explains why the behavioural form (that results in conformity to the proximal rule) has proliferated.

The behaviour of the male hoverfly, Millikan claims, is a paradigm case of unexpressed purposing to follow a rule, for while it is unlikely that male hoverflies perform the complex computations involved in conforming to the proximal hoverfly rule, they follow it nonetheless. And they do not simply follow it by accident, as I follow a quaddition-like rule accidentally whenever I compute a sum. Since the proximal hoverfly rule does not discriminate between female hoverflies and other objects that could just as easily fit the angular size and velocity criteria, there are all sorts of distal *quoverfly* rules that males happen to conform to just as my computation of sums accidentally conforms to the rule for quaddition. But rules such as 'if you see a bird, catch it' are not part of the proper evolutionary explanation for the proliferation of the males' behavioural pattern.

Male hoverflies purpose to follow the proximal rule because conformity to it has historically resulted, often enough, in conformity to the distal rule 'if you see a female, catch it'; and because it resulted in conformity to *that* rule is why this particular behavioural form has been selected for. The hoverfly's root biological purpose in following the proximal rule is to follow the distal rule. Since proliferation of the behavioural form which brings hoverflies into conformity with the proximal rule is ultimately due to the fact that it has historically brought them into conformity to the distal rule, it is *this* function (i.e., conformity to the distal rule) that represents the *purpose* or *meaning* of the hoverflies' behaviour.

Now we might be inclined to object that there is an essential difference between the rule-following behaviour of hoverflies, on the one hand, and the explicitly represented rule-following involved in linguistic performances and other instances in which we humans purport to follow rules. The objection I have in mind begins by noting that it is obvious that the male hoverfly is *hardwired* to follow the proximal hoverfly rule. But, the objection continues, the sort of expressed rule-following that is at issue in the Kripke-Wittgenstein paradox is about rules that we *learn* to follow. Recall Kripke's explication of what seems to be a fairly ordinary notion of rule-following:

> By means of my external symbolic representation and my internal mental representation, I 'grasp' the rule for addition… 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. This is the whole point of the notion that *in learning to add I grasp a rule*: my past intentions regarding addition determine a unique answer for indefinitely many new cases in the future.[21]

[21] Kripke, *WRPL*, 7-8. Italics added.

Although Kripke does not make the distinction explicit in his exposition of Wittgenstein, the rule-following paradox seems to concern solely the kind of rule-following behaviour involved in which the rules are *learned*. Thus, we might conclude, hard-wired or 'automatic' rule-following of the kind exhibited by the hoverfly is of an entirely different species than that involved in such acts as arithmetical computations or linguistic performances, and hence has no bearing on the sceptical argument.

But Millikan sees this objection coming, and the second example she highlights describes how even learned rule-following behaviours can be rooted in biological purposing. It so happens that if a rat has gotten sick immediately after having eaten something, it will never again eat anything with the same taste. Thus, a rat which falls ill shortly after eating some soap will never again eat anything that tastes like soap, even if it was in fact some nesting material and not the soap which was the actual cause of the rat's illness. Thus the learned proximal rat rule 'do not eat soap' is explained by the distal rat rule 'do not eat poisonous substances'. Millikan claims that according to her account of biological purposing, even the behaviour of circus poodles that have learned how to ride bicycles in order to procure their dinners exhibits non-accidental rule-following that is ultimately explained by the selection and proliferation of particular biological functions. To these we can add our own examples of the unexpressed rules governing our actions when swimming and walking, for the abilities to follow the unexpressed rules concerning air in the lungs and pushing with the toe are learned as well.

Bringing her discussion back to human linguistic activity, Millikan argues there is good reason to think of linguistic rules on the biological model she has presented. She suggests that verificationist assertibility conditions are to realist correspondence truth rules as the proximal hoverfly rule is to its distal counterpart:

> The first thing to note is that if truth rules were distal rules they would surely have to be *backed* by proximal rules, rules about how to respond to our thoughts (inference) and to the immediate fruits of our perceptual explorations (perceptual judgment). They would have to be *backed* by rules that determined assertability conditions, the innermost of these conditions being within the mind or brain or at the interface of mind or brain and world.[22]

On some of the more specific ways in which the relationship of assertibility conditions to correspondence rules reflects her conclusions on proximal and distal rules, she has the following to say:

> Conforming to the proximal hoverfly rule and the proximal rat rule often fails to bring hoverflies and rats into conformity to the distal hoverfly and rat rules. Similarly, conforming to proximal assertability rules might often fail to bring humans into conformity to truth rules. One can unknowingly say what is false even though one has good evidence for what one says… Also, whether conformity to the proximal hoverfly and rat rules helps to produce conformity to the distal hoverfly and rat rules on this or that occasion depends upon factors in the hoverfly's or rat's external environment over which it has no control. Similarly, whether conformity to proximal assertability rules would bring us into conformity to truth rules in this case or that might depend upon factors over which we had no control. For example, circumstances responsible for most perceptual illusions are circumstances outside the observer which, normally, he neither controls nor needs to control.[23]

We purpose to conform to certain proximal verificationist (or justification) rules because doing so has often enough resulted in conformity to distal correspondence rules, and success in accurate mapping or picturing (i.e.,

[22] Millikan, "TRHKWP," 346.
[23] Ibid., 346-7.

correspondence) is what explains why we purpose to conform to assertibility rules. And since we only purpose to conform to the proximal assertibility rules because our ultimate purpose is to conform to distal correspondence rules, it follows (on Millikan's view of meaning as function or purpose) that correspondence rules are what define the semantics of language and thought.

In spite of its apparent complexity, we can see now that the basic form of Millikan's solution to the Kripke-Wittgenstein paradox is very simple and straightforward. At heart, it is simply an inference to the best explanation in light of 1) the assumption that there *are* facts of the matter regarding our intentions[24] and 2) the considerations raised by the sceptical argument. She avoids the regress of interpretation by locating meaning-facts in non-accidental behaviour of organisms in which the rules to be followed are not explicitly represented.

§ III: Are Hoverflies and Rats Adequate to the Task?

My reason for presenting of Millikan's biological account is that I believe it will help us to get clearer on what is demanded of a solution to the Kripke-Wittgenstein paradox, and further, that it will move us towards a full understanding—and hence dissolution—of the paradox. Her account presents some important ideas with respect to rule-following behaviour that are surely relevant to the conceptual and linguistic abilities of human beings; but, I believe,

[24] When viewed in this light, Millikan's response seems to simply beg the question against the meaning sceptic, but this is not an objection that I want to press (at least, I am not interested in pursuing it quite yet). The reason for this will become clearer throughout the remainder of this chapter and the next, but in the meantime I will simply say that I think to merely discard the biological account as question-begging against the Kripke-Wittgenstein paradox would be un-instructive; we would be prevented from appreciating its positive contribution as one component in a full dissolution of the paradox.

it fails to fully address Kripke's sceptical paradox. In what follows, I will raise several distinct but related criticisms against the biological account. While some are rather probative and speculative in nature, there is one in particular that I believe both highlights the underlying source of the paradox and reveals the need for more than what Millikan has given us.

In his reconstruction of Wittgenstein's argument, Kripke emphasises repeatedly that any putative fact proffered by a straight solution—that is, a fact about me that would show that I intended to follow the rule for addition rather than quaddition—would have to be one that is available to consciousness:

> Even now as I write, I feel confident that there is something in my mind—the meaning I attach to the 'plus' sign—that *instructs* me what I ought to do in all future cases.[25]
>
> The idea that we lack 'direct' access to the facts whether we mean plus or quus is bizarre in any case. Do I not know, directly, and with a fair degree of certainty, that I mean plus?[26]

and finally,

> …there is a condition that any putative candidate for such a fact must satisfy. It must, in some sense, show how I am justified in giving the answer '125' to '68+57'. The 'directions'…that determine what I should do in each instance, must somehow be 'contained' in any candidate for the fact as to what I meant. Otherwise, the sceptic has not been answered when he holds that my present response is arbitrary.[27]

In § IV of the previous chapter, I argued that Kripke's epistemic criterion is an essential feature of the sceptical argument. There would be no paradox if we were not aware that we intend to follow rules and were not in some sense also aware of the content of our intentions. And while Millikan might agree with this

[25] Kripke, *WRPL*, 21-2.
[26] Ibid., 40.
[27] Ibid., 11.

last claim, she thinks Kripke's epistemic requirement is just plain wrong. In the expository section of her paper, she argues that his emphasis on finding something 'before my mind' that would serve as a fact which justifies my answer of '125' to 68+57 is misguided precisely because Wittgenstein *is right* in arguing against the Brentanian thesis that intentionality is "a *sui generis* feature given to consciousness":

> What lies before consciousness does not determine its own significance; knowing what one means is not a matter, merely, of apprehending the contents of one's mind. In short, meaning is neither a state of awareness nor an epistemological given. It does not occur encapsulated within consciousness; it is not a state that simply *shows* its content or its significance.[28]

This passage conveys what Millikan thinks the lesson of Wittgenstein's rule-following considerations *ought* to be. In the section it comes out of, she is keen to stress that she *agrees* with the observations made by Kripke's sceptic; her disagreement concerns what is taken to follow from the argument's premises. From the indeterminacy found in past behaviour and introspectible mental states, the sceptic concluded that meaning-facts do not exist. Millikan, on the other hand, argues that the very same considerations are properly understood as showing that meaning-facts are not *given* or *available to consciousness*. What is more, it is her very acceptance of Wittgenstein's arguments against the irreducibility of intentionality which opens the way for her solution; that meaning and intentionality are *not* epistemically given allows for the fact constituting my meaning 'plus' rather than 'quus' to be *closed* to consciousness. The argumentative import of this preliminary move is not to be overlooked, for it, and

[28] Millikan, "TRHKWP," 326.

not the details of the biological account itself, is the aspect of Millikan's argument that deflates the air from the Kripke-Wittgenstein paradox.

In his review of Kripke's book, Brian Loar argues that anyone who follows Wittgenstein in rejecting Brentano's irreducibility thesis and yet maintains that intentionality is an objective feature of thought owes a solution to the Kripke-Wittgenstein paradox.[29] At first glance, this would seem to be a demand for a straight solution. But consideration of the biological account reveals that matters are not so straightforward: recall that a straight solution must not only provide the fact about me that constitutes my meaning one thing rather than another (e.g., 'plus' rather than 'quus'), but that it must also *dissolve* the paradox—it must explain the self-awareness of everyday agents. Or (what might amount to the same thing), it must explain how we were conscious of rules and that we were following them in the first place.

Now there is obviously a sense in which the biological account *does* justify our ordinary conception of rules and ourselves as following them, for it provides a kind of basis for claims that we mean, think, and intend. It reveals how there can be determinate content to my intention such that I am justified in answering '125' to 57+68, and it also establishes the basis of an account of what that content is. But there is another sense in which the biological account fails to justify, for it makes the possession of explicit intentions utterly mysterious. The biological model re-affirms the propriety of claims that I am purposing to follow a determinate rule, but it simultaneously renders problematic how I, you, or anyone

[29] Brian Loar, "Critical Review of Saul Kripke's *Wittgenstein on Rules and Private Language*," *Nous* 19, no. 2 (1985): 280.

could ever be *aware* of our thoughts and intentions. What makes it such that we are *not* like hoverflies who do not explicitly represent the proximal hoverfly rule?

In the same sense that the molecular identification of water as H_2O (in addition to facts about our bio-chemical make-up) explains why we think of water as we do—why it tastes, feels, looks, etc. a certain way—Millikan's story purports to explain why we have the concepts pertaining to rule-following that we do. But at the very same time, her account renders problematic how we came to *possess* them. If meaning is not simply given to consciousness, how did we ever come to have an idea of it *at all*? Consider the following excerpt from Millikan's paper:

> To understand what it is to have an explicit purpose that one represents to oneself we must first understand what it is to have a purpose the content of which is *not* [explicitly] represented. Basic or root purposes must be *unexpressed* purposes…"Intend" strongly suggests an explicitly represented purpose, that is, a purpose that is thought about. So let me use the verb "to purpose"…to include this more basic way of having a purpose.[30]

It seems that Millikan may very well be *modelling* unexpressed purposes on the categories of explicitly represented acts of intending and the normativity involved therein. For one claim she seems to be making is that unexpressed purposes are in every way *similar* to explicitly represented intentions except that they are *not* explicitly represented. So there is an ambiguity in the first sentence of the above quotation, since there is a sense in which *we already have to have an idea of what it is to have an explicitly-represented purpose* in order to *understand* her appeal to unexpressed purposes.

[30] Millikan, "TRHKWP," 329.

Now there is certainly no problem with modelling atoms on billiard balls and subsequently claiming that the latter are composed of the former; and to the extent that the biological account performs a similar job with respect to normativity and intentionality, Millikan has achieved her purpose. However, even though atomic theory aids in the explanation of why we conceive of billiard balls as we do, it does not *fully* explain our cognizance of billiard balls in the first place. Likewise, while Millikan's theory plausibly explains *why* we represent our explicit intentions as we do (i.e., as having determinate content or satisfaction conditions), it does not suffice as a complete explanation of how we might have come to conceive of what it is for us to have intentions and follow rules.

Allow me to press this point a little further. Not only is the possibility of normativity at the heart of the sceptical paradox, but so too is our *awareness* of it. Notwithstanding Millikan's arguments against Kripke's epistemic requirement, it is because we are *aware* (or so we thought before encountering the sceptic) that we mean '*plus*' by the plus sign that the sceptical conclusion *is* paradoxical—indeed, that we are able to formulate the sceptical argument at all. It is because we already have an idea of what it means to follow a rule that we are able to consider and understand both the sceptical argument and Millikan's response to it. That we ordinarily conceive of following a rule as involving explicit intentions is, I believe, implicit in Millikan's work.

Seen in this light, it seems that the biological account presupposes an understanding of the very categories upon which Kripke's Wittgenstein sought to cast doubt. The biological account provides a plausible ground for acts of

meaning and intending—it answers the question concerning how it is that our mental states represent; but since it presupposes an understanding of the categories for which it is an explanation—which is, in essence, to presuppose an answer to the epistemological problem of the sceptical paradox—it is not a complete philosophical account of meaning and intentions.

That Millikan's account presupposes a prior understanding of these concepts is suggested by her stipulated sense of the verb 'to purpose'; but it is also evidenced by her metaphorical use of normative language that is ordinarily employed for the description and evaluation of the actions of *agents*—terms whose literal senses she stretches to apply to acts that are not deliberately performed consequent to the formulation of explicit intentions. Her biological account may be helpful in showing us how it is possible that we are able to form intentions according to which our use of a word or application of a rule may be judged as correct or not, but it fails to adequately address the *awareness* involved in explicitly represented intentions. It fails to explain how we could have come to have concepts pertaining to normativity and intentionality in the first place. In failing to address the epistemic component of the sceptical paradox, the account is self-defeating, for it has the implication that we would already have to know the theory before we could ask the question for which it was intended to provide an answer.

The arguments levelled up to this point in my critical assessment of the biological solution are by no means air-tight or decisive, and it is not my intention to advance them as such. But, collectively, they have brought us to a place where

we can fully appreciate one of the central theses of this essay, namely, that having a thought does not entail that one knows that (or what) one is thinking. Millikan's rejection of Brentano's irreducibility thesis clearly involves a denial that intentionality is a characteristic or property that is simply given to consciousness; as I pointed out a few pages back, this is the moral she would have us draw from the considerations raised by the meaning sceptic. And the epistemological significance of such a conclusion should not go unnoticed: intentionality does not simply present itself *as* itself—*possession* of it does not guarantee conscious *awareness* of it. In terms of rule-following, it does not follow from the fact that one imbibes and follows rules that one is at all aware that one does so; nor does it follow that one is in any sense directly aware of the content of the rules one intends to follow. Clearly, then, Millikan is right to reject Kripke's demand that facts about meaning and intentions be given to consciousness. Whatever fact it is that provides the normative ingredient which determines whether one's actual behaviour accords with her or his intentions *must be*, as the rule-following considerations show, something that is *not* given to consciousness.

While the biological account successfully explains how there can be correctness conditions with respect to what we mean or intend (that is, why we think of rules and intentions to follow them as we do), it does not adequately account for the awareness involved in explicitly represented acts of intending. On Millikan's account, it remains a mystery how we could be conscious of meaning and intentions, and thus of ourselves as rule-following beings. In countenancing the first component of Loar's challenge, it seems we are precluded from meeting

the two criteria demanded of any account purporting to give a straight solution to the paradox, namely, to provide the requisite meaning fact *and* to account for the awareness we have of rules and ourselves as creatures that follow them.

We are now in a position to see that the Kripke-Wittgenstein puzzle bifurcates into two distinct questions: in virtue of what do our mental states (including acts of meaning and intending) represent, and how can we come to be able to know or be aware of their content (and content generally)—the latter being the 'epistemological problem' to which I have been adverting throughout. If we reject the Cartesian epistemology of the mind—that is, if we deny that we know our thoughts simply in virtue of having them—it follows that a metaphysics of intentionality cannot yield, without further explanation, an epistemology of intentionality. As a metaphysical account, surely something like Millikan's story must be right. Her biological account constitutes a plausible explanation of the basis of intentional* (and intentional) phenomena; but since it does not account for the way in which we are ordinarily aware that we mean 'plus' by + (or that we mean anything at all, for that matter), it cannot be the whole story. Conscious meaning and intending does not seem to be entirely accounted for by an appeal to evolved biological purposes. Thus, the biological account requires a concomitant epistemological story in order to give some idea of how we came to possess normative and intentional* concepts.

Though they may be slight misnomers, I hope it will not be too misleading to continue referring to these two criteria that a straight solution to the sceptical paradox must satisfy under the headings 'metaphysical' and 'epistemic'. The

metaphysical criterion is simply the demand for the facts which the sceptic concluded did not exist; meeting it involves explaining how our mental representations can have determinate content. As Kripke showed, any such fact must have normative force—it must somehow determine the standard according to which one's actual behaviour might be judged as according with their intentions or not. I take Millikan's theory of biological purposing to be a viable attempt at meeting the metaphysical criterion. The epistemological criterion, on the other hand, demands an explanation of how we might have acquired concepts pertaining to intentionality and rule-following or how awareness of content is possible. While the criterion *is* epistemological, it does not specifically concern how we can be justified in holding beliefs about our thoughts. Rather, it is a demand for an explanation of the acquisition of concepts that make self-knowledge possible.

To discount Millikan's theory as entirely incorrect would be a mistake. If intentionality is *not* a *sui generis* feature which is simply given to consciousness—if "meaning is not a state that simply *shows* its content or its significance"—it follows that any adequate response to the paradox must not merely show *how intentionality arises in nature* (the metaphysical component), but it must also show *how we could come to be aware of it* (the epistemic component). Under the influence of a Cartesian theory of mind, the sceptic's mistake was to think that an answer to one must be an answer to both. But if Wittgenstein is correct in denying the epistemological implications of Brentano's

irreducibility thesis, then an adequate philosophical account of rule-following behaviour must comprise more than what Millikan's account has given us.

§ IV: Thought-about-Thought, Again

What are the implications of all this for thought-about-thought? At the beginning of the present chapter, I claimed that the arguments levelled by Kripke's Wittgenstein were compatible with a certain view of thinking and thinking-about-thought. More specifically, I claimed that the considerations constituting the premises of the sceptical paradox and the observations of the sceptical solution were compatible both with the idea *that* we think and the idea that we can *know that*—and *what*—we think.

Consider again the question "how do you walk?" Imagine that an average, competent walker (who is unfamiliar with the biomechanics of bipedal motion) somehow finds herself at a biology conference, thrust into a conversation with a group of specialists on bipedal mechanics. Having gotten a general sense of the discussion taking place, she hears and understands the question according to the biomechanical interpretation. For her to answer that she simply places one foot in front of the other would be incorrect—not merely because it would be a social faux pas that makes her appear foolish, but because it is not the answer to the question being asked. In such a context, the *correct* answer would be that she simply does not know. But there are, of course, other contexts in which it would be absurd for her to say she does not know how she walks, and while the drift of this essay is pressing toward the point that her self-understanding is not merely a

coarser-grained version of the biomechanical explanation of bipedal motion, there is surely some relation between the two. Since something like our common understanding of walking was necessary before we could formulate the questions that biomechanical theories seek to answer, our understanding of the latter presupposes our prior grasp of the former.

Similarly, since we had to know (or at least think) that we think before we could formulate the questions that a theory such as Millikan's is designed to answer, here too, an account of the latter presupposes some understanding of the former. We had to know (or at least think) that we think before we could formulate the Kripke-Wittgenstein paradox and before a theory such as Millikan's could be developed. If we had not been aware that we think, there would have been no paradox in the first place.

By this point, it should be clear that the picture of thought in play here is one on which it is a form of rule-governed behaviour. But this is not to say that all mental acts are actions in the sense that they are performed with an explicit intention to do so. If they were, we would be taken on a regress much the same as the one outlined in the sceptical paradox. Forming an intention involves the making of a decision. But if all mental acts were actions, one would have to intend to decide to intend something, and this would itself require its own decision, and so on. In part, what an account such as Millikan's accomplishes is that it shows us how rational thought can be rule-governed activity without the undesirable consequence that all mental acts are deliberate actions in the full-

blooded sense.[31] By calling thought 'rule-governed', my goal is simply to emphasize the fact that it is something that can be done well or poorly, correctly or incorrectly.

§ V: Conclusion

If meaning is not simply given to consciousness, then not only do we need to understand how it is possible that this is so, but we *also* need to understand how it is possible that we *are* somehow conscious of it. Even though the biological account failed as an explanation of the genesis of concepts pertaining to rule-following, it provided a plausible explanation for how our intentions can have determinate contents and how there can be facts of the matter concerning what we mean by our words. It also provided the basis of an account for what content is. It provided an explanatory basis for the possibility of thought and linguistic behaviour being truly rule-governed activity. This is, obviously, strictly opposed to the conclusion of the sceptical argument and subsequent sceptical solution according to which talk of meaning and intending—of following a rule 'correctly'—is *merely* a feature of the practices of our linguistic community. But if facts about what we mean or intend are not given to consciousness, it remains a mystery as to how we came to have notions of them at all. *If* (as has been an overarching assumption of this project) intentionality and rule-governed behaviour are intrinsically related, and *if* we construe thinking about thought as meta-intentionality or rule-governed behaviour *directed at* or *about* rule-governed

[31] Though I am highlighting the non-active variety of thoughts here, there are obviously mental acts that *are* full-blooded actions. Choosing to attend to a puzzle or solve a mathematical proof are examples of such.

behaviour, how it is that we ever developed concepts pertaining to thought—how we could come to be able to think about thought—remains to be adequately understood.

In the subsequent chapter, I will outline an account of the origins of intentional* and normative concepts inspired by the work of Wilfrid Sellars. In keeping with the context provided by the Kripke-Wittgenstein paradox, I will initially present the argument as answering the epistemological criterion demanded by the sceptical argument. I will argue that a synthesis of Sellars' and Millikan's accounts is required for an adequate response to the Kripke-Wittgenstein paradox. While Millikan provides a metaphysical explanation for thought, Sellars ideas are needed to understand how we might have arrived at our concepts pertaining to rule-following and intentionality in the first place. This will also serve to provide some clarity and support to what may have appeared thus far as an assimilation of the categories of intentionality and normativity.

Chapter III: Awareness of Meaning, Normativity, and Thought

§ I: Introduction

The moral of the previous chapter was that the capacity to follow rules that are explicitly represented is not exhaustively accounted for by an explanation of the mere capacity for intentional* thought and behaviour. For a realist about intentionality such as Millikan, the significance of accepting the observations of the sceptical argument was that purposes or intentions that are not explicitly represented are what uniquely determine the content of the explicit or conscious variety. If any rule is open to myriad interpretations of which only one is correct, it would seem one must have a meta-linguistic formulation of the rule instructing them which of the interpretations that is. But, as we saw in chapter one's exposition of the sceptical argument, this can only lead to a regress since the meta-linguistic formulation is in turn subject to as many interpretations as the original rule, thus one would need to know the meta-meta-language containing the rules for the meta-language, and so on. While the sceptic took these considerations to show that there can be no such thing as following a rule or meaning something by a word, the conclusion drawn by Millikan was that basic intentionality or rule-following must be rooted in processes that are not open to consciousness. The regress argument shows that the factor that ultimately determines our intentions and thus the 'rules' we purport to follow cannot be something of which we are directly aware; it therefore demands a *theory* which

postulates basic intentional* facts such that they could uniquely determine the content of the explicit intentions of which we *are* directly aware.

What Millikan failed to explicitly acknowledge, however, is the second, equally important, consequence of accepting the Wittgensteinian arguments against a Cartesian epistemology of mind and meaning. For the same reason that basic intentional* facts must be *closed* to consciousness, it follows that a metaphysical theory of the basis of intentionality does not also suffice to account for the consciousness of intentional* and normative categories that gives rise to the paradox in the first place; without already being aware of what it is to follow a rule or mean something by what we say, we could not have conceived of the paradox. While it seems quite likely that the capacity for root-purposing as articulated in Millikan's biological account is a prerequisite for the ability to explicitly represent what one intends (clearly, one must be able to follow rules in order to be aware that one is so able and to be aware of what those rules are), the arguments of the previous chapters have shown that possession of the former does not guarantee possession of the latter. Indeed, we saw that this was the very claim that the success of the biological account hinged upon: for the theory's purposes, the cogency of the hoverfly example is dependent on the reasonability of the conjecture that male hoverflies do not explicitly represent the complex measurements and calculations involved in conforming to the proximal hoverfly rule.

In the previous chapter, I suggested that the biological solution constituted a viable answer to the metaphysical problem posed by the paradox. That is, it

showed how our intentions could have determinate content. But it failed to also answer the problem of what explicit awareness of our intentions consists in—it failed to provide an explanation of how we ever came by normative and semantic concepts. By acknowledging the need for a stipulated sense of the verb 'to purpose', Millikan's theory appeared to contain tacit admission of the primacy of our usual conception of rule-following as involving explicitly represented intentions. Her response to Kripke's Wittgenstein was lacking because, while the sceptical paradox is not specifically epistemological, the reason for its paradoxical appearance is that it contains an epistemological component that an account such as hers does not explicitly provide.

Those sympathetic to Millikan's project may have long been thinking that I am not criticizing her on her own terms, that a naturalist such as herself would feel entirely content to refrain from providing an analysis of our ordinary rule-following concepts. Indeed, the idea behind her account may be to *replace* or *amend* the concepts that Kripke's sceptic has shown to be incoherent. As one commentator has put it, "[b]eing judged a revolutionary might not worry Millikan. She thinks she has good reasons for deviating from the intuitive picture. Is this not what philosophical naturalism is often forced to do?"[32] It should be clear that the basic idea behind my objection is not simply that Millikan's is an inadequate *analysis* or naturalistic reduction of normative concepts. Her biological account is based on the perfectly reasonable assumption that there are indeed facts concerning our acts of meaning and intending; it then explains—in

[32] Martin Kusch, *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein* (Montreal: McGill-Queen's UP, 2006), 73.

light of the arguments advanced by Kripke's Wittgenstein—where such facts must be located. My principal objection to her account is based on a similarly reasonable, well-evidenced assumption: our very ability to discuss the rule-following paradox strongly suggests that we are, in some sense and to some extent, *conscious* of meaning and our intentions, and the biological account of rule-governed behaviour has not given a full explanation of how this is so.

§ II: Intentionality and Normativity

At long last, it is time to say a few words about the nature of the relationship between intentionality and normativity. Although it will not be until after we have examined the Sellarsian approach to possession of intentional* and normative concepts that the picture will be entirely clear, some comments at this stage are necessary to alleviate what may have been an ongoing worry that intentionality and normativity are being inappropriately run together. This will also provide a useful introduction to the portions of Sellars' thought that will be presented in the following sections.

In short, the defining trait of intentional* phenomena is simply that they *are* rule-governed. In order for part of the world to be *about* part of the world—in order for something to be a representation *of* something else—the representation must be governed by some sort of rules of projection. Although this is equally true for artefacts that exhibit derivative intentionality such as metre sticks and maps, of greater interest is its relevance to the intentionality exhibited by the mental and linguistic representations of organisms like us that represent various

features of their environments, that are able to learn new ways (i.e., rules) to represent, and that are even able to learn new ways to learn. By this stage in our current endeavour, it should be clear that in saying that, as intentional* phenomena, all mental events are rule-governed, I do not mean to say that all thought episodes are instances of deliberate, rule-following behaviour. If I were saying that, I would be committing myself to the view of mental events that Kripke's Wittgenstein has shown to be incoherent. As the arguments of the preceding chapter were intended to highlight, Kripke's sceptical paradox shows that there *must* be rule-governed acts that are *not* deliberate—that is, such acts are not the products of explicit intentions. Millikan's response to the paradox effectively showed how evolutionary biology can be used to explain the development and proliferation of such rule-governed behavioural forms. On her model, we can understand how norms of rationality can legitimately apply to *all* thought episodes, even though in their simplest form, thoughts are not actions that are deliberately undertaken. Intentionality is exhibited in rule-following phenomena as disparate as when male hoverflies follow the proximal hoverfly rule and when I compute a sum with the expressed intention of following the rule for addition. On the view of intentionality that is slowly taking shape through the arguments of this essay, to possess a concept would be, generally, to be able to represent something *as* something. And if, in order to represent something *as* something, the representation must be governed by rules of projection, then it follows that a concept—the paradigmatic intentional* object—is essentially rule-governed.

But such ways of talking about intentionality and rule-governed behaviour in the ontological order presupposes our original awareness of rules and what it is to intend to follow them. It is explicit intentions to which the Kripke-Wittgenstein paradox is immediately addressed, and while naturalists like Millikan might think we have good reason for supplanting ordinary notions with something like her idea of biological purposing, there remains the need to explain the awareness we have of rules and ourselves as rule-following beings. In what does this awareness consist?

§ III: Language and Thought in the Order of Knowing

The question with which the previous section ended—the epistemological question for which the Kripke-Wittgenstein paradox demands an answer—subdivides into two subsidiary questions, for not only do we need to understand how an awareness of meaning is possible, but the paradox also demands an explanation of how knowledge that *we* mean is possible. That is, to fully account for the epistemological dimension of the rule-following considerations, we also need to know how knowledge of minds—both our own and those of others—is possible. In the following sections, I propose to sketch a view that responds to both of these questions.

It would seem reasonable to say that one can not deliberately follow a rule (i.e., in the full-fledged, explicitly intended way) unless one possesses concepts pertaining to rule-following.[33] Possession of such concepts would involve,

[33] Even though I have not seen this claim explicitly formulated by others writing on the subject of rule-following, it seems to me that it is implicit both in Kripke's exposition of Wittgenstein and in

among other things, the capacity to represent rules *qua* rules which in turn would involve a grasp of such normative categories as obligation, permissibility, correctness, and the like. Possession of rule-following concepts would be equivalent to the ability to explicitly represent one's intentions, an account of which was found to be missing from Millikan's theory.

It would also seem reasonable to say that concepts are, by-and-large (if not wholly), developed through language acquisition.[34] Accordingly, concept possession would be achieved through the mastery of a word or expression—it would amount to a kind of ability or know-how.[35]

It was views along these lines that led Wilfrid Sellars directly into what is essentially the same paradox as Kripke ascribed to Wittgenstein: if a language is taken to be a system of expressions which are governed by certain rules, then it would appear that in order to learn to follow the rules of an object language, one would first have to know the meta-language in which those rules are explicitly formulated; but, obviously, to have learned the meta-language would then require that we first know *its* meta-language (the meta-meta-language of the object language), and the regress would be on.[36] In the domain of the mental, the

---

much of the secondary literature that it has spawned. See, for instance, Crispin Wright, "Rule-following Without Reasons: Wittgenstein's Quietism and the Constitutive Question," *Ratio* 20, no. 4 (2007): 481-502.

[34] I realize that this assumption runs thoroughly counter to the grain of certain contemporary philosophies of mind of which Jerry Fodor's is perhaps the most pre-eminent. But since the Kripke-Wittgenstein framework is incompatible with the basic tenets of Fodor-esque causal theories of mind, I am going to further assume that I am not speaking to Fodorians since they would not have picked up this title in the first place.

[35] On the general view that concept possession is an ability rather than items of propositional knowledge, Millikan and I are in full agreement. See, for instance, Ruth Millikan and Andrew Woodfield, "Knowing What I'm Thinking Of," *Proceedings of the Aristotelian Society, Supplementary Volumes* 67, (1993): 91-124.

[36] Wilfrid Sellars, "Some Reflections on Language Games" reprinted in Sellars, *Science, Perception, and Reality* (Atascadero, California: Ridgeview Publishing Company, 1963): 321.

paradoxical conclusion is that in order to acquire concepts, one must already have them. Recall Kripke's exposition of the rule-following paradox: what makes it the case that we are adding rather than quadding cannot be that we have given ourselves a meta-linguistic rule for interpreting and following the rule for addition, for the instructions themselves are rules which may be interpreted in both standard and non-standard ways, and there is no ultimately authoritative level in the meta-meta-…-meta-language hierarchy.

Sellars begins his solution to the paradox in much the same way as Millikan. He suggests that linguistic behaviour is, in the first instance, rule- or pattern-governed: though there is a sense of 'because' in which fledgling language-learners speak as they do *because* of the rules governing their language, it is not because they *know* the rules and have explicit intentions to conform to them. Instead of hoverflies, Sellars' example of choice is the dance of bees. While it is unlikely that bees *deliberately* or *consciously* follow the rules for communicating the location of nectar relative to their respective hives, they wiggle and turn as they do *because of the pattern of the dance*. In the case of human beings, it is only once language learners become competent with the meta-linguistic vocabulary of their language that they acquire the ability to follow the rules of the language in a deliberate (i.e., explicitly intended) fashion. From this brief introduction and the overview presented in the previous section, it may already be clear how Sellars might think the epistemology of rule-following ought to be filled-in: our concepts of one's *meaning something* by a word or *intending*

to follow a rule are originally built-upon the concepts that are imbibed in the process of learning to speak meta-linguistically.

The first and probably best-known presentation of Sellars' views on the origins of intentional* concepts finds expression in the final sections of "Empiricism and the Philosophy of Mind".[37] The stage of his argument where he introduces his myth of the psychological genius Jones is similar to the one we are currently at in our discussion of the Kripke-Wittgenstein paradox.[38] Up to that point, Sellars had presented a theory of mind which was meant to detail how knowledge of public objects and events is possible; but, had he left it there, his theory would carry drastic behaviouristic consequences. If ended prior to the introduction of his myth, Sellars' theory would imply that development of intentional* and psychological concepts would be impossible. In acknowledging the existence of intentional* phenomena (of which mental phenomena are a part) while yet denying the possibility of knowledge of it, his account would amount to a sort of epistemic behaviourism. And it ought to be pointed out that this is quite similar to the place that Millikan's solution to the Kripke-Wittgenstein paradox left us. Since she did not explain how intentional* and normative concepts might be developed—how we could come to recognize or know meaning or that we have intentions—her biological account failed to make contact with the phenomena of immediate concern in the rule-following paradox. Consciousness of meaning and intentions was left unaccounted for.

---

[37] Wilfrid Sellars, "Empiricism and the Philosophy of Mind" [hereafter "EPM"] reprinted in Sellars, *Science, Perception, and Reality* (Atascadero, California: Ridgeview Publishing Company, 1963): 127-196.

[38] I am inclined to think that Wittgenstein's rule-following considerations and Sellars' infamous attack on the myth of the given are, in basic measure, the same.

At this stage in his (and our) dialectic, Sellars seeks to make what he has elsewhere called both an historical and logical point with a creative bit of anthropological science-fiction.[39] He asks us to imagine a community of our pre-historic ancestors who speak a purely 'Rylean' language; that is, we are to imagine a group of people who speak an entirely behaviouristic language (and hence think in purely behaviouristic terms). In spite of not possessing intentional* or normative concepts, the linguistic repertoire of our Rylean ancestors is quite rich: they make use of both elementary logical vocabulary such as conjunction, disjunction, negation, quantification, and counterfactuals, as well as looser logical vocabulary characteristic of the 'vagueness' and 'open texture' of ordinary discourse. Through our discussion of the biological account of the previous chapter, we can see that such a scenario should be both logically and historically possible. Recall that one of the primary points of emphasis in our discussion of Millikan's theory was that intentionality does not simply present itself *as* itself—*possession* of it does not guarantee *consciousness* of it. Thus, it is perfectly possible that the Ryleans could have developed and used a language before conceiving of themselves as having developed and employed a system of linguistic expressions and rules.

Having established this basic starting point, Sellars asks what resources would have to be added to the Ryleans' behaviouristic language in order for them to be able to speak of thoughts—to attribute intentional*, 'inner' episodes to their fellow community members, as well as to report directly upon their own. I

[39] Roderick Chisholm and Wilfrid Sellars, "The Chisholm-Sellars Correspondence on Intentionality," in *Intentionality, Mind, and Language*, ed. Ausonio Marras (Chicago: University of Illinois Press, 1972), 245.

propose that the solution he presents to the problem of this hypothetical ancestral community constitutes an answer to the epistemic problem posed by the Kripke-Wittgenstein paradox.

As a first step, Sellars suggests that what is needed is a meta-language that enables the Ryleans to speak of both types and tokens of overt linguistic events. It is important that we spell out the implications of this move since the Ryleans' adoption of a meta-language will be the key ingredient in our answer to the epistemological problem posed by the Kripke-Wittgenstein paradox.

Recall that the meaning sceptic was not questioning my knowledge of arithmetic, but rather the meta-linguistic representation of my intention.

> [My computation] 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 a function which, when applied to the numbers I called '68' and '57', yields the value 125.
>
> Now suppose I encounter a bizarre sceptic. This sceptic questions my certainty about my answer, in what I just called the 'metalinguistic' sense.[40]

This suggests a different way of articulating the portion of our epistemological problem which concerns the possibility of our awareness or knowledge of content: if intentionality is not epistemically given, if—as Millikan said—meaning does not simply *show* its content or significance, then how are we able to *talk* about the meaning of our thoughts and words? How could meta-linguistic discourse (and concepts) have originated?

From the perspective of the present project, the essential feature of language is that it is intentional* (as opposed to, say, that it is a vehicle for communication and cooperative endeavours—though of course these features are

[40] Kripke, *WRPL*, 8.

all intimately related). And furthermore, the essential feature of intentionality is that it is rule-governed. Thus, the semantic meta-language would contain the resources for explicit formulation of the rules of its object language. Once they adopted it, our fictitious ancestors would be able to speak of one another's linguistic productions as true or false; as meaning this, not meaning that, or not meaning anything at all; and even of one statement being an interpretation of another. Correlatively, the Ryleans would also be able to say of linguistic types and tokens that they are enjoined, permitted, or proscribed in particular contexts or universally.

§ IV: the Meta-language – verificationist or correspondence rules?

But we might be inclined to wonder—especially given that the truth rules of the biological account are not things we have 'direct' epistemic access to—whether it is at all plausible to think that introducing a meta-language to the original Rylese is *actually* logically and historically possible, and whether it could provide the epistemic element deemed missing in the biological account. For Sellars thought that Wittgenstein's language game metaphor had a great deal of merit. Much of his writing on the topic of linguistic rules involved comparing and contrasting them to the kind of examples of rule-governed activity that more typically come to mind when one thinks of a game, such as chess. As with Wittgenstein's reflections on language, there is a strong verificationist undercurrent in Sellars' account of the rules underlying linguistic practices. He typically distinguished linguistic rules as being of three types: language-entry (observation contexts),

language-language (inference), and language-exit (action-enjoining contexts). So this might naturally lead us to wonder if there is any way for a Sellarsian account of the epistemology of rule-following to be compatible with the biological theory rehearsed in the previous chapter. As we saw, the semantics which seemed consequent to an account of intentionality and normativity such as Millikan's were a version of the correspondence theory of meaning and truth.

There are two reasons why I believe this does not pose a problem for invoking a Sellarsian account of the epistemology of intentionality and normativity, and both can be drawn out of Millikan's response to the paradox. First, it is because meaning and truth rules are not simply given to consciousness that we need to construct a theory with respect to them:

> But precisely because truth rules are at bottom unexpressed rules, introspection can give us no handle on what kind of rules they are. Rather, it is necessary to develop a *theory* about truth rules, an explanatory hypothesis about what rules we are purposing to follow when we make sincere assertions.[41]

If the biological theory is correct, then whatever our intentional* and normative concepts provide for awareness of *cannot be* correspondence truth rules. Indeed, an explanation of our awareness of intentional* and normative facts and properties along the lines of a verificationist framework almost seems to be implied by Millikan's account. Recall what she had to say about the relationship of proximal verification rules to distal correspondence rules:

> The first thing to note is that if truth rules were distal rules they would surely have to be *backed* by proximal rules, rules about how to respond to our thoughts (inference) and to the immediate fruits of our perceptual

[41] Millikan, "TRHKWP," 345.

> explorations (perceptual judgment). They would have to be *backed* by rules that determined assertability conditions…[42]

Thus, there does not seem to be any contradiction in suggesting that the intentional* and normative concepts we possess originally derive from a verificationist rendering of rules in our meta-language.

Indeed, if the introduction of a meta-language to the original Rylese *were* the introduction of a system of correspondence truth rules, we would certainly be begging the question of present concern. But with the linguistic (and conceptual) resources already at their disposal, augmenting the Ryleans' language to allow for verificationist rules does not necessarily raise the same problem, for in its most primitive form it need only allow for assent or dissent to language-entry, language-language, and language-exit transitions. If we grant that creatures can follow rules before acquiring the ability to form explicit intentions to follow them, there should be no problem with our introduction of meta-linguistic forms to the language of the people in our myth.

§ V: A Brentanian Objection

By adopting a meta-language, our fictitious ancestors would acquire the capacity to use rule-governed linguistic representations (i.e., overt speech acts) to say things *about* rule-governed linguistic representations (i.e., other overt speech acts). In doing so, they would come to speak and hence explicitly conceive of the rule-fulness of their overt linguistic performances. In learning to speak in reference to speech, they would have gained the ability to speak semantically, for

[42] Ibid., 346.

the meta-linguistic vocabulary would contain the resources for an explicit formulation of the rules governing the original Rylean object language. By augmenting the original Rylese with the semantic meta-language, the Ryleans would expand their conceptual framework to include proto-concepts of the categories of intentionality and normativity. ('Proto' because as yet, the Ryleans have not developed a conception of the mind and thus, in this early epoch of our myth, their concepts—unlike ours—are still purely behaviouristic.)

At this stage of the science-fiction, I imagine that any dyed-in-the-wool Brentanian would decry as non-sensical the possibility of possessing semantical concepts applied solely to overt speech *prior* to mental ones. Roderick Chisholm, for instance, conveys a similar sentiment in his correspondence with Sellars:

> If the people of your myth were to give just a little bit of thought to the semantical statements they make, wouldn't they then see that these semantical statements entail statements about the thoughts of the people whose language is being discussed?[43]

While the history embodied in Sellars' myth is *conceivable*, Chisholm finds it highly fanciful and dubitable. Although Sellars had seen this objection coming and sketched a theory of meaning statements in order to answer it, a digression into his views on meaning would not only open up a whole other can of worms but is, I believe, unnecessary. In order to appreciate the possibility of a linguistic community in which semantical statements in reference to overt speech can be made *before* fully-fledged mental concepts have been acquired, one need only reflect upon the lessons of the previous chapters and paragraphs. We have seen

[43] Chisholm and Sellars, "Correspondence," 248.

that an organism can follow rules—even rules that are *learned* (as opposed to 'hard-wired')—*even if* it does not explicitly represent them.

Even though only two simple examples were outlined in our discussion of Millikan's biological theory, a further example of a typically human activity will help to seal the point. It is a brute fact about the dynamics of bicycles and motorcycles that a rider must counter-steer in order to turn. In order to make a right turn, for instance, a rider must either turn the front wheel slightly to the left or push down on the right handlebar (which has the same effect) in order to induce a right-ward lean in the cycle. The point should be obvious: even if most experienced cyclists and motorcyclists consciously intend to follow the rule for counter-steering (though I imagine even this is unlikely), surely the vast majority of us learn to ride and steer a bicycle long before we have any notion of what counter-steering is—we know *how* to counter-steer before we know *that* we counter-steer. That is, we typically learn how and are able to counter-steer before we could form a conscious intention to do so. This is just one more way of reinforcing the notion that has been emphasized throughout the preceding sections and chapters: from the fact that an organism exhibits intentionality—is able to follow rules—it does not follow that it *knows* or *is aware* that it is so able. There should then be no problem in the notion of a community in which speech applied to speech develops prior to speech (and hence thought) applied to thought.[44, 45]

[44] Indeed, this may even suggest a criterion of adequacy for theories of semantics which seek to explain the semantical relationship between thought and language (as opposed to other varieties of meaning; see Gilbert Harman, "Three Levels of Meaning," *The Journal of Philosophy* 65, no. 19 (1968): 590-602). Contrary to the dominant epistemological theme in the Cartesian and Brentanian traditions, an adequate semantics of language should be delineable with no recourse to the intentionality of thought.

For our purposes, the adoption of meta-linguistic talk is the crucial augmentation to Rylese that paves the way for the development of fully-fledged intentional* and normative concepts. In this way, the Ryleans acquire proto-rule-following concepts, for they have acquired the semantic meta-language that contains the explicit rules governing their original object language.

But even with the augmentation of Rylese to include a meta-linguistic vocabulary, our task is not yet complete. The epistemological problem of the paradox also demands—in addition to a story about how we could come to have intentional* and normative concepts—an account of the awareness we (at least sometimes) possess of the content of our *mental* states. If we do not explain how the Ryleans might have also come to develop concepts pertaining to their 'inner' lives, we will not have fully addressed the epistemological problem posed by the sceptical paradox.

§ VI: Enter the Genius Jones

With the addition of the (admittedly verificationist) meta-language in place, Sellars introduces the character of the psychological genius Jones. Jones is a Rylean himself, and as such is also confined to speaking and thinking in the terms and concepts of the strictly behaviouristic language. In spite of this, it comes about that Jones makes a significant observation: he notices that he and his fellow

[45] There are findings from developmental psychology and cognitive science that provide some support for the Sellarsian view that we do not gain the ability to think about our thoughts until well after we have acquired fluent speech. See, for instance, John H. Flavell, "Cognitive Development: Children's Knowledge About the Mind," *Annual Review of Psychology* 50, (1999): 21-45; Inge Bretherton and Marjorie Beeghly, "Talking About Internal States: the Acquisition of an Explicit Theory of Mind," *Developmental Psychology* 18, no. 6 (1982): 906-921; Peter Carruthers, *Language, Thought and Consciousness: an Essay in Philosophical Psychology*, (Cambridge: Cambridge UP, 1996).

Ryleans behave intelligently not only when their actions are accompanied by corresponding overt speech (e.g., talking oneself through a particular problem, say, how to get a sack of corn, a chicken, and a fox all across a river in a rowboat when the boat is only large enough to transport one of the three at a time), but also when no overt verbal commentary accompanies the behaviour. In an attempt to explain the observation that his compatriots act as they would if they were in (what *we* would call) a thinking-out-loud frame of mind even when they are silent, Jones develops a theory in which he postulates inner, private processes (let us assume he dubs these theoretical episodes 'thoughts') of which overt speech acts are but the culmination, and uses overt speech itself as the *model* for these inner episodes. As is typically the case with scientific theorizing where models are put to use, Jones' theory is accompanied by a commentary—gappy and brief to begin with—that serves to outline which features the postulated processes share in common with the previously understood overt speech being used as a model. ('Previously understood' due to the development of the meta-language.) The postulation of inner, intentional* episodes as *analogous* to overt speech acts is an extension of the Rylean conceptual repertoire that builds upon other concepts already in play in the Ryleans' language. Thus, Jones' commentary would stipulate that, for instance, it is not the verbal or inscriptive features of language which these inner processes possess. "[T]he episodes in question are not the wagging of a hidden tongue, nor are any sounds produced by this 'inner speech'."[46] For our purposes, the key positive feature of the analogy is that Jones

[46] Sellars, "EPM," 187.

uses the categories of *intentionality*, made explicit in the meta-linguistic semantic vocabulary adopted by the Ryleans, as a model for the intentionality of thought.

Having come this far in his historical reconstruction, Sellars' stops to point out some of the implications of the philosophy of mind embodied in the story of Jones, as well as to warn against potential misunderstandings. These points range in topic from metaphysics to epistemology to language and the mental; given the focus of the current project, I'll highlight one in particular:

> Although the theory postulates that overt discourse is the culmination of a process which begins with 'inner discourse', this should not be taken to mean that overt discourse stands to 'inner discourse' *as voluntary movements stand to intentions and motives*. True, overt linguistic events *can* be produced as means to ends. But serious errors creep into the interpretation of both language and thought if one interprets the idea that overt linguistic episodes *express* thoughts, on the model of the use of an instrument.[47]

The thought expressed here resonates with points made previously in this and other chapters. On the account of thought and rule-following that is taking shape here, we can make sense of thoughts (*qua* conceptual events) as being rule-governed—as subject to appraisal—*without* making the mistake of thinking that they must all be deliberately intended *actions*.

Once he has developed—albeit in a very rough and rudimentary form—his theory of thinking on which the intentionality of thought is analogous to the semantic characteristics of overt linguistic acts, the next step Jones takes is to teach the theory to his compatriots. When Jones sees Smith mending the roof on her hut after a violent thunderstorm even though he has not overheard her say anything like "My roof is damaged and I shall repair it", he infers—on the basis

[47] Ibid., 188.

of his theory—that Smith is having the thought, "My roof is damaged and I shall repair it." And, even when Smith *has* candidly expressed her thought, Jones' theory suggests that it is the thought and not the overt statement itself which is the real cause of Smith's repairing her roof. What is more, it turns out that Smith can be trained to *directly report* her own thoughts. That is,

> Smith can be trained to give reasonably reliable self-descriptions, using the language of the theory, without having to observe [her] overt behaviour. Jones brings this about, roughly, by applauding utterances by Smith of 'I am thinking that p' when the behavioural evidence strongly supports the theoretical statement 'Smith is thinking that p'; and frowning on utterances of 'I am thinking that p', when the evidence does not support this theoretical statement. Our ancestors begin to speak of the privileged access each of us has to his [or her] own thoughts. *What began as a language with a purely theoretical use has gained a reporting role*.[48]

One can see how Sellars was a forerunner of the so-called 'theory-theory' of folk psychological discourse. But with respect to this association, it is important to note that although he sought to illuminate our understanding of mentalistic discourse by introducing mental events as *theoretical* entities, as was emphasized by the italicized sentence at the end of the quote above, we can be trained to directly report on our own mental states. Thus, on this account, the epistemic authority of avowals is saved.[49]

Although the myth of Jones potentially explains how we come by the ability to be conscious of our intentions, it remains to be explained how Jones was able to make the observation that inspired his theory in the first place. On the picture of concept acquisition at play in this project, for him to notice that his compatriots acted *intelligently*, Jones must already possess the concept of

---

[48] Ibid., 189.

[49] I am not suggesting such authority is infallible, but merely that there are at least some occasions when avowals are highly authoritative. More on this in the final chapter.

intelligent behaviour, and surely this is part of the family of psychological concepts the origins of which we were hoping Sellars' myth would explain. To classify behaviour as intelligent, Jones would already have to possess the concept 'intelligent'. And if a concept of intelligence is to be consistent with the arguments of the preceding sections and chapters, it must certainly involve an awareness of rules and what it means to follow one.

We can answer this objection by building on Sellars' own response. After a brief discussion of behaviourism, he characterises the Rylean language "as not only a *behaviouristic* language, but a behaviouristic language which is restricted to the *non-theoretical* vocabulary of a behaviouristic psychology".[50] Let us consider what a "behaviouristic language which is restricted to the non-theoretical vocabulary of a behaviouristic psychology" might look like. Prior to internalising Jones' theory, he and the rest of the Ryleans would have something that we might call a 'thinking-out-loud' theory of mind. That is, they possess (proto-) concepts of what it is to be confronted with a decision, deliberate, choose a course of action, and fail or succeed in following through on an intention, but as of yet, they are like children who have not yet learned to keep their thoughts to themselves. The Rylean conception of thought—of what it is to deliberate, to form an intention, to think through a problem, etc.—is limited to those episodes involving overt verbal performances.

While Jones' postulation of inner episodes whose intentionality parallels that of overt linguistic productions is significant, the key ingredient to answering the epistemic problem posed by the paradox is located in the Ryleans' adoption of

[50] Ibid., 186.

a meta-language. In so doing, they introduced ways of speaking *about* their linguistic performances: assuming that a language is a system of expressions governed by certain rules, by augmenting their prior language so that they are now also able to speak of word tokens and types, our mythical ancestors have adopted an *explicit system of rules*. They have thus developed *concepts* pertaining to following a rule—of making a mistake or doing something correctly, of an action being proscribed, permitted, or enjoined, and of verbal performances meaning or not meaning what they were *intended* to. As such, even though they were following rules all along, they are now consciously aware that they are (and were) doing so. With Jones' help, they have become consciously aware that they are consciously aware. It may be helpful to outline the key moves in our paraphrase of Sellars' myth in a series of steps:

1) At first encounter, the Ryleans spoke a language that allowed them to speak of public objects and events.

2) They subsequently developed a semantic meta-language that allowed them to expand their conceptual repertoire such that they were able to speak and think of their overt linguistic performances as being governed by certain rules.

3) Next, with the help of Jones they became able to notice and report their own intentional*, inner episodes.

4) Lastly, we can add the development in step (3) to the Ryleans' originally behaviouristic framework such that they develop a full battery of terms (and concepts) for intentions, volitions, and the like.

To sum up: the essential point to see is that a prior consciousness of the rule-fulness of overt linguistic behaviour (embodied in the story by the behaviouristic version of the concept of intelligence that derived from the introduction of the semantic meta-linguistic vocabulary) can provide the

precursors of the concepts of meaning something by a word and explicit intentions that are under attack in the Kripke-Wittgenstein paradox. More precisely, in the process of extending their language to include meta-linguistic discourse, our fictitious Rylean ancestors acquired concepts pertaining to intentionality and normativity. Concepts pertaining to intentions, meanwhile, could only have acquired their full meaning—that is, the meaning they ordinarily have for us as applicable to *mental* entities—once Jones had successfully taught his theory to his compatriots.

In the epilogue to the Myth of Jones, we can imagine that after teaching his theory to his fellows, the hero of our story disappears without a trace. He is remembered for a while, but as time moves on and subsequent generations come and go, memory of him and his accomplishments slowly fades from the collective Rylean consciousness. It eventually happens that young (neo-) Ryleans simply learn the modified language as if it had always been the way ours is now: they learn to speak (and think) of public objects, to make semantic statements in response to linguistic acts, and to give relatively reliable self-descriptions (that is, to possess self-knowledge). They forget that meta-linguistic categories and awareness of their 'inner lives' were originally a creative enrichment of the original Rylean framework of language and thought rather than merely *data* that were always simply *given* to their minds as conscious beings.

§ VII: Conclusion

We have now come full circle in our discussion of the Kripke-Wittgenstein

paradox. If our version of the Myth of Jones is an adequate reconstruction of the origins of our concepts of intentionality and normativity, we can now understand how we arrived at the place where Kripke's Wittgenstein started. We can now see how the development of an awareness of meaning and intentions could occur *prior* to the elaboration of a theory such as Millikan's. Indeed, we can see how the awareness *must* have predated the metaphysical account, for it had to be in place before we could ask the metaphysical question in the first place. And, most importantly, we can also see how the development of a meta-language along with Jones' theoretical achievement might have led us to the puzzle expounded by Kripke's Wittgenstein.

Conclusion

§ I: Introduction

The scepticism about meaning and intentions got off the ground with the observation that there is no meta-linguistic level to which we can appeal as the ultimate authority for questions concerning what we mean or intend. The sceptic's original query put pressure on the supposition that there is indeed a fact of the matter as to whether my intention to follow the rule for addition has determinate content. We saw that specifying the content of my intention by way of an appeal to a mathematical proof or algorithm would get us nowhere, because the sceptic could simply re-direct his questioning onto the content of the proof or set of instructions.

The arithmetical example was simply a special case of the more general argument that runs as follows: 1) in order for me to *intend* to follow a rule or *mean something* by a word (which is a special case of intending to follow a rule), there must be a *determinate content* to my intention; otherwise it would make no sense to say that my actual behaviour—the behaviour that is supposed to 'make good' on my intention—is correct or incorrect; that is, it would make no sense to say that it accords with my intention, or that I said what I *meant* to say, 2) since any rule is open to myriad interpretations of which only one can be correct, it seems that whenever I intend to follow a rule, I must have a meta-linguistic rule in mind that determines which interpretation is in fact the correct one, 3) but this can only lead to a regress since the meta-linguistic rule is in turn subject to as many interpretations as the original rule, thus I would need to know the meta-meta-

linguistic rule to specify the correct interpretation of the meta-linguistic rule, and so on, 4) therefore, Kripke's Wittgenstein concluded, since there can be nothing about my occurrent mental states—no ultimately authoritative meta-linguistic representation—there is no fact of the matter as to what I intend when I mean to follow a rule; anytime I purport to follow a rule or mean something by a word, my action is in fact arbitrary.

By grounding acts of meaning and intending in evolutionary history, Millikan's solution to the paradox showed that there could indeed be facts about what we mean by an expression or intend when purposing to follow a rule. In doing so, it salvaged the belief that our intentions and thoughts *do* have determinate content, and it set out the basis of an account of what that content is; but almost paradoxically, it rendered possession of that very belief problematic. The primary point of emphasis in chapter two not only opened the way for Millikan's particular solution to the paradox but also revealed that a singular account such as hers does not fully dissolve the Kripke-Wittgenstein paradox. Recall what Millikan took to be the upshot of the rule-following considerations:

> What lies before consciousness does not determine its own significance; knowing what one means is not a matter, merely, of apprehending the contents of one's mind. In short, meaning is neither a state of awareness nor an epistemological given. It does not occur encapsulated within consciousness; it is not a state that simply *shows* its content or its significance.[51]

In other words, the contents of our vocalizations, inscriptions, intentions—of our representations generally—*are not epistemically given*. Having a thought is not sufficient for being consciously aware of that thought. From here, we saw that the

[51] Millikan, "TRHKWP," 326.

sceptical paradox arose out of an illicit conflation of the metaphysics and epistemology of content, and that the Kripke-Wittgenstein puzzle thus bifurcates into two distinct questions. The metaphysical question—to which the biological account was an answer—asks how or why it is that our mental states (including acts of meaning and intending) represent. But since meaning is "neither a state of awareness nor an epistemological given", the epistemological question—the question of how we can be able to have knowledge of the content of our mental states—remained to be answered.

The epistemological problem was our main concern in the previous chapter. The problem was to explain—in light of the biological account of the metaphysics of rule-following—how we could develop concepts pertaining to thought, norms, rules, etc. such that we could be aware that our thoughts have content and have some sense of what that content is. On the Sellarsian picture, such concepts begin to develop through the course of learning to say things about public linguistic acts. Concurrent to learning to make semantical statements, we develop notions of correctness and incorrectness and of what it is for linguistic performances to mean things. In addition, we learn to apply the intentional* categories of the semantic meta-language to our thoughts (initially perceived as verbal propensities), thus becoming able to make reliable self-reports. Taken together, these two developments in our conceptual framework allow for being aware of ourselves as creatures that have intentions to follow rules and that mean things by the words we speak *prior* to the elaboration of the biological theory which explains the metaphysics of intentionality.

§ II: What of the Sceptical Solution?

Before we conclude, let us briefly return to Kripke's presentation of Wittgenstein's paradox and solution. If we accept Millikan's and Sellars' views with respect to the metaphysics and epistemology of content respectively, it seems that we have endorsed both a straight *and* a sceptical solution to the paradox. Insofar as it demonstrates how our thoughts can have determinate content, Millikan's theory is certainly a straight solution. But the Sellarsian theory bears a striking resemblance to the sceptical solution attributed to Wittgenstein, for it locates our original awareness of the categories of normativity and intentionality in the rule-ful and representational (i.e., semantic) properties of overt linguistic expression, and on Sellars' view, language games and the conceptual thinking that comes with them are *essentially* social phenomena. Consider the following passages from two series of lectures he gave:

> We can imagine a child to learn a rudimentary language in terms of which he can perceive, draw inferences, and act. In doing so, he begins by uttering noises which *sound like* words and sentences and ends by uttering noises which *are* words and sentences. We might use quoted words to describe what he is doing at both stages, but in the earlier stage we are classifying his utterances as *sounds* and only by courtesy and anticipation as *words*.
> Only when the child has got the hang of how the sounds function in the language can he be properly characterized as saying 'this is a book', or 'it is not raining', or 'lightning, so shortly thunder', or 'you spanked me, so you don't love me'.[52]

> Now the more we know about a person, the better we are able to judge what…he would be likely to say (think-out-loud)—if he were in a thinking-out-loud frame of mind. It is obviously difficult to be accurate about this, particularly when we are dealing with sophisticated minds. But even here the difficulty is one in *practice* rather than of *principle*. And

[52] Wilfrid Sellars, "the Structure of Knowledge" [hereafter "SK"], in *Action, Knowledge and Reality*, ed. Hector-neri Castaneda (Indianapolis, Indiana: The Bobbs-Merrill Company, 1975), 320.

> when it is children in the initial stages of learning a language who are the subjects, the difficulty in *practice* is substantially less than it becomes subsequently when they have learned to lie, deceive, and conceal their thoughts.
> Can we not, as children, be trained by those who know us intimately (our parents), and who therefore know (*ceterus paribus*) what our short-term verbal propensities are (i.e., what we are thinking), to *respond* reliably to our own short-term propensities to say that-*p*, as well as to respond to our actual sayings of '*p*'?
> And can not this ability be generalized in such a way that we can reliably respond to new propensities, i.e., to thoughts other than those in terms of which we have been trained?[53]

and finally,

> [T]here is…a sound score to the idea that while reality is the 'cause' of the human conceptual thinking which represents it, this causal role cannot be equated with a conditioning of the individual by his environment in a way which could in principle occur without the mediation of the family and the community. The Robinson Crusoe conception of the world as generating conceptual thinking directly in the individual is too simple a model.[54]

Sellars thought that linguistic and conceptual norms were initially imbibed by an individual through the conditioning and training role played by her or his linguistic community, and moreover, that the community's practices were also the source of the individual's *awareness* of those norms. Recall Kripke's interpretation of Wittgenstein's sceptical solution:

> [Wittgenstein] simply points out that each of us *automatically* calculates new addition problems (without feeling the need to check with the community whether our procedure is proper); that the community feels entitled to correct a deviant calculation; that in practice such deviation is rare, and so on. Wittgenstein thinks that these observations about sufficient conditions for justified assertion are enough to illuminate the role and utility in our lives of assertion about meaning and determination of new answers. What follows from these assertability conditions is *not* that the answer everyone gives to an addition problem is, by definition, the

[53] Ibid., 326-7.
[54] Wilfrid Sellars, "Philosophy and the Scientific Image of Man," in Sellars, *Science, Perception, and Reality*, 16.

> correct one, but rather the platitude that, if everyone agrees upon a certain answer, then no one will feel justified in calling the answer wrong.[55]

If Kripke's Wittgenstein had not already concluded that intentions could not have determinate content (and thus that there really were no such things), he might very well have used the observations highlighted here as a basis for an epistemological theory akin to Sellars'. Essential to both the Sellarsian and sceptical solutions is the pride of place they give to the public, social element involved not only in the course of developing an *awareness* of the rule-governed, representational features of language and thought, but also in *structuring* that same space.

Although Kripke never explicitly rejects the possibility of a dual-component solution, such is clearly implied by the opposing, dichotomous nature of his presentation of the two forms of solution. Does this point to an underlying incompatibility between the metaphysical and epistemological components to the solution that I have endorsed? The answer to this question may depend on implications of both that have not been filled in. Although the theory of truth rules that Millikan elucidates from her response to the Kripke-Wittgenstein paradox is definitely incompatible with the variety of verificationism that seems implicit in the Sellarsian and Wittgensteinian accounts of linguistic rules, this too—i.e., the apparent incompatibility—would seem to stem from a conflation of the metaphysics and epistemology of content. Recall how Millikan stressed the necessary role of proximal assertibility rules to *back* distal correspondence rules. On the biological model, the awareness we have of assertibility conditions is the evidence upon which the theorized grasp of correspondence rules is based.

---

[55] Kripke, *WRPL*, 111-2.

Taking this in conjunction with our oft emphasised point that awareness of meaning or content is not guaranteed by the mere capacity to represent, there is no immediately obvious problem in uniting the metaphysics of Millikan's straight solution with the epistemology of the Sellarsian/Wittgensteinian sceptical solution.

§ III: Thought-about-thought, One Last Time

If the Kripkensteinian rule-following considerations are in fact an appropriate lens with which to focus a discussion of the metaphysics and epistemology of thought, and if the arguments I've presented with respect to the paradox are sound, then we are now in a position to better appreciate the introductory comments made at the outset of the first chapter. The sceptic's conclusion did not merely threaten philosophical preconceptions concerning our ability to follow rules. If it had proven irrefutable, the sceptic's focussed attack on the content of intentions would have been devastating to the very possibility of representation. It turned out, however, that his conclusion depended on an illegitimate assumption, namely, that *if* our mental states *do* have determinate content, *then* that content must be epistemically given. Denying this assumption on the grounds of the very same rule-following considerations relied-upon by the sceptic paved the way for Millikan's biological solution, but further reflection showed that Kripke's epistemic requirement could not simply be left aside. If the content of our mental states can only be uniquely determined by factors that are not available to consciousness, the possibility of self-knowledge is severely threatened. Not only

did the biological solution make it entirely mysterious how we could have *conceived* of the sceptical paradox in the first place, but also, to echo Kripke again, "[t]he idea that we lack 'direct' access to the facts whether we mean plus or quus is bizarre in any case. Do I not know, directly, and with a fair degree of certainty, that I mean plus?"[56] With the aid of Sellars' myth, we saw how Kripke's epistemic requirement could be met without relying on the dubious, Cartesian assumption that to have thoughts is to know them.

If, as Sellars' myth suggests, explicit awareness of intentional*, rule-governed phenomena originally stems from learning to speak about speech, the implications for self-knowledge (or, as I've been calling it, thought-about-thought) are quite significant. In my introduction, I claimed that this essay would help us to make sense of the fact that people often do things for reasons other than those they themselves would give even when speaking (or thinking) sincerely, and this may seem at odds with my efforts to show how self-knowledge is possible. However, the positive Sellarsian account of self-knowledge is compatible with the former claim. That is, there is room here for a sort of strong fallibilism with respect to self-knowledge. Although fully-fledged language users will nearly always know that they mean to follow the rule for addition or that they are thinking of a pink ice cube (such mundane cases surely account for the vast majority of thoughts, thus our fallibilism should not be thought of as 'strong' in the proportional sense), accurate self-knowledge about other, arguably more significant, dimensions of experience can be much more difficult. For instance, knowing one's own motives in cases of interpersonal conflict can often be

[56] Ibid., 40.

difficult to ascertain—even when one is trying to be honest with her or his self. (Indeed, it's these and other such difficulties in knowing oneself that keep clinical psychologists and counsellors and therapists of all kinds in business.)

Or consider a more political example: think of a man who regularly and fairly consistently treats his wife as inferior and subordinate to himself. Depending on his mood, he either humours her or outwardly belittles or silences her when she voices her opinion in a conversation. If in need of something (perhaps a utensil or condiment) when already seated down to supper, he asks his wife—who herself may have just sat down after placing the final dish on the table—to retrieve it for him. In general, he treats her as an assistant or supporter for his endeavours; he treats her as though it is *her* purpose to assist him in achieving *his* purposes. If asked, such a man might very well say that he considers his wife to be his equal—that she is every bit as valuable a person as he. If pressed with particular examples of his oppressive, subordinating behaviour, he might say that while he and his wife are equal, they have different roles. He might say that he is the bread-winner and head of the family, not only in the sense that he is in charge, but also in the sense that, as a man, he is less emotional and better equipped for tasks requiring rational thought. He might go on to say that, complementarily, his wife's role (due to her feminine attributes) is to nurture and raise the children, keep the house, and prepare the meals. It's not insignificant that in most cases, wives of men such as this share virtually the same outlook—they would say both that they are their husbands' equals and also that they and their husbands are appropriately suited to their respective roles. And what's

more, both husband and wife could very well say all of this with complete sincerity: if confronted with the obvious lack of true commitment to their egalitarian espousals, they might *honestly* respond with confusion, disbelief, indignation, or a combination of the three. It's not at all uncommon for people who verbally espouse equality between different sexes (or races or socio-economic groups or what-have-you) to live and think in ways that belie their egalitarian espousals.

Rather than merely attack the inconsistencies exhibited in such ways of thinking, many feminists (especially those with a Marxist bent) have persuasively questioned the veracity of such sincerely made avowals and shown that gender-essentialist ways of thinking and speaking are more accurately described as a tool for maintaining status quo power relations. It's not too big a stretch to construe the feminist arguments as a challenge to the self-knowledge of people with such ways of thinking.

Let's reflect on this example in light of the intentionality considerations brought to the fore by this project. Parallel to the example of the male hoverfly, our being able to grasp and follow proximal assertibility rules is never a guarantee for successful conformity to distal correspondence rules. So it is entirely possible for us to systematically make sincere but false avowals when, for whatever reason, we are obtuse to the actual purpose or function of a particular dimension of discourse.

And this is what I take to be one of the most significant lessons of the Sellarsian and Wittgensteinian epistemologically-oriented views on language and

thought: the language games we play have deep and systematic effects on the ways we perceive ourselves, others, and the world. In this project I hope to have shown how this is possible while still countenancing something along the lines of Millikan's realist approach to intentionality and normativity.

There are many more implications for the epistemology of the mind sketched in the preceding pages—full discussion of which will have to wait for other occasions. Self-knowledge does not merely consist in data that are *given*; rather, knowing oneself is a skill—something that can be done well or poorly, correctly or incorrectly. In turn, honest self-representation is not simply a matter of saying what one believes to be true about oneself, for it is dependent upon (among other things) one's *aptitude* for self-knowledge. On a somewhat different note, this account of the metaphysics and epistemology of thought also raises questions concerning the precise relationship between language and thought—especially regarding how the former might mould the latter and thus have an integral place in a metaphysics of self-consciousness.

Bibliography

Blackburn, Simon. "The Individual Strikes Back." in Miller and Wright, 28-44.

Boghossian, Paul A. "The Rule-Following Considerations." *Mind* 98, no. 392 (1989): 507-549.

Bretherton, Inge and Marjorie Beeghly. "Talking About Internal States: the Acquisition of an Explicit Theory of Mind." *Developmental Psychology* 18, no. 6 (1982): 906-921.

Carruthers, Peter. *Language, Thought and Consciousness: an Essay in Philosophical Psychology*. Cambridge: Cambridge UP, 1996.

Child, William. "Memory, Expression, and Past-Tense Self-Knowledge." *Philosophy and Phenomenological Research* 73, no. 1 (2006): 54-76.

Chisholm, Roderick M. and Wilfrid Sellars. "The Chisholm-Sellars Correspondence on Intentionality." in *Intentionality, Mind, and Language*, edited by Ausonio Marras, 214-248. Chicago: University of Illinois Press, 1972.

Flavell, John H. "Cognitive Development: Children's Knowledge about the Mind." *Annual Review of Psychology* 50, (1999): 21-45.

Fodor, Jerry. "Having Concepts: a Brief Refutation of the Twentieth Century." *Mind & Language* 19, no. 1 (2004): 29-47.

---. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press, 1990.

Forbes, Graeme. "Skepticism and Semantic Knowledge." in Miller and Wright, 16-27.

Hattiangadi, Anandi. *Oughts and thoughts: rule-following and the normativity of content*. New York: Oxford University Press, 2007.

Harman, Gilbert. "Three Levels of Meaning." *The Journal of Philosophy* 65, no. 19 (1968): 590-602.

Kripke, Saul A. *Wittgenstein on Rules and Private Language*. Oxford: Basil Blackwell, 1982.

Kusch, Martin. *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein*. Montreal: McGill-Queen's UP, 2006.

Loar, Brian. “Critical Review of Saul Kripke’s *Wittgenstein on Rules and Private Language*.” *Nous* 19, no. 2 (1985): 273-280.

McDowell, John. “Wittgenstein on Following a Rule.” *Synthese* 58 (1984): 325-63.

Miller, Alexander & Crispin Wright, eds. Rule-Following and Meaning. Montreal: McGill-Queen’s UP, 2002.

Millikan, Ruth Garrett. “How We Make Our Ideas Clear: Empiricist Epistemology for Empirical Concepts.” *Proceedings and Addresses of the American Philosophical Association* 72, no. 2 (1998): 65-79.

---. *On Clear and Confused Ideas*. New York: Cambridge UP, 2000.

---. “The Father, the Son, and the Daughter.” *Pragmatics and Cognition* 13, no. 1 (2005): 59-71.

---. “The Price of Correspondence Truth.” *Nous* 20, no. 4 (1986): 453-468.

---. “Thoughts Without Laws: Cognitive Science with Content.” *The Philosophical Review* 95, no. 1 (1986): 47-80.

---. “Truth Rules, Hoverflies and the Kripke-Wittgenstein Paradox.” *The Philosophical Review* 99, no. 3 (1990): 323-353.

---. *Varieties of Meaning*. Cambridge, MA: MIT Press, 2004.

Millikan, Ruth Garrett & Andrew Woodfield. “Knowing What I’m Thinking Of.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 67, no. (1993): 91-124.

Schwitzgebel, Eric. “The Unreliability of Naive Introspection.” *Philosophical Review* 117, no. 2 (2008): 245-273.

Sellars, Wilfrid. “Empiricism and the Philosophy of Mind.” in Sellars (1963), 127-196.

---. “Philosophy and the Scientific Image of Man.” in Sellars (1963), 1-40.

---. *Science, Perception and Reality*. Atascadero, California: Ridgeview Publishing Company, 1963.

---. “Some Reflections on Language Games.” in Sellars (1963), 321-358.

---. “the Structure of Knowledge.” in *Action, Knowledge and Reality*, edited by Hector-neri Castaneda, 295-346. Indianapolis, Indiana: The Bobbs-Merrill Company, 1975.

Wittgenstein, Ludwig. *Philosophical Investigations*. translated by G.E. Anscombe. Oxford: Blackwell, 2001.

Wright, Crispin. “Kripke’s Account of the Argument Against Private Language.” *The Journal of Philosophy* 81, no. 12 (1984): 759-778.

---. “Meaning and Intention as Judgment Dependent.” in Miller and Wright, 129-140.

---. “Rule-following Without Reasons: Wittgenstein’s Quietism and the Constitutive Question.” *Ratio* 20, no. 4 (2007): 481-502.

---. “Wittgenstein’s Rule-following Considerations and the Central Project of Theoretical Linguistics.” in *Reflections on Chomsky*, edited by Alexander George, 233-64. Oxford: Basil Blackwell, 1989.