

WITTGENSTEIN ON MIND AND BODY

WITTGENSTEIN ON THE MIND-BODY RELATIONSHIP

By

JESSICA MURPHY, B.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts

McMaster University

© Copyright by Jessica Murphy, September 2006

MASTER OF ARTS (2006)

McMaster University

(Philosophy)

Hamilton, Ontario

TITLE: Wittgenstein on the Mind-Body Relationship

AUTHOR: Jessica Murphy, B.A. (University of Ottawa)

SUPERVISOR: Dr. Jill LeBlanc

NUMBER OF PAGES: v, 119

Abstract

The purpose of this thesis is to examine Wittgenstein's treatment of the mind-body relationship and the problems associated with articulating this relationship. It is important to note that Wittgenstein does not attack Cartesian dualism in the typical materialist fashion, by arguing against the likelihood that there might be causality across metaphysical gaps, or by insisting that psychological terms in fact refer to the body. His strategy, rather, is to reject the conception of the body that is at the heart of both dualism and materialism: a picture of the human form and its behaviour as expressively vacant, with no significant role to play in our account of mentality, of the mind as a realm of private experience behind that behaviour, and of our experience of one another as mediated by hypotheses. I argue that Wittgenstein calls into question the assumptions of the mechanical body and the hidden mind by reminding us of our ability to just *see* others as minded, to see their behaviour as expressive of a mind. Recognizing this expressive relation between our mental states and our outer behaviour, a relation that is neither that of a contingent causal connection nor that of identity, I suggest, is vital to dissolving the whole complex of mind-body problems, at least insofar as this can be done.

Acknowledgements

I would first like to thank my supervisor, Dr. Jill LeBlanc, for her support, guidance, and patience. I would also like to thank my second reader, Dr. Nick Griffin, for his helpful corrections in the short time I allotted him.

Dr. Andrew Lugg at the University of Ottawa first ignited my interest in Wittgenstein with his level-headed enthusiasm for his later work. Without Dr. Lugg's well-timed support and kind words I could not have made it this far.

I would also like to express my gratitude to Ronnie Abou-Abssi, for never failing to make me smile.

Finally, I would like to thank God, and St. Jude, for carrying me through.

Table of Contents

Introduction

Chapter 1: Wittgenstein and Skepticism

1.1 Kripke on Private Language.....	8
1.2 Wittgenstein and “The Argument of the Ordinary”.....	17
1.3 The “Intellectualist” Picture in Philosophy.....	23
1.4 A Craving for Generality.....	34
1.5 Conclusion.....	41

Chapter 2: The Private Language Argument

2.1 Language and Behaviour.....	45
2.2 “There is not enough regularity to call it a ‘language’”.....	47
2.3 Private Objects.....	53
2.4 The “No Stage-Setting” Argument.....	56
2.5 Private Language and the Problem of Other Minds.....	69
2.6 Conclusion.....	75

Chapter 3: The “Hard Problem” of Consciousness

3.1 The Hard Problem.....	78
3.2 “Explaining” Consciousness.....	83
3.3 Consciousness and Behaviour.....	89
3.4 Behaviourism.....	96
3.5 Functionalism.....	100
3.6 Conclusion.....	112

Conclusion

Bibliography

Introduction

The purpose of this thesis is to examine Wittgenstein's treatment of the mind-body relationship. The problem of reconciling our ordinary experience as embodied agents with scientific accounts of the world, and in particular, the human body, has been the source of much drama in the philosophy of mind. This problem, at least in its modern form, can be traced back to the seventeenth century and the rise of mechanistic physics. Prior to Descartes, Charles Taylor writes, the material universe was regarded "as a kind of medium, in which psychic contents like heat and pain, or the supposed Forms or Species of scholastic tradition, could be lodged or embodied or manifest themselves."¹ Meaning, according to this expressive conception of nature, was unproblematic; it was something that could simply be *perceived*. Against this pre-modern view, however, Galileo and his contemporaries (including Descartes) argued that the material universe should be understood mechanistically; as Galileo famously put it, the natural world is "written in the language of mathematics." With this came the idea that the human body, too, might be understood in a purely scientific way, according to the law-governed regularities of cause and effect, without any mention of mentality. As it turns out, however, our first-person experience of consciousness proves stubbornly resistant to such treatment; our conscious experience simply does not seem reducible to the causal processes of nature. The question then arose as to how the conscious mind might fit into the universe as described by natural science. Given the gulf between mind and matter in

¹ Charles Taylor, *Sources of the Self: The Making of Modern Identity* (Cambridge, MA: Harvard University Press, 1989) 146.

these areas, Descartes quite sensibly proposed a dualist scheme, holding that the mind and body were two independent yet causally-interacting substances. Thus he had saved the mind from the reductive net of efficient causation, but at the price of making the mind a radically private thing, bearing only a contingent, external relation to the body. Two questions then arose. First, how can we know that there are minds inside these mechanical bodies? Second, given that there is mind as well as body, how are the two related? Rather than de-mystifying the relation of mind to matter, Descartes had opened an evidential gap between the two which seemed to render the mind-body problem irresolvable.

The conflict between a subjectivism which fades ultimately into solipsistic isolation and an objectivism which sees everything through the lens of impersonal science has dominated the philosophical agenda ever since, and nowhere so obviously as in the philosophy of mind. Wittgenstein, I suggest, can be seen as trying to dissolve the classic problems of mind and body by starting, not with first-person introspection nor third-person science, but with second-person inter-subjectivity – that is, with a careful look at our ordinary human interaction in which these problems simply do not show up. It is this attention to our everyday engagement with one another that prompts Wittgenstein to propose a *conceptual* connection to hold between the mind and the body. His insistence that “an inner process stands in need of outward criteria” (PI 580) is often taken to imply such a connection, but there is little agreement as to just how it is to be understood. I suggest that this relation should be construed as an *expressive* one. In particular, Wittgenstein attempts to call into question the Cartesian assumptions of the

mechanical body and the detached mind by reminding us of our ability just to *see* others as minded, to see their behaviour as expressive of an inner life. Recognizing this expressive relation between our mental states and our outer behaviour, a relation that is neither that of a contingent causal connection nor that of identity, I suggest, is vital to resolving the whole complex of mind-body problems, at least insofar as this can be done.

Chapter one examines Wittgenstein's view of language and meaning while clarifying his attitude toward skepticism. Kripke's controversial interpretation of the rule-following remarks in the *Philosophical Investigations* provides a good opportunity to do both. In *Wittgenstein on Rules and Private Language*, Kripke reads Wittgenstein as mounting a powerful skeptical challenge to the possibility of meaning and understanding, and then going on to resolve it himself by way of a "skeptical" solution. His mistake is to presume that Wittgenstein's association of the meaning of a word with its use in the language was meant to suggest a theory of language in which meaning is explained in terms of some independently available notion of use or practice. The rule-following remarks, however, are an investigation into our criteria for ascribing the understanding of a word or of an expression of a rule to someone; they are not a search for what justifies those criteria. For Wittgenstein, our practices are permeated by meaning – they are not something more basic in terms of which meaning can be justified. By failing to recognize the reciprocity between the rule and the practice, we come to imagine that meanings are either autonomous in a super-empirical, "Platonic" sense or else altogether illusory.

Chapter two focuses on the private language argument and its significance for the problem of other minds. As mentioned above, this problem gets started with the assumption that I know my own mind directly and others' only by analogy; thus I arrive at an understanding of other peoples' mental states via a route of circuitous inference. On this view, psychological terms acquire the meaning they have by referring to inner states of private awareness. The effect of the private language argument is to expose this picture as a distortion of our ordinary understanding of ourselves and of each other. The mental states of others are accessible to us not because they constitute the best explanation of outer behaviour, but because they make themselves manifest in behaviour. It is just for this reason that we are able to have a common language of mental states at all. It is only by recognizing this language as essentially tied to patterns of practical activity that mysteries about mental concepts, and our inner lives in general, begin to unravel.

In chapter three, I examine the significance of these considerations to contemporary philosophy of mind. Despite mainstream philosophy's official rejection of substance dualism, the Cartesian picture of the person continues to have a powerful hold on our imaginations. The view of the physical world as purely mechanical, and of the self as autonomous, able to shrug off its contexts at will, remains deeply entrenched in the discipline. Indeed, what the prevailing materialist theories of mind often quite explicitly offer is a "naturalized" Cartesianism. It should not be surprising, then, that many of the classic problems which plagued Descartes have survived the transition to naturalism. In its updated form, the mind-body problem shows up as "the hard problem

of consciousness”. The problem is aptly summed up by David Chalmers: “Consciousness”, he writes, “is the largest outstanding obstacle [to] a scientific understanding of the universe.”² That is, while the brain and nervous system can be subjected to the methods of natural science, our subjective experience seems to resist such treatment. This, for many philosophers, seems to be enough to qualify consciousness as “a mystery”; the thought that we might have some non-scientific understanding of ourselves and of each other – that there might be some things about which it is “true to say that our considerations could not be scientific ones” (PI 109) – does not even show up in contemporary philosophy of mind as an option to be argued against. In response, I argue that a perfectly intelligible explanation of consciousness can be given by showing how we ordinarily employ the word.

² David Chalmers, *The Conscious Mind*, (Oxford University Press, 1996) xi.

Chapter 1: Wittgenstein and Skepticism

The purpose of this chapter is to clarify Wittgenstein's attitude toward skepticism. It will be helpful to do this straight away, I think, because despite the interpretive attention that has been lavished upon Wittgenstein's later work, his precise position on the matter remains elusive. Indeed, he has been claimed to have held just about every attitude that one might possibly have toward it, from "common sense realism" in the spirit of G. E. Moore to "contemporary idealism".³ Such claims might already sound odd to one who knows Wittgenstein as a philosopher who saw his task as one of "assembling reminders",⁴ like this one:

[T]he idealist will teach his children the word "chair" after all, for of course he wants to teach them to do this and that, e.g. to fetch a chair. Then where will be the difference between the idealist-educated children and the realist ones? Won't the difference only be one of battle cry? (Zettel 414).

In this passage, and others,⁵ Wittgenstein's message is that the divide between the realist and idealist camps is, practically speaking, uninteresting. Thus, while I do not suggest that Wittgenstein endorsed skepticism in any of its forms, nor do I suggest, as some philosophers do, that he attempted some straightforward refutation of it.⁶ Rather, his strategy was to remind us that in ordinary contexts, we are – groundlessly – taking for

³ In *The View From Nowhere*, for example, Thomas Nagel describes Wittgenstein as "one of the most important sources of contemporary idealism" (Oxford: Oxford University Press, 1986) 105.

⁴ Oswald Hanfling, *Wittgenstein and the Human Form of Life* (New York: Routledge, 2002) 74.

⁵ For example, Zettel 413, *On Certainty* 476, and the *Blue Book* p. 48: "...the common-sense man ... is as far from realism as from idealism..."

⁶ G. N. Schlesinger is one philosopher who takes Wittgenstein to have tried (unsuccessfully) to mount a straightforward "criterial" argument against other-minds skepticism. See his *Metaphysics* (New Jersey: Barnes and Noble Books, 1983) 147.

granted a framework of belief and practice which already precludes the truth of skeptical scenarios, and it is only for this reason that I am able to speak intelligibly about language and its relation to the world in the first place. For Wittgenstein, then, the most fundamental level at which we can discuss why we can understand each other – why our meanings line up – is normative. Statements at this level are not ordinary assertions, but rather grammatical remarks that present to us the form of our discussion. If we get clear on the role of these remarks, then we will not be tempted to construe them as necessary empirical truths, nor will we be puzzled when our ordinary assertions seem to lack the necessity that belongs to them. Rather, we will recognize that there is a point at which our “explanations” cease to actually explain anything and we ought to just say “this language game is played” (PI 654).

My discussion of skepticism in this chapter will focus mainly on Kripke’s controversial interpretation of Wittgenstein’s discussion of rule-following in the *Philosophical Investigations*. In his book *Wittgenstein on Rules and Private Language*, Kripke suggests that meaning, for Wittgenstein, is explained in terms of conformity to community practices, which are not themselves understood as embodying meanings. Section one of this chapter will be devoted to explaining Kripke’s interpretation of the rule-following remarks. This interpretation, most philosophers agree, is wrong, but it is informatively wrong. Drawing upon criticisms of his interpretation by several philosophers, most notably John McDowell and Stanley Cavell, in section two I will examine the source of Kripke’s misreading and what it can teach us about Wittgenstein’s intention in the rule-following remarks. Following Cavell, I argue that Kripke’s error lies

largely in his failure to take seriously the argumentative and rhetorical style of the *Investigations*, the fundamental feature of which Cavell calls “the argument of the ordinary”. As a result, Kripke takes what Wittgenstein means to be a rejection of the picture of language that frames the dispute about the nature of rules as a move within it. In section three, I argue that the rule-following remarks in the *Investigations* can be seen as a philosophical response to what Wittgenstein takes to be one of the main coordinating themes of modernity, what Charles Taylor calls the “intellectualist” or “disengaged” view of human intelligence, or, to borrow a phrase from Cavell, the idea that our primary relation to the world is one of knowing. In the rule-following remarks, this picture shows up as a demand for an account of language which presupposes nothing about us, which operates independently of the pattern of actual responses that characterize a participant in that activity. When our actual practice fails to live up to the requirements imposed by this picture, we are thrown into skepticism. In the final section of this chapter, I support my suggestion that picture this is the real target of the rule-following remarks via a disagreement with Cavell. Cavell suggests that Wittgenstein viewed skepticism as a natural and inescapable feature of our life with language. By contrast, I think that there is textual support to suggest that Wittgenstein viewed skepticism as an outgrowth of the scientific worldview and the craving for generality that so often accompanies it.

1.1 Kripke on Private Language

The rule-following remarks are best read in relation to Wittgenstein’s later philosophy of language, and, in particular, his association of meaning with use. One of

the earliest expressions of his later thought – the *Blue Book* of 1933-34 – opens with the question “What is the meaning of a word?” (BB p. 1). In the *Tractatus*, he offered the following answer: “A name means an object. The object is its meaning” (T 3.203). By now, however, Wittgenstein had grown wary of the question. Like “What is length?” and “What is the number one?”, the question “What is meaning?” produces in us a mental cramp, because it sends us immediately in misguided pursuit of some substance or other that corresponds to the noun. At once, “we feel that we can’t point to anything to reply ... and yet ought to point to something” (BB p. 1). It is as if the “meaning” were something alongside the sound or the mark on the paper, something immaterial that must be added to the dead sign to bring it to life. Around the time that Wittgenstein was writing, for example, philosophers like Frege, Meinong, Husserl, and Moore (along with Wittgenstein himself, in his early work) conceived meaningful words as “names” for something else (objects, persons, values, and so on).⁷ Each of these theories is rooted ultimately in the “Augustinian” picture that Wittgenstein attacks throughout the *Investigations* – the idea that every word has a meaning correlated with it, and that this meaning is the object for which the word stands. This view recalls aspects of Wittgenstein’s own early theory of meaning developed in the *Tractatus*, which portrayed all propositions as ultimately composed of simple names whose meaningfulness is determined by the nature of the simple object for which each name stands.

In his later work, Wittgenstein abandoned this view, famously stating that, “[f]or a large class of cases – though not for all – in which we employ the word ‘meaning’ it

⁷ In the *Tractatus*, Wittgenstein had already begun to move away from the theory of naming; it was maintained only with respect to the “objects” that may occur as elements in a proposition.

can be defined thus: the meaning of a word is its use in the language” (PI 43).⁸ Much of the *Investigations* devoted to replacing his early theory of meaning – and the notion of meaning in general – with a description of the use or meaning of words in terms of their function in the various articulations of human forms of life. For example, certainly “Fred”, can be the name of the dog Fred and “N.C.C.” can be the name of the organization that looks after federal lands and buildings in Ottawa, but even within such language-games of giving names it is clear that resemblances are often remote, that “Fred” functions very differently as a name than “N.C.C.”. While a cursory look at surface grammar would have us suppose that these words were equivalent in meaning, the differences become evident as soon as we turn our attention to their use and rules of application. Wittgenstein’s point in turning our attention in this way, however, is not to suggest an alternative theory of meaning to be summed up under the slogan “Meaning is Use” – for the fact that words are used is obvious – but to release us from the intellectual fetish of *the* meaning itself; “it is that there is *no more* to the meaning of a word than its use in the language.”⁹

⁸ This statement is often taken to indicate that Wittgenstein held a “Use Theory” of meaning; however, in the same passage, Wittgenstein quickly goes on to state that “the meaning of a name is sometimes explained by pointing to its bearer” (PI 43). The fact that he states that it is the case often but not always that the meaning of the word is given by its use and then immediately notes an exception suggests that he did not hold that, as a general theory, the meaning of a word is always identified with its use. What he does say is that, while in many instances one can come to know the meaning of a word by coming to understand how it is used, in some cases – for example, cases in which the object is close by – the meaning can be explained by simply pointing and saying “That is...”. The passage is better read as warning against the temptation to look outside the language game for the meaning of a word, for example, to the object that it names – in which case the meaning could be divided, destroyed, etc. – or to some psychological feeling – in which case we would have to be lucky enough to all possess same feeling in order for communication to be possible.

⁹ Hanfling (2002: 90).

If the meaning of a word is its use in the language, then understanding that word is just the ability to go on using it in the right way. This proposal flies in the face of the intuitively tempting thought that understanding is something more basic in terms of which this ability can be explained. We might insist, for instance, that the ability to apply a word properly stands in need of an explanation in terms of one's mind being in contact with something – the “meaning” itself – that enables her to go on in this way. Language is a rule-governed activity, and it is essential that words can be used either rightly or wrongly, in accord or discord with the rules for their use. How is it, we might ask, that we are able to follow a rule – to go on, not just in any way, but *correctly*?

In *Wittgenstein on Rules and Private Language*, Kripke reads Wittgenstein as arriving finally at a “skeptical paradox” which arises from his rejection of a series of attempts to answer this question and which purports to show “all language, all concept formation, to be impossible, indeed unintelligible.”¹⁰ Most of the time, Kripke assumes that this fact would have to be a mental state of some sort, and concludes that there is simply no such mental state as “understanding a rule” or “grasping the meaning of a concept”. Such a state would have to stand in normative relation to our behaviour, in the sense that it would have to embody some standard directing which courses of action count as in accord with the rule or meaning and which count as going against it. Kripke interprets Wittgenstein as arguing that no mental state could have these normative

¹⁰ Saul Kripke, *Wittgenstein on Rules and Private Language* (Cambridge: Harvard University Press, 1982) 62.

implications, since any such state would be open to an indefinite variety of conflicting interpretations.

Kripke's argument goes roughly as follows: the skeptic asks how I can be sure that I have been acting in accordance with the function "plus" in my addition calculations, rather than the function "quus", where "quus" is defined as, "*x quus y = x plus y, if $x, y < 57$; otherwise $x quus y = 5$* ". So far, all of my calculations have involved variables that were less than 57; thus my performance is consistent with either function. Kripke's point is that there is no discernible fact associated with my mental act of meaning "plus" which would preclude it being interpreted as "quus". To put the problem more generally, the result is that a given rule in fact rules out nothing, since for any course of action there is some interpretation of the rule such that my action is in accord with it. This will be true no matter how meticulously the rule is stated; any additional explanation that I might give myself can itself be variously interpreted. According to Kripke, the argument comes to a head in PI 201, where Wittgenstein writes, "[t]his was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule" (PI 201). The result is that nothing which might enter our minds in the course of an attempt to understand a rule could have determinate implications for what it is to be in accord or conflict with it. Thus no mental state or process could amount to our understanding the rule in one way rather than another, and "the entire idea of meaning vanishes into thin air".¹¹

¹¹ *ibid.* 22.

Kripke argues that even as Wittgenstein finds this skepticism unavoidable, he also recognizes it to be “insane and intolerable” and attempts to console us with a “skeptical” solution to the paradox posed in PI 201.¹² He suggests that Wittgenstein copes with the dissolution of genuine meaning and understanding not by showing the skepticism to be uncalled-for, but rather by offering what he calls a “skeptical” solution to the paradox. Unlike a “straight” solution, which attempts to expose the skepticism as in the end unwarranted, a “skeptical” solution “begins on the contrary by conceding the skeptic’s negative assertions are unanswerable”, but that shows that “nevertheless, our ordinary practice or belief is justified because – contrary appearances notwithstanding – it need not require the justification the skeptic has shown to be untenable.”¹³ In this way, Kripke says, Wittgenstein’s rejection of “private language” is comparable to Hume’s rejection of “private causation” and Berkeley’s rejection of the “metaphysical myth of matter”. In the *Dialogues*, for example, Berkeley examines the way in which philosophers tend to employ the word “matter” and finds that this word, taken as such, cannot meet his empiricist criteria of meaning. In doing so, his intention is not to dispel the word from our ordinary vocabulary, but rather to strip it of its metaphysical baggage and return it to us in a less confused and misleading form. In the same way, Kripke takes Wittgenstein’s paradox not as an attempt to banish the words “meaning” and “understanding” from everyday usage, but to cleanse them of their metaphysical confusion so that they might cease to trouble us philosophically.

¹² *ibid.* 60.

¹³ *ibid.* 66.

According to Kripke, Wittgenstein's solution is to offer an account of meaning based on communal agreement in behaviour. As we have seen, Kripke's Wittgenstein rejects the idea that there is a genuine fact of the matter concerning which rule a student is following – say, "plus" as opposed to "quus" – nor any real question whether the student has genuinely understood which rule his teacher meant. The only thing that might function as a substitute allowing us to preserve the concepts of meaning and understanding is the brute fact of the instructor's approval or disapproval of the student's performance; this is just what it is for the student to have understood the meaning of the formula for the rule. The instructor can fulfill this role because she is a representative of the community; she has been found to be in agreement with it in these and similar instances.

Kripke calls this meaning in terms of "assertability conditions" rather than in terms of "truth conditions". By this, he is referring to the roughly circumscribable conditions under which a certain utterance is considered correct. For a student to be acting in accordance with the rule "plus", rather than "quus" or any other, for example, the relevant considerations are just her inclination to act in a certain way and the instructor's approval or disapproval. More generally, an utterance is meaningful not in virtue of its standing in a certain relation to the world, as on a truth-conditional account of meaning, but rather in virtue of it being endorsed by the majority of those who make up the rule-following community as a correct use of the words that make up the utterance. The result, of course, is that there can be no private language – any meaningful utterance

must carry with it at least the possibility that it might be subject to communal approval or disapproval.

Kripke's attribution of skepticism to Wittgenstein has been met with criticism from many commentators. This is likely due in large part to a simple lack of attention to the text. Perhaps the most curious thing about Kripke's reading is his failure to notice that in the latter half of PI 201, Wittgenstein clearly goes on to withdraw the skeptical paradox as vitiated by "a misunderstanding". Here is 201 in full:

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

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

Since we do manage to understand each other (which implies that we use our words meaningfully), and since the rule + interpretation model has been shown to be incoherent, Wittgenstein suggests that the proper response is not to embrace the paradox but to correct the misunderstanding on which it relies. Supporting the paradox is "the thought that there is always a gulf between the statement of a rule – a string of words – and the rule's execution or application", as if the "dead" signs had to be brought to life by some

interpretation before they could be put to use.¹⁴ If we think of following a rule as something like consulting a sign-post – which does not intrinsically stand in a relation of accord or conflict with anything else – we quickly realize that the rule, like the directions on any sign-post, can be variously interpreted, and that every subsequent interpretation now stands in need of a further interpretation. An arrow on a post pointing “right”, for example, does not inherently have a normative impact on our behaviour; what gives it this significance is just our taking it in a certain way. Similarly, if items in the mind lack any normative force of their own, then they are given significance only by our taking or interpreting them in some way. Thus any item entering our minds stands in need of an interpretation, even the interpretation itself; and so on *ad infinitum*.

On John McDowell’s reading, the paradox at PI 201 refers to this threat of regress: we conceived of understanding as something like being in possession of an interpretation that cannot be interpreted, but found that every interpretation produces some item which now stands in need of an interpretation itself. Meaning, then, was construed as the “final interpretation” of the rule – the interpretation which in some peculiar way contains its own predetermined verdicts of right and wrong use – and so our grasp of rules was pictured as something like intuiting a Platonic Form. Wittgenstein’s suggestion in the latter half of 201, however, is not that we give up the very idea that we understand and mean things but rather to abandon the “sign-post” conception of the mind’s contents, according to which mental items are intrinsically “normatively inert”, and instead, as Wittgenstein says, realize that there must be “a way of grasping a rule

¹⁴ D. H Finkelstein, “Wittgenstein on Rules and Platonism”, *The New Wittgenstein*, eds. Alice Crary and Rupert Read (New York: Routledge, 2000) 62, 67.

which is *not* an *interpretation*". Having exposed the dilemma as a false one, the question to which Kripke is concerned to give a positive response no longer seems troubling. MacDowell concludes: "There seemed to be problems about the normative reach of meaning, but since they depended on a thesis we have no reason to accept, they stand revealed as illusory."¹⁵

1. 2. Wittgenstein and "The Argument of the Ordinary"

Whether McDowell's declaration that we can now be at peace with questions about meaning is accurate is, of course, up for debate. My concern here is the question of how Kripke could have missed the thrust of Wittgenstein's remarks on rule-following, especially 201, so completely. As Cora Diamond notes, "[w]hen competent and careful readers of a philosopher come up, quite confidently, with accounts of what he says which are not merely wrong but obviously wrong, and obviously wrong about essentials, it may be important to see how this happened."¹⁶ And indeed, one is tempted to say that it is rather *obvious* that Wittgenstein is less than sincere in proposing the paradox. His so-called solution to it is, in Kripke's words, more "straight" than "skeptical"; we see "that on closer examination, the skepticism proves to be unwarranted",¹⁷ the result of a false dilemma which presented an ultimately incoherent conception of meaning as the only alternative to it. As McDowell writes, "Wittgenstein's point is that this dilemma seems compulsory only on the assumption that understanding is always an interpretation: his

¹⁵ John McDowell, "Meaning and Intentionality in Wittgenstein's Later Philosophy", *Mind, Value and Reality* (Cambridge: Harvard University Press, 1998) 274.

¹⁶ Cora Diamond, *The Realistic Spirit* (Cambridge: MIT Press, 1991) 243.

¹⁷ Kripke 66.

aim is not to shift us from one horn of the dilemma to the other, but to persuade us to reject the dilemma by discarding the assumption on which it depends”.¹⁸ Rather than reading Wittgenstein as endorsing the paradox, we should read him as seeking to jettison the supposition which gives rise to the dilemma: that understanding consists in interpretation.

No doubt the subtlety of Wittgenstein’s rhetorical style has some role to play in Kripke’s misreading of the rule-following remarks. In the case of PI 201, it may be particularly easy to mistake Wittgenstein’s interlocutor’s voice for Wittgenstein’s own, if only because Wittgenstein does not guide the reader with quotation marks as he often does. As Stanley Cavell notes, the subtlety of Wittgenstein’s writing can sometimes be taken for deliberate obfuscation, provoking and irritation and impatience on the part of the reader:

Why does he write that way? Why doesn’t he just say what he means, and draw instead of insinuate conclusions?¹⁹

This feature of the text cannot be completely to blame, however. At several points, Kripke acknowledges and willfully dismisses Wittgenstein’s repeated requests that he not be interpreted as putting forth a theory of his own, and speaks of a Wittgenstein who might perhaps “cagily” reject the metaphysical views attributed to him.²⁰ For example, anticipating the accusation that in attributing skepticism about meaning to Wittgenstein he has ascribed to him a philosophical thesis of the kind he wishes to avoid, Kripke declares, “[n]evertheless I choose to be so bold as to say: Wittgenstein holds, with the

¹⁸ McDowell, (1984: 338).

¹⁹ Stanley Cavell, “The Availability of Wittgenstein’s Later Philosophy”, *Must We Mean What We Say? A Book of Essays* (Cambridge: Cambridge University Press, 2002) 70.

²⁰ Kripke 273.

skeptic, that there is no fact as to whether I mean plus or quus.”²¹ Statements like these show that Kripke is certainly aware that Wittgenstein did not consider his own philosophical project to be one of affirming or conclusively rebutting theses; yet he is twice guilty of failing to honour Wittgenstein’s methodological dicta. First, he attributes to Wittgenstein a form of skepticism which involves the denial that there are any facts of meaning or understanding, and second, he takes Wittgenstein to be offering a skeptical solution to a paradox that he never affirms to begin with.

Cavell suggests that one reason for Kripke’s misinterpretation of the remarks on rule-following might be his assumption that the style of the *Investigations* is incidental or secondary to the content of the writing; “not just that it fails sometimes, which is inevitable, but simply that it *cannot* matter *that* much, as much, say, as its fervor continually seems to declare”.²² Indeed, given Kripke’s refusal to respect Wittgenstein’s wishes concerning how to view his work, it is hardly surprising that his interpretation strays from the text. Not only does he dismiss Wittgenstein’s own instructions for how to read the book, in the very passage (PI 192) of the *Investigations* where he takes Wittgenstein to be *denying* the existence of the superlative fact that would constitute meaning, Wittgenstein in fact says, “You have no model of this superlative fact, but you are seduced into using a super-expression” (PI 192). Kripke does not see that Wittgenstein is trying to prod the reader into abandoning the framework which made questions about such a fact tempting in the first place.

²¹ *ibid* 70.

²² Stanley Cavell, “The Argument of the Ordinary”, *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism* (Chicago: University of Chicago Press, 1990) 79.

Cavell suggests in response that we ought to see Wittgenstein's style of writing as indicative of what he hoped to achieve as a philosopher, and therefore crucial to his message. This thought gains support from the preface to the *Investigations*, where Wittgenstein suggests that, after an initial struggle with the presentation, he found it rather fitting:

It was my intention at first to bring all this together in a book whose form I pictured differently at different times. But the essential thing was that the thoughts should proceed from one subject to another in a natural order and without breaks.

After several unsuccessful attempts to weld my results together into such a whole, I realized that I should never succeed. The best that I could write would never be more than philosophical remarks; my thoughts were soon crippled if I tried to force them on in any single direction against their natural inclination. – And this was, of course, connected with the very nature of the investigation.

According to Cavell, one of the book's crucial stylistic and argumentative devices is "the argument of the ordinary", a term he uses to highlight the way in which an apparent stalemate between two opposing views in the *Investigations* is met finally with a plea from a place beyond the dispute for each to abandon the assumption that they share:

These voices, or sides, in the argument of the ordinary, do not exhaust the space of the *Investigations*, or the tasks of its prose. There is the space not party to the struggle of the sides (I do not think of it as a further voice) often containing its most rhetorical or apparently literary passages – as for example, about the icy region of the sublime, or the keyboard of the imagination, or turning our investigation around as around a still point, or repairing a spider's web – that are gestures of assessment from beyond this struggle.²³

²³ *ibid.* 83. In the same essay, Cavell writes, "[i]n taking rules as fundamental to Wittgenstein's development of skepticism about meaning, Kripke subordinates the role of criteria in the *Investigations*, hence appears from my side of things to underrate drastically, or to beg the question of, the issue of the ordinary, a structure of which is the structure of our criteria and their grammatical relations" (p. 66). This remark suggests that Cavell understands our rule-following activities to be somehow more formal than those of employing criteria when we use a word, and thus more distant from the ordinary. I see no reason to view the remarks of rule-following as in any way removed from Wittgenstein's concern with the

The exchange in the rule-following remarks can be seen as an example of this type of argument. In this case, Wittgenstein responds to the alternating Platonism and skepticism of his interlocutor by urging them to reconsider their commitment to the assumption which made the dilemma seem unavoidable in the first place: the “sign-post” model of understanding. His intention is to remind us of the facts of our following rules in everyday life before engaging in philosophical reflection. Kripke’s misreading of the rule-following remarks can in large part be attributed to his having overlooked this feature of Wittgenstein’s method; in failing to acknowledge what Cavell calls “the space not party to the struggle of the sides”, Kripke interprets the voices that seem to affirm skepticism and propose its “solution” as speaking to one another from within the same conflict. If Cavell is correct in suggesting that “[w]hy Wittgenstein creates remarks to call out interpretations which lock in this way, is a question of style, and hence of his philosophical motivation”,²⁴ then for Kripke to have missed this is for him to have misunderstood Wittgenstein’s aim in writing the *Investigations*.

In the rule-following remarks, Wittgenstein’s interlocutor proposes a view of understanding as something like being in possession of an interpretation that is not itself subject to further interpretation. We supposed that the interpretation was required to convert the general formula into a specific order to be followed. It is this assumption that Wittgenstein wants to call into question in the remarks on rule-following, by drawing our

ordinary; indeed, many of his examples are of everyday things like street signs, suggesting that Wittgenstein meant these remarks to apply to a wide range of activities and not merely our more formal ones.

²⁴ Stanley Cavell, *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy* (Oxford: Clarendon Press, 1979), 96.

attention to the discrepancy between how we tend to regard these features of our lives when doing traditional philosophy, and how we relate to them in our everyday lives with words. The relevant difference between how rules appear to us when we step back from our actual rule-following practices and inquire what, in general, following a rule might consist in, and how rules are present for us when we are engaged in following them, is that in the former case we might say that there is always a question whether someone understands, while in the latter, everyday case, there is usually no question at all.²⁵ Of course, there are occasions when we find the meaning of a rule unclear and may need an interpretation, but successful interpretations are nothing more than informative restatements of the original rule, a “substitution of one expression of the rule for another” (PI 201); they do not have any intrinsic power to guide our actions that was lacking in the first statement. As with the rule itself, the power of an interpretation to direct our behaviour does not reside in the interpretative formula per se, but in the particular circumstances in which the formula is learned and employed.

Kripke’s reading (or misreading) of the rule-following remarks therefore indicates an unwillingness – not rare or unique to Kripke – to consider the possibility that philosophical significance might be found in the everyday, in occupying a perspective from which the problem simply does not show up. His dismissal of Wittgenstein’s own instructions for how to read the book, in particular, suggests that he is unable to see philosophy as anything but a particular sort of activity in which philosophical theses are put forth and then proved or rebutted. If this is true, then it is unsurprising that Kripke

²⁵ Warren Goldfarb, “Kripke on Wittgenstein on Rules”, *Journal of Philosophy* 82 (1985): 485.

has overlooked what Cavell calls “the argument of the ordinary”, as well as its importance for Wittgenstein’s intention to have his own voice come from beyond the dispute. His inability to see in the rule-following remarks an “inside” – the voice of the interlocutor that alternates between Platonism and skepticism – and an “outside” – Wittgenstein’s own voice – to the argument has the effect of conflating Wittgenstein’s motivations with those of his interlocutors, leaving Kripke with the impression that Platonism and skepticism stand as the only alternatives to one another. Indeed, since what leads us so surely into skepticism is just Kripke’s unwillingness to admit into evidence those facts which would make it obvious that we commonly do understand the meaning of rules, Cavell claims that the conception of philosophy that frames Kripke’s investigation already *is* skepticism: “[W]e do seek, and therein comes the skeptical conclusion, or solution: the demand for a philosophical solution is the skepticism”.²⁶ The question now becomes: what is the alternative to this picture? And what does this entail for what Wittgenstein wants to say about meaning? It is to these questions that I now turn.

1. 3. *The “Intellectualist” Picture in Philosophy*

In order to answer these questions, it will help to get clear on picture on which the paradox arises. McDowell characterizes it as follows:

A succession of judgments or utterances, to be intelligible as applications of a single concept to different objects, must belong to a practice of going on doing the same thing. We tend to be tempted by a picture of what that amounts to, on the following lines. What counts as

²⁶ Cavell, (1990: 97).

doing the same thing, within the practice in question, is fixed by its rules. The rules mark out rails along which correct activity within the practice must run. These rails are there anyway, independently of the responses and reactions a propensity to which one acquires when one learns the practice itself; or, to put the idea less metaphorically, it is in principle discernible, from a standpoint independent of the responses that characterize a participant in the practice, that a series of correct moves in the practice is really a case of going on doing the same thing. Acquiring mastery of the practice is pictured as something like engaging mental wheels with these objectively existing rails.²⁷

Both Platonism and skepticism share an eagerness to view rules in this way, as containing within themselves their own application, needing, as it were, no help from us. Indeed, in our ordinary rule-following activities, it is easy to forget that there is never only one application that can be imagined in a given situation. Ordinarily when we are following a rule, the kind of skeptical considerations that concern Wittgenstein's interlocutor never occur to us. To one who has learned algebra, for example, the formula for addition, ' $x + y = z$ ', seems to leave no question about how the calculation is to be carried out for any set of values; it seems to take care of its own application, so to speak. Of course, as we have seen, it does no such thing – the rule just stands there, "like a sign-post" (PI 85) – no formula applies itself, leaves no question about how to go on. But abandoning the idea that the rule determines its own application seems to spoil the notion that our actions might be normatively constrained by a rule: "If one is wedded to the picture of rules as rails, one will be inclined to think that to reject it is to suggest that, say, in mathematics, anything goes: that we are free to make it up as we go along."²⁸ If a rule does not contain within itself its own application such that the pattern of its future uses is in some super-

²⁷ McDowell, "Non-Cognitivism and Rule-Following", *Wittgenstein: To Follow a Rule* (Boston: Routledge and Kegan Paul, 1981) 145-146.

²⁸ *ibid.* 150-51.

empirical sense determined in advance, only individual will could determine normative standards, and this amounts to an absence of normativity.

On the one hand, then, the remarks on rule-following engage our ability to imagine a situation in which a student goes wildly off course in, say, writing out a numerical series, making it seem as though Wittgenstein intends to raise skeptical concerns about meaning and understanding. On the other, however, he attempts to call attention to a different feature of our experience, that is, the feeling of naturalness and ease that pervades many of our rule-following activities. This point is related to a tension that Kripke claims to find in the remarks on rule-following. According to Kripke, this tension arises as a result of the way in which Wittgenstein first develops the paradox and then proposes the skeptical solution to it. Recall that we arrived at skepticism due to our inability to settle on the meaning of a rule – where the meaning was supposed to be the “final interpretation” of the rule – since a further interpretation is always conceivable. Wittgenstein’s alternative account of meaning, developed in his solution to the paradox, however, is supposed to rest on communal agreement, which in turn relies on the brute fact that given a certain training, most members of a community will simply find it inevitable to react in a certain way when then told to act in accordance with a given rule. It is this inevitability of response that gives rise to the tension that Kripke finds in Wittgenstein’s argument, for it was just the fact that we were able to conceive of all kinds of different responses in interpreting a given rule that gave rise to the skeptical paradox about meaning and understanding in the first place. On the one hand, Kripke claims that Wittgenstein sees “no a priori reason” why an intelligent creature could not follow the

rule for, say, addition in a deviant way. On the other hand, however, Kripke argues that Wittgenstein's skeptical solution undermines the intelligibility of this prospect:

If such a possibility [that a capable person might grasp the plus function in a deviant way] were really completely intelligible to us, would we find it so inevitable to apply the plus function as we do? Yet this inevitability is an essential part of Wittgenstein's own solution to his problem.²⁹

Kripke's argument melds the idea that our ordinary rule-following practices exhibit conceptual features – like normativity – that must be preserved with the observation that those practices cannot themselves justify these ascriptions, since they do not foreclose on the possibility of alternative interpretations of the expression of rule. Here we see the source of the tension; as Goldfarb writes, “these two ideas pull in opposite directions: it is no wonder that nothing can meet the demands imposed by the two of them”.³⁰ Kripke can only find this tension here, though, if he sees Wittgenstein as coming from within the conflict between Platonism and skepticism to propose a positive thesis of his own. And as we have seen, this is not Wittgenstein's intention.

Rather seeing him as “party to the struggle of ideas”, we might see Wittgenstein as trying to negotiate a path between Platonism and skepticism, to find a position that is able to accommodate both the impersonal authority of the rule and its element of personal judgment. This is not to say that we should see him as proposing a theory of his own, but rather as investigating our actual rule-following activities to see how these two features of our experience might co-exist. As he observes, the statement of a rule “sometimes leaves room for doubt and sometimes not. And this is no longer a philosophical

²⁹ Kripke 98, ft. 78.

³⁰ Warren Goldfarb, “Kripke on Wittgenstein on Rules”, *Journal of Philosophy* 82 (1985): 479.

proposition, but an empirical one” (PI 85). The tension that Kripke finds in Wittgenstein’s writing, we saw, is the result of his inability to reconcile these two ideas; one suggests that we should be altogether skeptical about the notion of normativity, while the other leads us to believe that our experience of being normatively constrained by a rule is genuine. Both, however, share the view that a true understanding of a rule consists in something like “engaging mental wheels with ... objectively existing rails”. When our actual practice fails to live up to this requirement, we are left to question the very possibility of normativity. Perhaps if we could see what is amiss with this picture of rule-following, then, the skeptical considerations that trouble Wittgenstein’s interlocutor will cease to seem so daunting.

The following passage by McDowell gives us a hint as to what the problem might be:

We can find this picture of genuine truth compelling only if we either forget that truth-bearers are such only because they are meaningful, or suppose that meanings take care of themselves, needing, as it were, no help from us. This latter supposition is the one that is relevant to our concerns. If we make it, we can let the judging subject, in our picture of true judgment, shrink to a locus of pure thought, while the fact that judging is a human activity fades into insignificance.³¹

In this passage, McDowell touches upon what is perhaps one of the main features of what Charles Taylor calls the “intellectualist” picture in philosophy: the demand for a way to account for the rationality manifest in our various activities which is completely distinct from those activities. As Cavell writes, “I must empty out *my* contribution to words, so that language itself, as if beyond me, exclusively takes over the responsibility for

³¹ McDowell (1984: 351-52).

meaning”.³² Warren Goldfarb identifies this requirement as one of Wittgenstein’s principle targets in the remarks on rule-following:

The demand, however, is for a fixing of the correct continuation that does not rely upon us or take for granted anything about us, at all. What Wittgenstein principally wants to suggest is that we do not have any real conception of what this comes to. We have, as Wittgenstein is wont to say, ‘no model of it’.³³

Charles Taylor, too, emphasizes the inextricability of the rule from the practice:

The reciprocity is what the intellectualist theory leaves out. In fact, what it shows is that the ‘rule’ lies essentially *in* the practice. The rule is what is animating the practice at any given time, and not some formulation behind it, inscribed in our thoughts or our brains or our genes, or whatever. That’s why the rule is, at any given time, what the practice has made it.³⁴

Finally, Wittgenstein confirms Taylor’s suggestion himself in *On Certainty*:

Not only rules, but also examples are needed for establishing a practice. Our rules leave loop-holes open, and the practice has to speak for itself (OC 139).

Under the spell of the intellectualist picture, we imagined that understanding a rule consisted of being in possession of its final, decisive interpretation; this was supposed to satisfy the requirement that our understanding of a rule transcend its particular instantiations, since it presented an account of the understanding of the rule that was conceptually independent from the pattern of actual responses that we accept as being in accordance with it. Anything short of this ideal is seen as artificial, arbitrary, a reluctant concession for the sake of expediency; it will not be in accord with rationality per se. As Taylor notes, the intellectualist picture offers us no other alternative: “[I]ntellectualism

³² Cavell, “Declining Decline”, *The Cavell Reader* (Oxford: Blackwell Publishers Ltd., 1996) 339.

³³ Goldfarb 487.

³⁴ Charles Taylor, “To Follow a Rule”, *Philosophical Arguments* (Cambridge: Harvard University Press, 1995) 178.

leaves us only with the choice between an understanding that consists of representations and no understanding at all”.³⁵

Taylor goes on to argue that freeing ourselves from the intellectualist impasse involves recognizing a third option: “Embodied understanding provides us with the third alternative we need to make sense of ourselves”.³⁶ It is by overlooking what Taylor here calls “embodied understanding” that we come to imagine that a satisfying account of rationality must bear no trace of us. I believe that Wittgenstein, too, had something like this in mind in attempting to expose the rule-following dilemma as a false one. The intellectualist view demands that we search for the essence of normativity, which is supposed to be buried beneath the miscellany of our everyday rule-following activities. But if, as Wittgenstein suggests, we rotate the axis of our investigation from vertical to horizontal – from what we supposed to be hidden to what lies open to view – we see that the statement of the rule is bound up with the pattern of its correct applications only within a certain context: that of a regular use of the expression of the rule, a custom.

Let me ask this: what has the expression of a rule – say, a sign-post – got to do with my actions? What sort of connexion is there here? – Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it.

But that is only to give a causal connexion; to tell how it has come about that we now go by the sign-post; not what this going-by-the-sign-post really consists in. On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom (PI 198).

The training that we receive when we are initiated into a practice of reacting to expressions of a rule in a certain ways and not in others is not a mere conventional

³⁵ *ibid.* 173.

³⁶ *ibid.*

overlay on the phenomenon of normativity, a causal profile which presupposes an already-present connection between the expression of the rule and its application, but rather the very means by which this connection is established and sustained. Only in such a context can there be “a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call ‘obeying the rule’ and ‘going against it’ in actual cases” (PI 201).

For one locked into the intellectualist picture, acknowledging the reciprocity between the rule and the practice will seem a disappointing compromise, for it means that the connection between the two is not an objectively existing, super-empirical fact but rather a hard-won convention forged through training. It was in digging for such a fact, however, that we arrived at a level where it seemed as though all sorts of problems for meaning and understanding arose. Wittgenstein’s point is that we do not need to dig at all – that at the level of our actual rule-following activities we have a “fully satisfying” perspective in which we are not required to surrender our use of the term “correct” in all cases simply because we can imagine situations in which it might be called into doubt.³⁷

As McDowell writes,

[m]y right to claim to understand [another person] is precarious, in that nothing but a tissue of contingencies stands in the way of my losing it. But to envisage its loss is not necessarily to envisage its turning out that I never had the right at all.³⁸

The fact that a practice is *ours* does not mean that we are thereby unable to distinguish actually obeying a rule from merely thinking that one is obeying it. Though we could

³⁷ McDowell (1981: 150-51).

³⁸ McDowell (1984: 349).

not, for example, talk of the rule for addition as determining its own applications for a group of people who respond wildly or with confusion to the command “Add 2” – and this itself would give us reason to conclude that the order does not mean the same thing to them as it does for us – this does not entail that our own feeling of being normatively constrained by a rule is illusory. It simply means that our normative practices are what they are only against a background of general facts about us.

Wittgenstein therefore resituates normativity within the bounds of time and place; to become aware of ourselves as embodying practices is to become aware of ourselves as embedded in a particular historical and cultural setting. The “tissue of contingencies” of which McDowell writes involves a history of training and socialization into established practices which we receive as a part of our growing up in a particular social setting. Thus, according to Taylor,

[m]y embodied understanding doesn't only exist in me as an individual agent, but also as the coagent of common actions. This is the sense we can give to Wittgenstein's claim that obeying a rule is a practice. He means by this a social practice.³⁹

Because this understanding lies largely unformulated in the background of our activities, it is easy to overlook in favour of more explicit theoretical reasoning. Of course, we often are able to step back from what we are doing and formulate explicit theories about some limited range of phenomena that we wish to explain, but this type of theoretical activity makes sense only against a background of all sorts of facts and assumptions concerning the practice in question that are not made part of an explicit theory. This

³⁹ Taylor, (1995: 178).

involves a sense of how to go on in certain instances, of which facts are relevant, of how others are supposed to respond, and so on.

One of Wittgenstein's aims in the rule-following remarks (and beyond), I take it, is to stress just how much implicit understanding is at work behind the scenes in our rule-following activities. The intellectualist view in philosophy takes our ability to formulate explicit theories in particular circumstances and distorts it into a general requirement for rationality in all of our activities. What the view requires, then, is that the background understanding that normally is presupposed when we, say, write out an addition table be made explicit so that our theoretical understanding of what counts as going on in the same way be able to encompass every possible case, regardless of circumstance. But since we "have no model" for what this comes to, it can seem that we are burdened with the skeptical considerations that Kripke reads Wittgenstein as putting forth as unavoidable. In speaking from beyond the conflict between Platonism and skepticism, however, Wittgenstein's aim is to show that neither position is inevitable, that the picture itself is incoherent and should be abandoned as such. Taylor neatly sums up Wittgenstein's approach:

[Wittgenstein] undermines the picture by bringing out the background we need for the operations described in the picture to make sense, whereby it becomes clear that this background can't fit within the limits that the disengaged view prescribes. Once understood against its background, the account shows itself to be untenable.⁴⁰

Earlier we envisioned rules as something like "objectively existing rails" which race ahead of us to encompass every possible circumstance. When no candidate could meet

⁴⁰ Charles Taylor, "Lichtung or Lebensform: Parallels Between Heidegger and Wittgenstein" *Philosophical Arguments* (London: Harvard University Press, 1997) 71.

this requirement, we were thrown into skepticism. Now, however, we see that this is not a genuine failure, because the requirement was an illusory one. If we “have no model” for meeting the requirements that the intellectualist view imposes upon us, then the proper response is to retire it, to change our way of looking at things so that it no longer appears compulsory. According to McDowell, “[t]he right response to ‘How is meaning possible?’ or ‘How is intentionality possible?’ is to uncover the way of thinking that makes it seem difficult to accommodate meaning and intentionality in our picture of how things are and to lay bare how uncompulsory it is to think that way.”⁴¹ In coming to see this picture as just one way to see things, we can come to see it for what it is: particular conceptualization of the area of investigation which makes certain demands on us – about which questions are important and which not, about what would constitute a good answer to those questions, about what we are like as human beings, and so on. By encouraging us to let go of this picture Wittgenstein is not suggesting that we stop questioning, but that we might be led to ask different questions; for example, about the sources this picture has in our lives, about what commitments on our part make it seem so inevitable.

Cavell suggests that Wittgenstein had little to say in response to these questions, and regarded skepticism as an inescapable feature of our life with language. He argues that in order to understand the later Wittgenstein as a philosopher of culture, we must see the *Investigations* as concerned with the inheritability of philosophy, in general, and with maintaining our relationship to the possibility of philosophical skepticism in particular.

⁴¹ McDowell “Meaning and Intentionality in Wittgenstein’s Later Philosophy”, *The Wittgenstein Legacy*, eds. Peter A. French, Theodore E. Uehling Jr. and Howard K. Wettstein (Notre Dame: University of Notre Dame Press, 1992) 47.

In the final section of this chapter, I suggest that, although Wittgenstein does not concern himself much in the *Investigations* with questions about the source of the intellectualist picture and the skepticism that so often accompanies it, he does touch on it elsewhere. Thus it will also serve to offer further textual support for my claim that this picture should be seen as the proper target of the remarks on rule-following.

1. 4. *A Craving for Generality*

According to Cavell, the return to the ordinary in the *Investigations* amounts to our coming to see that what we need in philosophy is not the proffering of ever more elaborate theories, but instead the recognition that the possibility of philosophical skepticism constitutes a basic feature of our human condition, in particular, of our life with language. Here I part ways with Cavell, for I have my doubts that Wittgenstein would want to tie the fate of philosophy to the possibility skepticism, and therefore to the intellectualist view in philosophy. Similarly, I am inclined to doubt that Wittgenstein held philosophical skepticism to be a valuable and inescapable part of our humanity. Indeed, Wittgenstein continually warns his reader to beware of such philosophical “musts”; in a well-known remark at PI 66, he urges us: Don’t say “must,” but “*look and see*”. While this warning occurs in a discussion of language, other remarks suggest that he felt it appropriate in other areas of philosophy as well. Drawing on Wittgenstein’s criticism of Frazer’s *The Golden Bough*, as well as G. H. von Wright’s article, “Wittgenstein in Relation to His Times,” I suggest that Wittgenstein was just as wary of attempts to characterize human beings as *necessarily* some way or another.

Like McDowell, Cavell takes Wittgenstein to be claiming that the attempt to justify our practices from an external standpoint is misguided, since such detachment involves losing sight on those conceptual and perceptual capacities in virtue of which we understand that practice as engaged participants. Unlike McDowell, however, Cavell rejects the suggestion that philosophical clear-sightedness might free us from the fear of groundlessness which motivates the search for external justification. The divergence here is due to what each takes to be the source of philosophical disengagement. For McDowell, the temptation to alienate oneself from familiar practices is contingent upon certain events in relatively recent intellectual history. In *Mind and World*, he suggests that the philosophical anxieties he discusses require the background of a conception of nature as “disenchanted”, a conception which “was made available only by a hard-won achievement of human thought at a specific time, the time of the rise of modern science.”⁴² Similarly, Taylor connects dominance of the intellectualist view in both philosophy and our broader cultural habits to “the hegemony of institutions and practices that require and entrench a disengaged stance: science, technology, rationalized forms of production, bureaucratic administration, a civilization committed to growth, and the like, the kind of thinking of which both are variants has a certain counter-cultural significance, and inherent thrust against the hegemonic forms of our time.”⁴³ Cavell, on the other hand, tends to locate the source of alienation in those features of our lives which are less contingent, more indelibly human, for example, in our engagement with language: “This

⁴² John McDowell, *Mind and World* (Cambridge: Harvard University Press, 1996) 70.

⁴³ Charles Taylor, “Lichtung or Lebensform: Parallels Between Heidegger and Wittgenstein”, *Philosophical Arguments* (Cambridge: Harvard University Press, 1995) 75-76.

means that I think the griefs to which language repeatedly comes in the *Investigations* should be seen as normal to it, as natural to human natural language as skepticism is.”⁴⁴

As a result, Cavell’s view seems to be that to hope for protection against the feeling of groundlessness is fundamentally misguided, something akin to denying our own finitude: “What challenges one’s humanity in philosophy, tempts one to excessive despair or to false hope, is named skepticism. It is the scene of a struggle of philosophy with itself, for itself.”⁴⁵ This natural, human inclination towards skepticism connects to what Cavell takes to be the theme of the *Investigations* – Wittgenstein’s cultural concern that philosophy survive: “[F]or I take its pervasive theme of the inheritance of language, the question, the anxiety, whether one will convey sufficient instruction in order that the other can go on (alone), as an allegory of the inheritability of philosophy – which is after all what the isolated, all but unnoticed child in Augustine’s description of his past, did also inherit.”⁴⁶ For Cavell, however, this recognition of our human relation to skepticism is tied up with the realization that only from an appreciation of our fundamentally human relation to the possibility of skepticism does the human endeavor of philosophy have a future: “Then why can’t [skepticism] be ignored? For Wittgenstein that would be akin to ignoring philosophy, and surely nothing could be so easily ignored – unless false hope and excessive despair are signs and effects of unobserved philosophy.”⁴⁷

For Cavell, then, Wittgenstein’s attitude toward the type of philosophical problems that arise in the remarks on rule following is that such problems are part of our

⁴⁴ Cavell, (1996: 337).

⁴⁵ *ibid.* 327.

⁴⁶ *ibid.* 347.

⁴⁷ *ibid.*

fate as speakers. Precisely why this is the case, Wittgenstein does not say: “Why intellectual bewitchment takes the form it takes in the *Investigations* we have not said – Wittgenstein speaks of pictures holding us captive, of unsatisfiable cravings, of disabling sublimizings. He does not, I think, say very much about why we are victims of these fortunes, as if his mission is not to explain why we sin but to show us that we do, and its places.”⁴⁸ It is true that Wittgenstein does sometimes write as if it is language itself that seduces us into philosophical confusion. In PI 109, he writes that “[p]hilosophy is a battle against the bewitchment of our intelligence by means of language”; and elsewhere that “[w]e remain unconscious of the prodigious diversity of all the every-day language-games because the clothing of our language makes everything alike” (PI p. 224). These remarks, and others, make it sound as if Wittgenstein locates the source of philosophical bewilderment in the language itself. However, it is also true that Wittgenstein often notes our reluctance in philosophy to delve beneath the coarse weave of surface grammar to examine the particular uses to which we put our words. In “Wittgenstein in Relation to His Times”, G. H. von Wright suggests that Wittgenstein saw this reluctance as symptomatic of the intellectual and cultural climate of the time. He writes:

Wittgenstein thus thought that the problems with which he was struggling were somehow connected with the ‘*Lebensweise*’ or the way people live, that is, with features of the culture or civilization to which he and his pupils belonged.⁴⁹

Von Wright writes that Wittgenstein found the spirit of the European and American civilization both alien and distasteful, and “deeply distrusted” its organizing idea of

⁴⁸ *ibid.* 338.

⁴⁹ G. H. von Wright, “Wittgenstein in Relation to His Times”, *Wittgenstein and His Times*, ed. B. McGuinness (Oxford: Basil Blackwell, 1982) 118.

progress via the technological harnessing of science, with its tendency toward self-destruction and its capacity to cause “infinite misery.”⁵⁰ This aspect of Wittgenstein’s thought, he suggests, mediates a link between “the view that the individual’s beliefs, judgments, and thoughts are entrenched in unquestioningly accepted language-games” and “the view that philosophical problems are disquietudes of the mind caused by some malfunctioning in the language-games and hence in the life of the community.”⁵¹

Contrary to Cavell’s claim that Wittgenstein does not “say very much about why we are victims of these fortunes,” we can find support for von Wright’s claims in works other than the *Investigations*.⁵² In the *Blue Book*, Wittgenstein writes of our “craving for generality” which has as its main source “our preoccupation with the method of science”:

I mean the method of reducing the explanation of natural phenomena to the smallest possible number of primitive natural laws.... Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer questions in the way science does. This tendency is the real source of metaphysics, and leads the philosopher into complete darkness (BB p. 18).

And in *Culture and Value*, he writes:

Science: enrichment and impoverishment. *One* particular method elbows all others aside. They all seem paltry by comparison, preliminary stages at best. You must go right down to the original sources so as to see them all side by side, both the neglected and the preferred (CV p. 60-61).

Remarks like these suggest that Wittgenstein believed it is a craving for generality, and not merely a problem with language, which makes us content to remain on the surface of

⁵⁰ *ibid.*

⁵¹ *ibid.*

⁵² Cavell’s reading is unique in that, unlike other authors who write on Wittgenstein as a philosopher of culture, it remains entirely within one text. The *Investigations*, he suggests, should be read as a work whose cultural teaching is internal to its structure.

our grammar. This craving has as its source our modern fascination with the scientific method, a method that explains things by subsuming diverse phenomena under general causal laws. As David Pears writes,

The scientific construal of the world ... forced people to raise questions about their place in it as observers and to wonder how they really knew about it. At this point it was natural to assume that these questions too would be answered by theories of greater generality – metaphysics after physics. If a certain technique has enabled you to climb halfway up a mountain, you naturally expect the same technique to take you to the top.⁵³

The astonishing success and explanatory power of this method for the purposes of explaining certain phenomena seduces us into thinking that everything might be explained this way, by stepping back from one's subject matter and elaborating theories, and we are bewildered when a problem is not susceptible to an answer in causal scientific terms.

These comments make it difficult to maintain Cavell's view that Wittgenstein thought that the skepticism addressed in the *Investigations* is natural to human language. I have suggested that Wittgenstein believed that it is possible that we might recognize the causal point of view *as* a point of view – as just one way of looking at things. This seems to be the message of Wittgenstein's criticism of Sir James Frazer's anthropological study of magic in *The Golden Bough*.⁵⁴ Wittgenstein objects to Frazer's characterization of the primitive rituals of his subjects as proto-scientific theories, such that we can take for granted that the performance of these rituals is, as it were, smoothly continuous in its

⁵³ David Pears, "Literalism and Imagination: Wittgenstein's Deconstruction of Traditional Philosophy", *International Journal of Philosophical Studies* 10 (2002): 7.

⁵⁴ Sir James George Frazer, *The Golden Bough: A Study in Magic and Religion* (New York: The Macmillan Company, 1935).

evolution to the goals of modern scientific explanation. By viewing the practices of his subjects in this way, Frazer is able to trace a progressive development from them to our modern scientific worldview. According to Wittgenstein, this is nothing more than a modern prejudice, a refusal to actually examine the facts in an attempt to understand how these practices showed up for these people, how they made sense to them from within a broader system of reference that differs from our own causal scientific way of looking at the world.

The historical explanation, the explanation as an hypothesis of development, is only *one* way of assembling data – of their synopsis. It is just as possible to see the data in their relation to one another as to embrace them in a general picture without putting it in the form of a hypothesis about temporal development (PO, p. 131).

Frazer's mistake, as Wittgenstein sees it, is in seeking *hypotheses* to explain the foreign practices of other peoples: "[O]ne must only piece together what one knows, without adding anything, and explanation follows of itself" (PO, p. 121). Instead of positing something hypothetical to be dug out from "behind" appearances (like the peoples' "wrong" beliefs), that is, we should seek the *Weltbild* that shapes their world, as it shows up in the grammar of their practices. In failing to recognize that these people lead their lives with a shared background understanding at all, we fail to realize that this understanding is different from our own. In assuming that their concerns are akin to ours, one is liable to construe the differences one finds entirely in terms of knowledge and the adequacy – or, more often, inadequacy – of what one takes to be their primitive scientific theories. Since our scientific knowledge is much greater than theirs, we then tend to suppose that there has been a more or less linear development from them to us, and

suppose that our progress in science might be generalized into progress on the whole. In short, we fail to consider the possibility that such practices may belong to an altogether different way of inhabiting the world. I think it follows from this that Wittgenstein would also have thought it a mistake to suppose that *we* cannot help trying to explain and control everything, that our craving for generality is somehow inherently human.

1. 5. Conclusion

In this chapter, I suggested that Wittgenstein's target in the rule-following remarks is what Charles Taylor calls the "intellectualist" view in philosophy, and the kind of scientific metaphysics that often accompanies the scientific worldview. This is not to say, then, that Wittgenstein found "the method of science" inherently unattractive, but only that he thought it misleading and inappropriate in philosophy, when we are apt to turn it on ourselves and neglect the fact that to the extent that disengagement is possible at all, it is only so against a background of shared practices.. There is nothing in scientific explanation itself which leads to the crowding out of other ways of looking at the world; however, the scientific worldview frequently becomes aligned with the intellectualist view in philosophy because we tend to assume that the theoretical, disengaged stance required for scientific theorizing and calculation is not just one way for humans to inhabit the world, but instead gives us a picture of the way human beings fundamentally are anyway: disengaged subjects gazing upon a world of objects. It thus shares with the intellectualist picture a neglect of the shared background practices that are a presupposition for its operation in the first place. The proper response is not to eschew

scientific investigation altogether, but to recognize the scientific picture itself as an abstraction from our richer, everyday understanding of the world and ourselves in it.

Chapter 2: The Private Language Argument

My purpose in this chapter is to examine the significance of the private language argument for Cartesian dualism, and, in particular, the problem of other minds. This problem gets started with the assumption that I know my own mental states directly and others' only by analogy; thus, at best, I can arrive at probable knowledge about the minds of others via "a route of shaky inference."⁵⁵ The private language argument can be read as an attempt to close this evidential gap by showing that the mind bears a *logical*, and not merely contingent or causal, relationship to the body. Although there is little agreement as to the precise form of Wittgenstein's reasoning, the argument is generally thought to call into question the Cartesian assumption that the human form has no significant role to play in our account of our cognitive and non-cognitive capacities and their exercise. It is a typical supposition that the argument does this by showing private language to be impossible due to its failure to meet some particular adequacy condition for anything that is to be called a language. On the account I offer here, however, the private linguist fails not because private language is impossible *per se*, but because the idea of such a language cannot be made fully intelligible. Unlike on Kripke's view, this yields no insight into the social nature of meaning, but rather returns us to our ordinary concepts of meaning and language while revealing a *logical* connection to hold between mental states and behaviour.

⁵⁵ R. C. Buck, "Non-Other Minds", *The Philosophy of Wittgenstein: The Private Language Argument*. ed. John Canfield (Garland Publishing, Inc.: 1986), p. 43.

My discussion will be divided into several sections. In the first, I compare Kripke's notion of privacy with the notion at stake in what is more commonly held to be the private language argument. This section will serve to illuminate what I take to be the real problem with private language: we cannot make sense of the claim that the private linguist, having been stripped of all the criteria which we normally take to be suggestive of language use, is nevertheless speaking a language. Section two expands on this idea by describing the relationship between ascriptions of language use and behaviour. In section three, I suggest that the private language argument can be read as targeting a certain, rather tempting, intuition about language and experience; this is the thought that ineffability is an indelible feature of our inner life. This will set the stage for section four, in which I put forth one argument against the possibility of private language. Though I do not suggest that the "No Stage-Setting" argument that I offer here is the only one to be found in the *Investigations*, I think it is the most powerful, and Wittgenstein thought it important enough to emphasize in the notes for the public lecture published in *Philosophical Occasions*.⁵⁶ The upshot of the argument is that our psychological vocabulary is not learned on the model of "object and designation", according to which I would learn meaning of words like "pain" by acquainting myself with the objects for which they stand. In the final section of this chapter, I will examine the significance of this conclusion for the problem of other minds. In short, if there is a *logical* connection between the inner and the outer, rather than a mere causal one which holds between two

⁵⁶ Ludwig Wittgenstein, "The Language of Sense Data and Private Experience", *Philosophical Occasions 1912-1951*, ed. James Klagge and Alfred Nordmann (Indianapolis: Hackett, 1993) 290-397.

logically-distinct, independently-specifiable entities, then it is false to say that we have only mediated access to one another's mental states.

2. 1. *Language and Behaviour*

The most obvious way in which Kripke's discussion of private language differs from Wittgenstein's own discussion is in what he takes to be the relevant sense of privacy at stake in the argument. According to Kripke, communal standards constituted by agreement in judgments are needed to place normative constraints on what is to count as being in conformity with a given rule. This *ipso facto* rules out the possibility of a private language, for no possible fact about a solitary linguist's intentions, insofar as these facts are construed individualistically, is enough to impose normative constraints on her linguistic behaviour. Thus the notion of privacy with which Kripke takes Wittgenstein to be concerned is a contingent one; the problem is not that the private linguist cannot be understood, but that there happens to be no one around to understand him.

Most philosophers agree that this is not the sense of privacy with which Wittgenstein is concerned in the discussion that is typically taken to constitute the private language argument. In PI 243, he introduces a private language as one whose "words refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language" (PI 243). Such a language would be "necessarily unteachable";⁵⁷ it would not be akin to a secret code

⁵⁷ David Pears, *Ludwig Wittgenstein* (New York: Viking Press, 1970) 158.

which I just happen to keep to myself, but a language which I could not possibly share with another person, because the things that define its vocabulary – my private, inner states – are accessible to me alone. Thus the crucial feature of this sense of privacy concerns the inability of others to understand the meaning of the private symbol: they cannot learn how to apply it. On this view, the possibility of private language signifies the possibility of a rule establishing the meaning of a symbol referring to a *quale* or private mental object that cannot be understood by third parties. Undermining this conception of privacy, we will see, involves showing that no sense can be given to a semantics that is necessarily unintelligible to others. In what follows, I suggest that in the private language argument, Wittgenstein's gradual removal and restoration of those features of behaviour and context which we normally take to be suggestive of normativity can be seen as an attempt to elucidate our criteria for ascribing language use to someone. I will argue that the private linguist fails not because her language fails to meet some *a priori* definition of language, but because the concept itself cannot be made fully intelligible.

Wittgenstein's investigation into our criteria for ascribing rule-governed linguistic behaviour to someone begins at around PI 206, when he asks the following question: "Suppose you came as an explorer into an unknown country with a language quite strange to you. In what circumstances would you say that the people there gave orders, understood them, obeyed them, rebelled against them, and so on?" (PI 206). In other

words: Under what circumstances would we say that they followed rules?⁵⁸ In the remark that follows, he offers a partial response:

Let us imagine that the people in that country carried on the usual human activities and in the course of them employed, apparently, an articulate language. If we watch their behaviour we find it intelligible, it seems ‘logical’. But when we try to learn their language we find it impossible to do so. For there is no regular connection between what they say, the sounds they make, and their actions: but still these sounds are not superfluous, for if we gag one of the people, it has the same consequences as with us: without the sounds their actions fall into confusion – as I feel like putting it. Are we to say that these people have a language: orders, reports, and the rest?

There is not enough regularity for us to call it a “language” (PI 207).

What is interesting about this passage is that while the activity in which the people are engaged meets Kripke’s criteria of being social, Wittgenstein nevertheless denies that it can be considered a language. Kripke, on the other hand, would take it for granted that they are speaking to one another. On his interpretation, it would presumably be possible for an isolated individual to display all of our ordinary criteria for employing a language and still not mean anything in his use of a term. This is because, as we saw in chapter one, for Kripke these criteria themselves stand in need of justification. The fact that one cannot be given was what led him to his “skeptical solution”; when none of the individualistic accounts of meaning that he tries out are satisfying, he takes this to be indicative of the social nature of meaning. In PI 207, however, Wittgenstein offers an example in which the apparent participation of other people is irrelevant. This suggests

⁵⁸ That these questions can be considered one and the same is shown by Wittgenstein’s remark that “[f]ollowing a rule is analogous to obeying an order” (PI 206).

that it is possible for there to be a society which nevertheless falls short of the criteria for employing a language.

If, on the other hand, an isolated individual's verbal behaviour did exhibit purposive, rule-governed activity which proved intelligible to us, then his solitude would also be irrelevant, for there would be something in his behaviour to suggest normativity. The above question can therefore be posed here as well: Under what circumstances are we willing to ascribe rule-governed behaviour to someone? I suggest that Wittgenstein's discussion of private language can be read as addressing this same question. Thus, on the view I will offer here, the argument is an investigation into our criteria for ascribing language use to some person or group; it is not a question of what justifies those criteria.

2. 2. "*There is not enough regularity to call it a 'language'*"

The crucial feature that the language described in PI 207 shares with a necessarily private language is its un-translatability. For, as we saw, there is not enough "regularity" in behaviour – in particular, not enough regularity in the relationship between what the people utter and how they behave – for an outsider to catch on to what they are doing. We cannot discover what they mean by observing how their use agrees with how they explain what they mean; nor can we do so by noting the role that their use of a given symbol plays in their actions. These are just two telltale signs of language use, but removing them from the example has the result of disqualifying what these people are doing from being a language. The effect of the example, then, is to reveal the *logical connections* that those criteria have to the concept of language. What determines whether

someone is following a rule is not some logically independent fact that happens to be accompanied by behaviour; nor is it the behaviour itself. Rather, that behaviour can only be identified as behaviour *expressive of* language use; rule-following activity is manifested *in behaviour*, in regular and public connections between verbal utterances and physical actions. When the private linguist severs these connections, he is not doing away with some contingent behavioural symptom of what is essentially a mental activity, but discarding something essential to the concept itself. It is for this reason that the notion of a necessarily private language cannot be made fully intelligible: we cannot make sense of the claim that, despite not meeting any of our criteria for ascribing language use to someone, one is nevertheless speaking a language.

This interpretation, I think, sits nicely with the little-acknowledged fact that Wittgenstein never explicitly *says* that private meaning is impossible. Many philosophers suggest that, after introducing the idea of a private language at PI 243, Wittgenstein goes on to argue that such a language would be impossible⁵⁹, meaningless⁶⁰, that someone attempting to establish a private language would in fact “have no meaning to communicate even to himself.”⁶¹ The text, however, never goes so far as to state such a conclusion; even the famous diary case of PI 258, which is often taken to be the crux of

⁵⁹ For example, “[t]here can be no such thing as a rule for the use of a word which cannot logically be understood or followed by more than one person” (P. M. S. Hacker, “The Private Language Argument”, *A Companion to Epistemology*, ed. J. Dancy and E. Sosa (Oxford: Blackwell, 1993) 316).

⁶⁰ As Pears puts it, an unteachable language cannot be meaningful, since “meaning is linked...with public criteria” (Pears (1970), 147).

⁶¹ A. J. Ayer, “Can There Be A Private Language?” *The Private Language Argument*, ed. John V. Canfield (New York: Garland Publishing, 1986) 65.

the argument, stops short of explicitly concluding that private meaning is impossible.⁶²

In this passage, we are asked “to imagine the following case”:

I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation. I will remark first of all that a definition of the sign cannot be formulated. – But still I can give myself a kind of ostensive definition. – How? Can I point to the sensation? Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation – and so, as it were, point to it inwardly. – But what is this ceremony for? For that is all it seems to be! A definition surely serves to establish the meaning of a sign. – Well, that is done precisely by the concentrating of my attention; for in this way I impress on myself the connection between the sign and the sensation. – But “I impress it on myself” can only mean: this process brings it about that I remember the connection *right* in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about ‘right’ (PI 258).

Nowhere in this section does Wittgenstein explicitly state that “S” would be altogether bereft of meaning, only that “here we can’t talk about ‘right’”. Indeed, it would seem an awkward departure from his overall deconstructive strategy to make such a sweeping generalization about the line between language and non-language – to plant, amidst his various admonitions against theorizing in philosophy, such a crisp, self-contained positive thesis.⁶³ Recently some philosophers have become more attuned to this fact, as well as to the diverse nature of the remarks themselves, and abandoned the attempt to reconstruct a single, cohesive “private language argument”. Instead, they more modestly suggest, Wittgenstein intended these remarks as a collection of disanalogies with public

⁶² Stanley Cavell seems to be the first to have noticed this (S. Cavell, *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy* (Cambridge: Cambridge University Press, 2003)), and Oswald Hanfling reminds us in his *Wittgenstein and the Human Form of Life* (New York: Routledge, 2002).

⁶³ Indeed, the following methodological reminder shows up in the middle of the argument at PI 255: “The philosopher’s treatment of a question is like the treatment of an illness.”

meaning – a series of challenges to the apparent ease with which one is supposed to be able to establish a private language. Judith Jarvis Thomson⁶⁴ suggested a similar move long ago, and it has been made more recently by John Canfield⁶⁵ and Oswald Hanfling,⁶⁶ amongst others. On this reading, the private language argument is not a cohesive, knock-down refutation of the possibility of private meaning but yet another example of Wittgenstein’s well-worn strategy of returning the language of philosophical problems to their natural setting in our everyday lives.

The second advantage of this type of reading is that, unlike some others, it does not require a strict line to be drawn between language and non-language via, say, some outdated verificationist principle. On the earliest version of the argument, labeled “Old Orthodoxy” by Stewart Candlish, such a principle – something like, “All genuine contents must, if true, be verifiable” – was supposed to mark just such a boundary; a private language, being unverifiable, would fall into the latter category. According to this interpretation, the private diarist in PI 258 might initially succeed in forging a private connection between the symbol “S” and a private sensation, but with without any independent standard of proper use she might, for all she knows, go on to apply it quite wildly; thus, the connection would become tenuous and ultimately collapse.⁶⁷ This type of reading has since been abandoned; not only is the verificationist principle on which it rests notoriously counterintuitive, but it is not even clear that one *could not* verify that her

⁶⁴ Judith Jarvis Thomson, “Private Languages”, *American Philosophical Quarterly* 1 (1964) 20-31.

⁶⁵ John V. Canfield, “Private Language: The Diary Case”, *Australasian Journal of Philosophy* 79 (2001): 377-394.

⁶⁶ Hanfling (2002).

⁶⁷ This interpretation traces back to Norman Malcolm, “Wittgenstein’s *Philosophical Investigations*”, *Knowledge and Certainty*, Ithaca, NY: Cornell University Press, 1963.

application of her private symbol “S” is consistent with her original intention. Perhaps she could do so indirectly, by checking the general reliability of her memory. She might, for example, test herself by studying the contents of a drawer, shutting it, writing down what she believes the items to be, and then reopening it to check whether she has remembered correctly. If her memory does not fail her in this test – or, for that matter, in the various less formal tests of everyday life – then we have no reason to suppose that it should in the private case.⁶⁸

Though most philosophers have dropped the stronger verificationism of the Old Orthodox interpretation, some philosophers still draw a strict line between language and non-language with a more modest version of this principle. This version states, in short, that someone can be said to follow a rule only if it is possible, at least in principle, to verify that he does so; if one cannot, then no rule can be said to exist.⁶⁹ While less obviously implausible than the stronger verificationist principle, this updated principle still runs counter to some of our intuitions.⁷⁰ According to this version of the argument, one need never check that the use of his rule is correct, but it must be at least possible that he do so. But why should it matter if, in principle, one might verify the correct use of his concept, if, in practice, he need never do so? Some independent reason must be given for adopting this claim; it cannot simply be that it allows us to rule out the possibility of a

⁶⁸ This point is made by Pears (1970: 161).

⁶⁹ For example, a rule can be said to exist only if there exists a “usable criterion of successful performance” (David Pears, *The False Prison*, Vol. 2 (Oxford: Blackwell, 1996) 132), “publicly checkable procedures” (Paul Johnston, *Wittgenstein: Rethinking the Inner* (New York: Routledge, 1993) 20), or an “operational standard of correctness” (Hans-Johann Glock, *A Wittgenstein Dictionary* (Oxford: Blackwell, 1996) 312); in short, a rule can exist only if it is possible, at least in principle, to verify that one follows it.

⁷⁰ There are, for example, statements which are unverifiable in principle but which we are still inclined to consider meaningful; e.g. statements about arbitrary times and places.

private language. On the interpretation that I will offer, however, we need not rely on any such *a priori* adequacy condition which would distinguish language from non-language. Rather, by looking at how language actually works, we will see that no sense can be made of a rule that is necessarily unintelligible to others.

2. 3. *Private Objects*

In keeping with the spirit of Wittgenstein’s method as described in chapter one, then, on the view I will offer here there is no one, cohesive private language argument, but rather a series such arguments spread throughout the first part of the *Investigations*. This is related to what is at stake in the discussion of private language; the object of criticism is not any particular developed theory, but rather a set of intuitions about experience and what language would have to be like in order to express our experience perfectly. The problem of private language might therefore be construed as a problem of reference. The question is: What sorts of intentional attitudes can we be said to take toward our private experiences? At the very least, we might insist, we should be able to name them. In this case, intentionality – thought, perception, belief, and so on – might be conceived as a relation between a judging subject and an intentional object – the “idea” in the language of early modern philosophy. One stands in relation to one’s mental states as one does objects in the physical world: “On the one hand the perceiver, on the other the entity, and between the two the relationship *being aware of*.”⁷¹ The difference is that, while my awareness of physical objects comes via a long causal chain leading out into

⁷¹ Canfield (2001): 379.

the world, my awareness of these private, mental objects is direct and indefeasible. If a person confronted with these entities were to name them, she could not possibly share the meanings of these names with another person, because their referents would be accessible to her alone.

Such a language, we might think, could be considered an “ideal” language, since it would be able to avoid many of the perceived defects of public language. These defects typically have to do with our ordinary vocabulary being too coarse or crude to articulate the finer shades of immediate experience. Because public language is geared toward interpersonal communication, it is not designed to capture the subjective aspect of experience; this could only be expressed in a private language which, released from the pragmatic demands of effective communication, would be free to acknowledge the rich detail of our inner lives. The assumption at play here is that “ineffability is a pervasive feature of experience, so that whenever we describe what we feel or perceive, there is always something left over – an ‘inner content of experience’, as it is sometimes put – that eludes description.”⁷² Sensation words – like “pain”, for example – will therefore have both a public and a private component. The public component may be shared by even those who are incapable of experiencing the sensation, if through training one learns our shared concept of what pain-behaviour is, along with the consequences of someone’s being in pain, and comes to use the word in appropriate ways under appropriate circumstances. The subjective component, on the other hand, is established by the nature

⁷² Hanfling (2002: 30). The same intuition is expressed by Wittgenstein in PI 273: What am I to say about the word “red”? – that it means something ‘confronting us all’ and that everyone should really have another word, besides this one, to mean his *own* sensation of red? Or is it like this: the word “red” means something known to everyone; and in addition, for each person, it means something known only to him? (Or perhaps rather: it *refers* to something known only to him.)

of the painful experience itself. This means that there will always be a gap between public and private meanings of sensation words; two people might be able to use a colour word, for example, in just the same way while nevertheless having different colour qualia. Thus, although they call the same things “red”, for instance, they *mean* different things by the word.

Presumably, though, if one is not inhibited by his desire to make himself understood by others, he might develop a phenomenologically precise language which could acknowledge the richness of his inner life. Instead of the various uses to which we put public words, this language would have but one purpose: to represent the intrinsic properties – the qualia – of his subjective experience. Having given himself private ostensive definitions of the words in his private language by assigning each to a corresponding quale, it is the inherent properties of the experience which fixes the meanings of the terms. That is to say, these properties themselves determine for him when and how to apply the terms of his language. They have the authority to do this because what lies before his awareness carries with it its own significance; it has a meaning independently of whatever purposes one might have in naming it. Such a language, being tailor-made to fit the properties of one’s own experience, would avoid the above epistemological difficulty of knowing what someone else means in using a word. Since one cannot be in the dark about what one means to refer to, there is no possibility of any discord between how a word is used and what it actually *means* to the person using it.

Refining language in this way, in the private language remarks, involves a gradual stripping away of all hints of a practice. Unlike public language, a private language is not supposed to be governed by publicly accessible behavioural criteria for meaning ascriptions; this is exactly why it cannot be understood by others. We might insist, however, that even if the private linguist cannot be justified in his claim to be following a rule by reference to his behaviour or any other public criteria, he might at least be able to justify it to himself. In order to do this, he must show that his use is governed by a criterion. As we will see, however, such a criterion is just what cannot be given; at this stage of bare introspection there is simply nothing left that would count as one. On the one hand, then, in claiming that he is following a rule, the private linguist is forced to posit a criterion for the correct use of the term. On the other hand, however, his attempts to characterize this criterion are hindered by the requirement that the meaning of the concept not be manifested in his behaviour, since this would enable other people to learn the language. As mentioned earlier, this has the effect of making what the private linguist is doing unintelligible as a form of language. In the next section, I move on to the details of what I take to be a persuasive argument against the possibility of private language.

2. 4. The “No Stage-Setting” Argument

As mentioned above, one motivation for wanting to develop a private language is that it would be free of the instrumental demands to which public language caters. A private language would be “a disembodied language, lacking any connection with the

objective world”;⁷³ its only task would be to trace the contours of one’s private, inner experience. This is to suppose the independence of the “contents” of self-consciousness from their natural bodily expression; given that the two bear only a contingent, causal relationship to one another, presumably the latter might be subtracted from our picture of language without any significant loss to its meaning. Then, instead of being refracted through the imperfect medium of public language, the contents of our private experience might be finally able to “speak for themselves”. As we will see, however, this is to misrepresent the nature of ostensive definition. Far from purifying or perfecting language, our stripping it of any instrumental use in our lives only drains the notion of all content. As mentioned earlier, this has the effect of revealing a *logical* connection to hold between the inner and the outer.

In PI 257, Wittgenstein introduces the example of a child who cannot be taught the meaning of the word “pain” in the usual way, because his pain is never outwardly manifested; we would never be in the position to inform him, for example, “You have a toothache”. Wittgenstein begins his remarks on private language by making it clear that ordinary our sensation vocabulary is not already private or semi-private in any of the ways described above. We do not teach a child the meaning of words like “pain” by directing his attention inwards, but by teaching him to replace the natural, behavioural manifestations of his feelings with verbal expressions. It is only because there is a great deal of agreement as to what constitutes pain-behaviour that we are able to teach the

⁷³ McGinn, *Wittgenstein and the Philosophical Investigations* (New York: Routledge, 1997) 118.

meaning of the word “pain”. A child who lacks these natural expressions will be unable to learn the meaning of the word.

But perhaps, Wittgenstein’s interlocutor suggests, the child is a genius who takes it upon himself to name his sensations. In the final lines of PI 257, Wittgenstein argues that the presumption that one could do such a thing oversimplifies the act of ostensive definition, which requires “stage-setting” within an already existing language:

... But what does it mean to say that [one] has ‘named his pain’? – How has he done this naming of pain?! And whatever he did, what was its purpose? – When one says “He gave a name to his sensation” one forgets that a great deal of stage-setting in the language is presupposed if the mere act of naming is to make sense. And when we speak of someone’s having given a name to pain, what is supposed is the existence of the grammar of the word “pain”; it shews the post where the new word is stationed (PI 257).

In this passage, Wittgenstein reminds us of a point developed earlier in the *Investigations*: successful ostensive definition requires that the term being defined have a pre-existent place in the language. The act of ostensive definition is not the naming of a pre-defined object, like a person or a pet, but the definition of a concept; this is done by pointing to objects to which the word refers and at the same time saying the word, so that the hearer learns to associate the word with the thing. On the surface it seems that the entire meaning of the word being defined, along with the rules for its use and therefore the possibility of using it wrongly, would be grasped when one understands the thing to which it refers; for it looks like the role of at least some words is just to refer to objects in this way.

The problem with this conception of how language is taught becomes clear when we note the ambiguity involved in giving something a name. Any object will have a

number of properties – colour, shape, consistency, and so on – and simple pointing will not settle which of these the word is to stand for. Wittgenstein makes this point in the following passage of the *Blue Book*:

But the ostensive definition will not suffice by itself, because it can always be variously interpreted.... I explain the word ‘tove’ by pointing to a pencil and saying “This is called ‘tove’”. I may be taken to mean “This is a pencil” or “This is round” or “This is wood,” etc. (BB 2).

The rules for the use of the word, and thus the possibility of using it incorrectly, are not set up by the act of pointing; in order for ostensive definition to be successful, rules specifying how the concept is to be used, what is to count as the same thing, and so on must be laid down in advance of the act in order to determine the nature of the thing being identified. One will understand his teacher when she says, for example, “This is the ‘king’” if he knows that she is talking about a piece in a game, but without such background knowledge her instructions can always be variously interpreted – “king” might be the definition of a colour, shape, or an entirely new concept. Thus the act of naming is not the first act in the defining process, but the final one after establishing the “post” where the new word is to be stationed; genuine acts of naming “take place within a complex social setting governed by an array of conventions which determine a particular language-game in which the name has a subsequent use.”⁷⁴ This point is overlooked by the putative private linguist, whose first mistake to believe that “[o]nce you know *what* the word stands for, you understand it, you know its whole use” (PI 264).

⁷⁴ P. M. S. Hacker, *Wittgenstein, Meaning, and Mind*, (Oxford: Blackwell Publishers, 1997) 101.

According to Wittgenstein, the presumption that the meaning of a word is secured by the thing to which it refers is a symptom of our “tendency to sublime the logic of our language” (PI 38) – our tendency, that is, to think of the collection of overlapping uses for our concepts as concealing a single, unique essence which holds this multiplicity together and must be dug out and purified by philosophical analysis. Naming, as one feature of the logic of our language, “appears as a *queer* connection of a word with an object,” because the philosopher is trying “to bring out *the* relation between the name and the thing by staring at an object in front of him and repeating a name or even the word ‘this’ innumerable times. For philosophical problems arise when language *idles*” (PI 38). Thus we arrive at the idea that the act of naming is a peculiar, distinct kind of connection between a word and a thing – a hidden relation whose nature, though ordinarily concealed by our informal and inexact everyday practices, may be revealed by removing it from its empirical context in our daily lives and considering it in the abstract. This way of thinking lends itself naturally to the idea that one might establish a private link between a word and a thing – that the bestowal of meaning is a “private” or “mental” process connecting a sign with an object, a process originating within each of us individually which we then choose whether or not to make public. While of course one might secretly connect a word with an object⁷⁵ – and in this case, it would be perfectly appropriate to call such a thing “private” (as in a private code), or even “mental” –

⁷⁵ Such a thing would be akin to a secret code rather than a necessarily private language. The difference between the two might be brought out by noting the difference between imagining one inventing a game which nobody ever happens to play, and imagining one inventing a game in a context in which no games exist, or ever have existed: “As things are I can, for example, invent a game that is never played by anyone. – But would the following be possible too: mankind has never played any games; once, however, someone invented a game – which no one ever played?” (PI 204).

excessive attention to such exceptional cases threatens to distract us from the broader context of rule following which forms the ground of the possibility of such an act, as well as the context by reference to which it has meaning. Thus the private language argument can be seen as part of a larger attack on the assumption that words and the circumstances of their use are analytically separable.

Some philosophers take the diary case of PI 258 to be an extension PI 257: where a word has no use, it has no meaning, so “here we can’t talk about right.”⁷⁶ Malcolm Budd writes:

[The private linguist’s] act of private ostensive definition does not give any content to the idea that it would be *correct* for him to write ‘S’ down on certain subsequent occasions and *incorrect* for him to write ‘S’ down on certain other occasions. For the combination of an act of attention to a sensation and the utterance of ‘This is called “S”’ does not determine the meaning of ‘S’: any ostensive definition can be variously understood. It is the way in which a sign is used, or is intended to be used, that determines its meaning, and the concentration of a person’s attention upon a sensation as he speaks or writes down the sign implies nothing about how the sign is to be used.⁷⁷

In the diary case, the private linguist seems at first glance to have established a limited use for the sign “S” – he wants to record “the recurrence of a certain sensation.” In order to accomplish this, it seems that all he has to do is, confronted by various mental states, judge some the same (or similar enough) and others different. But what do we mean by “the same” here? How similar is similar enough? The word “same” can also be variously applied, and to talk of using it rightly or wrongly will have to have reference to

⁷⁶ For example, McGinn (1997), Malcolm Budd (*Wittgenstein’s Philosophy of Psychology* (London: Routledge, 1989). Some later remarks by Glock can also be read along these lines, for example: “Concentrating one’s attention cannot establish criteria of IDENTITY for subsequent uses of ‘S’. Such criteria can be provided only by specifying what kind of thing is at issue through a sortal term” (1996: 312).

⁷⁷ Malcolm Budd, *Wittgenstein’s Philosophy of Psychology* (London: Routledge, 1989), pp. 59-60.

some purpose or broader context of use. The private linguist, however, lacks any such context; hence Budd's claim that no content can be given to the claim that the private diarist is on some occasions correct and on other incorrect.

Consider, for example, Wittgenstein's discussion of the concept of a "primary colour". Red, yellow, and blue are generally thought to belong to this category, but green is contentious because it is in some ways like, and in other ways unlike, the other three.⁷⁸ The painter Paul Klee had a system of five primary colours, and the philosopher C. L. Hardin, taking his cue from scientific experiments with light, insists that red, blue, and green are the only primaries.⁷⁹ It is tempting to think that we might look to the colours themselves to arbitrate disagreements about how the concepts are used; for "can't it be said that in some sense or other the grammar of colour-words characterizes the world as it actually is? ... Doesn't one put the primary colours together because there is a similarity among them, or at least put colours together, contrasting them with e.g. shapes or notes because there is a similarity among them?" (Zettel 331). The criteria for using the concept "primary colour", however, just identify the colours we want to pick out as members; so when we look to these included colours for verification of their membership, the similarity that we point to is these same criteria: "For if I say 'there is a particular similarity among the primary colours' – whence do I derive the idea of this similarity? Just as the idea 'primary colour' is nothing else but 'blue or red or green or yellow' – is not the idea of that similarity too given simply by the four colours? Indeed,

⁷⁸ It is, for instance, unlike other the other primaries in that it is a mixture of two colours, but at the same time unlike the other intermediaries in that it does not appear to be a mid-point between two primaries in the way that ochre is a yellowish-red and purple is a reddish-blue.

⁷⁹ "Infinitely many triples of light will qualify – there is no such thing as *the* light-mizing primaries – but, in order to maximize the range of producible hues, one member will always look mostly blue, one mostly green, and one mostly red" (C.L. Hardin, *Colour for Philosophers*, 2nd ed (Indianapolis: Hackett, 1993) 28).

aren't they the same?" (Zettel 331). When we point to the commonality amongst them, we are simply *recognizing* a concept's criteria – we are not *discovering* those criteria to be true. If the rules for membership in the concept "primary colour" were different, then we could look to these members and find the same similarity: "Then might we also take red, green, and circular together?' – Why not?!" (Zettel 331). For their similarity too would be just what the concept described for us.

Indeed, it would seem odd to suggest there is just one correct use of the concept, and that one person – either Hardin or Klee – must be using it wrongly. Is physics more suited to make such a pronouncement, or aesthetics? The answer, of course, is that the two are engaged in different language games oriented toward different purposes. Only within one of these linguistic frameworks are we able we make decisions about the truth or falsity of assertions about concepts. I can ask, for example, whether it is true that it is Thursday with reference to our calendar system, but I cannot meaningfully ask whether the calendar system itself is true, if it is *really* Thursday. It is similarly futile to try to justify the rules of language with something in the world that confirms them: "And saying that the rules of grammar are arbitrary is directed against the possibility of this justification, which is constructed on the model of justifying a sentence by pointing to what verifies it" (Zettel 331). A person engaged in scientific practices with an eye toward "maximizing the range of producible light hues" will be right to embrace Hardin's criteria, while a follower of Klee would do better adopt a flatter palette of mixed colours. It is only by reference to the context or way of life in which one employs the concept that we are able to make judgments about right or wrong use.

For the same reason, it is inappropriate to “talk about ‘right’” in the private diarist’s case in the absence of any practice for using “S”. The idea that we could do such a thing is rooted in the assumption that the meaning of the word “same” is self-evident – that, even while other words might be variously applied, this one, at least, is non-negotiable. As Norman Malcolm writes, “[t]here is a charm about the expression ‘same’ which makes one think that there cannot be any difficulty or any chance of going wrong in deciding whether *A* is the *same* as *B* – as if one did not have to be *shown* what the ‘same’ is”.⁸⁰ This may be due to an inclination to think, as Wittgenstein suggests, that “we have an infallible paradigm of identity in the identity of the thing itself” (PI 215). He goes on to challenge this assumption, however, asking: “Then are two things the same when they are what *one* thing is? And how am I to apply what the *one* thing shows me to the case of two things?” (PI 215). Depending on the situation, the statement “We live in the same house” might be used to convey a variety of things, for example, that we own remarkably similar homes, that we have the same phone number, that I can empathize with your family life, and so on. Like all words, the meaning of “same” is learned through training – according to Wittgenstein, it is taught through example and practice; we might, for instance, show a student a collection of samples of the same colour or of the same shape and then have him reproduce them. In doing so, Wittgenstein says, he masters a “technique”, which is to say that the activity is not one that can be carried out by simply following a set of prescribed rules. This technique will in turn show up as he exercises it in a practice.

⁸⁰ Norman Malcolm, “Wittgenstein’s Philosophical Investigations,” in *The Philosophical Review* 63 (1954) 536.

In the case of a private language, this practice is said to be the private one of recognizing a sensation or collection of sensations as the same as the one I had originally labeled “S”. But here we run into the same problem that confronted us in the primary colours example: how is one to justify his use of the word “same” here? As we just saw, Wittgenstein rejects any attempt to find justification for our concepts outside of the language-game in “the nature of things”. The perceived similarity between red, green, and yellow, for instance, is just that they meet chosen criteria. That is, their similarity does not *justify* their belonging to a particular concept – it is *the same thing*. In the same way, the justification for grouping two samples under the concept “S” will refer to the rules for using the concept rather than the sensations themselves. As Wittgenstein writes,

We imagine sometimes as though they were so many different things in a box – colour, sound, pain. But it makes no sense to say “I have this”, pointing within myself. ‘I am inclined to say’ that pain is a sensation – but I can’t justify this by a noticeable similarity among the contents of a box – colour, sound, pain. The similarity must be in the *concept* (WWLP, p. 62).

What the samples S_1 and S_2 will have in common is that they meet chosen criteria, which will in turn reflect some purpose or manner of living.⁸¹ But if the meaning of “S” is to be private, it cannot connect with one’s behaviour or circumstances in any regular way; for if another person, by noting this behaviour, could learn how to use “S”, then the meaning of the term would be rendered public. The result is that neither we nor the private linguist have any resources with which to justify his use of the term. The original ceremony did not establish a standard of use or “range of application”⁸² for “S”; it only

⁸¹ As Wittgenstein puts it, “to imagine a language means to imagine a form of life” (PI 19).

⁸² B. Goldberg, “The Linguistic Expression of Feeling,” *American Philosophical Quarterly* (1971) 88.

seemed to because we imagined similarity or “sameness” as something like an absolute identity that could be read off the samples themselves; this imagined identity is what led us to believe that we were already in possession of the truth conditions for whether the diarist remembered right. The claim that the rules of grammar are arbitrary – as opposed to forced upon us by “the things themselves” – is meant to undermine this conception of identity as anticipating reality in an absolute or exclusive way: “[A] sound is an expression only as it occurs in a particular language-game, which should now be described” (PI 261). With nothing but individual will to determine my use of the term, the notion of correct use that will exist within the private language will be such that if I *believe* that I am right then I am right.

In PI 270, Wittgenstein begins to restore a context of use to the example; in this passage the private linguist discovers that his entry of the term “S” in his diary corresponds to a rise in blood pressure. Once the term is put to use, we have some criteria by which to judge that his use of the word on one occasion is the same as on another. What S_1 and S_2 have in common is that they signal a rise in blood pressure, and “S” may be used wrongly if one violates this requirement. At this point, however, “S” names a sensation because of the way in which it is employed within a language-game. Once the conditions for ostensive definition are fulfilled, the result is to render the language public.

Wittgenstein compares the grammar of our everyday language with a pair of scales. Both are public, their meaning resting upon an agreement in society. We can weigh something on a pair of scales, and we can express something in language; we

might even do so privately, without telling anyone. But it is only the pair of scales or the grammar that allows us to do this. An ideal language, then, would be like an ideal unit of measurement. Various units of measurement can be used to determine the weight of an object, say, a brick of cheese. Though different units of measurement will yield different numerical results, we still consider them to be measurements of the same weight. Of course, depending on the situation, some will be more practical than others; “metric tonnes”, for example, would be out of place in a deli. But would it make sense to insist that, practical considerations aside, there is one form of measurement which is best? This would seem an odd suggestion, and is perhaps based on a false inference from the observation that there are more or less accurate or appropriate measurements in certain situations to the hypothesis that there must be a form of measurement that is ideally accurate in some absolute sense.

Perhaps a better analogy for an ideal language is that of an ideal clock: “‘The ideal clock would always point to the time “now”.’ This also connects us with the language which describes only my impressions of the present moment. Akin is the primal utterance that is only an inarticulate sound... The ideal name, which the word ‘this’ is” (*Remarks on the Philosophy of Psychology*, Vol. I, #721). Such a clock would point only to the present moment – not as one conventionally defined moment amongst others, but in some absolute sense which need not answer to any particular instrumental need. This might suggest to us that it is therefore more pure or more accurate, and we could then use that “ideal” measurement of the present to judge the accuracy of various time-keeping devices. In the same way, we might use a private qualia language to serve

as a criterion for the validity of ordinary concepts. Of course, this is clearly absurd. A clock which pointed only to the present moment would not only be useless – it would not be a clock at all. Analogously, a private language which employed concepts only by referring to the content of my immediate experience could not be used to say anything, and so would not be a language. Rather than refining or perfecting language, we have only emptied the notion of all content.

Briefly summarized, then, the argument is that ordinary criteria for ascribing language use to a person involves reference to a normative regularity, observed in terms of an agreement between what she utters and how she acts. In the private case, however, all we could say is the private linguist has undertaken to use “S” in a certain way. This suggests a connection between what someone means by a word and the role that word plays in her actions, which the ineffability intuition severs. If this intuition implies that there is a subjective realm of experience where the subjective component of a word’s meaning is privately constituted, then it is misguided. Meaning is not located in the mind or soul or cerebral cortex; it is not *located* anywhere. When we ascribe meaningful language use to someone, we do so on the basis of publicly observable criteria. The point is not that someone’s employing a language is something for which we need behavioural evidence, but that what we call language employment is something that is manifested *in behaviour*.

I want to end this section with a suggestion as to what the ineffability intuition is really getting at. As we saw, this intuition had to do with a sense of running up against the limits of language. Of course, no one will deny that this sometimes happens; we have

all had experiences that we could not describe to others. But from the fact that we *sometimes* struggle to make ourselves understood, it does not follow that we *always* do, that some portion of our experience always slips through the cracks of public language. Rather, if language is a technique – a knowing *how* rather than a knowing *what* – then one can be better or worse at it. Our experiences, we saw, do not “speak for themselves,” but can be characterized in an indefinite number of ways, which, depending on the circumstances, will strike us as more or less accurate, or fitting, or poetic. This is why we have poets and writers describe them for us. Consider this short love poem by Rilke:

Understand, I’ll slip quietly
away from the noisy crowd
when I see the pale
stars rising, blooming, over the oaks.

I’ll pursue solitary pathways
through the pale twilit meadows
with only this one dream:
You come too.⁸³

I dare say, it beats “S”.

2. 5. *Private Language and the Problem of Other Minds*

What the private language argument is supposed to establish has often been overstated. It might be helpful, then, to begin this section with what the argument does *not* do. In particular, it does not provide a knock-down proof of the existence of other minds. Wittgenstein is sometimes said to refute skepticism by proving, in short, that language is an inherently *social* practice that would not exist if not for the existence of

⁸³ Rainer Maria Rilke, *Rilke on Love and Other Difficulties*, translated by J. L. Mood (W. W. Norton & Company, 2004).

other people. This is to suggest that all of our social practices, “say, the sacrament of marriage, or the history of private property, or the ceremony of shaking hands, or ... ultimately the existence of language, constitutes a proof of the existence of others.”⁸⁴ Stanley Cavell is surely right to point out, however, that this begs the skeptical question: “Because if what we ‘accept’ as human beings ‘turn out to be’ automata or aliens, then can’t we take it that automata or aliens marry and own private property and shake hands and possess language?”⁸⁵ From the fact that these beings might participate in social forms, it does not follow that they are therefore human.

Wittgenstein’s point is that language is necessarily a practice, not that it is necessarily a social practice. While this does not provide the conclusive refutation of other-minds skepticism for which we had hoped, the argument does have important implications for it. Descartes gave the problem of other minds what is perhaps its clearest expression in the *Meditations*, when he described looking from a window and seeing what seem to be men crossing a square: “And yet what do I see from the window but hats and coats which may cover automatic machines? Yet I judge these to be men.”⁸⁶ His point is a simple one: I can observe other peoples’ outward appearance and overt behaviour – I can, in principle, even observe their brains and nervous systems – but I cannot observe their mental states. I am not able to identify another person as a subject of experience by *seeing* that she is one; rather, I do this by reasoning that she is “the subject of those experiences which stand in the same unique causal relation to body N as

⁸⁴ (Cavell: 1996) 329.

⁸⁵ *ibid.*

⁸⁶ René Descartes, *Meditations on First Philosophy* tr. John Cottingham (Indianapolis: Hackett, 1986) 21.

my experiences stand to body M.”⁸⁷ Thus I learn the concept of “pain” on the model of my own pain, as it stands in its special relation to my body, and then extend this model to other bodies which are not my own. But as Wittgenstein remarks,

this will be a none too easy thing to do: for I have to imagine pain which I *do not feel* on the model of the pain which I *do feel*. That is, what I have to do is not simply to make a transition in the imagination from one place of pain to another. As, from pain in the hand to pain in the arm. For I am not to imagine that I feel pain in some region of his body (PI 302).

My own pain is inextricably mine; thus if the meaning of the word “pain” were established on the model of “object and designation”, I could not possibly extend it to other people. Ironically, it would seem odd to even call a pain my own, since in no logically conceivable case could it be otherwise.

The Cartesian picture of language, according to which we learn the meaning of words like “pain” by acquainting ourselves with the object for which it stands, therefore leads to untenable consequences. Not only are we unable to speak of each others’ inner states, but we can not even talk about our own. How, then, are we to account for the fact that we do have a language for mental states? This is the point of Wittgenstein’s famous “beetle in the box” thought experiment:

Now suppose someone tells me that *he* knows what pain is only from his own case! – Suppose everyone had a box with something in it: we call it a “beetle”. No one can look into anyone else’s box, and everyone says he knows what a beetle is only by looking at *his* beetle. – Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. – But suppose the word “beetle” had a use in these people’s language? – If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a

⁸⁷ P. F. Strawson, “Critical Notice of Wittgenstein’s *Philosophical Investigations*”. *Mind*, LXIII (1954).

something: for the box might even be empty. – No, one can ‘divide through’ by the thing in the box; it cancels out, whatever it is.

That is to say: if we construe the grammar of the expression on the model of ‘object and designation’ the object drops out of consideration as irrelevant (PI 293).

If we were each to obtain our psychological vocabulary by pure introspection, then we could not possibly go on to use it as part of our common language; for if only I know what is in my box, then only I can know what the word I use to refer to it means. So if we do have a common language in which we talk of mental states – and we do – then the terms of the language cannot refer to private inner entities.

With this in mind, we can better understand Wittgenstein’s insistence that “an inner process stands in need of outward criteria” (PI 580). This remark is often read as claiming that there is a conceptual connection between the inner and the outer, but there is little agreement as to how this connection is to be understood. I suggest that it should be read as pointing to our ability to recognize behaviour as *expressive* of mind. It is for this reason that we are able to have a psychological vocabulary at all:

How do words *refer* to sensations? – There doesn’t seem to be any problem here; don’t we talk about sensations every day, and give them names? But how is the connexion between the name and the thing named set up? This is the same as: how does a human being learn the meaning of the names of sensations? – of the word “pain” for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour.

“So you are saying that the word ‘pain’ really means crying?” On the contrary: the verbal expression of pain replaces crying and does not describe it (PI 244).

What makes it possible for us to have a common language of mental states is that we are able to recognize those states as finding a natural expression in bodily behaviour. Our psychological vocabulary is learned from others in contexts in which they are aware of us expressing those states, and at the same time that we learn to apply it to ourselves, we learn to apply it to other people as well. The upshot is that behaviour of others is not *evidence* of their mental states, but the only possible criteria for what one means when one ascribes psychological predicates to them.

Now we can begin to see more clearly the implications of the private language argument for the problem of other minds. The Cartesian assumption is that the relation between sensation and behaviour is a contingent, causal one, and we can only hope that this relation can form the basis for an inference from behaviour to state of mind. But for Wittgenstein “this is precisely what we *do not* do.... We do not need to believe (with justification) that others are human beings to treat them as such. We relate to them as fellow thinkers and sufferers, which is to say we acknowledge them, their pains, their need for help.”⁸⁸ Even if we were to decide that such inferences from behaviour to mind are plausible, they are bound to betray the naturalness and certainty which characterizes our engagement with other people as conscious beings like ourselves and not as mere things. This engagement is not the result of our ability to make quick, imperceptible inferences about mental states, but of our capacity to just *see* the behaviour of others as expressive of a mind – to see discomfort in a certain posture, shame in a downcast face, love in a person’s eyes: “Consciousness in another’s face. Look into someone’s face,

⁸⁸ Alessandra Tanesini, *Wittgenstein: A Feminist Interpretation* (Malden: Polity Press, 2004) 112.

and see the consciousness in it, and a particular *shade* of consciousness. You see on it, in it, joy, indifference, interest, torpor, and so on. The light in other people's faces" (Zettel 220). On many occasions, we are able to read a particular shade of consciousness off a person's expression. This is why we can say things like, "His guilt was written all over his face", "I can see it in your eyes", "She is putting on a brave face", and so on. In describing another's behaviour, we do not start by identifying the mere physical movements of a physical thing and then try to infer inward to some concealed state of mind. Rather, we describe behaviour as joyful, sad, lively, and so on. This is not to say that people are completely transparent; some people may be easier to read than others, and it is always possible for a good actor to conceal his feelings. We must learn our way around both individuals and cultures, and, of course, we might fail. But when we do succeed in understanding another in this way, it is not because we have discovered how to make better inferences, but because we have become more familiar with her idiosyncrasies, more sensitive to the fine detail in her actions, more attuned to how these details fit into the looser patterns of human behaviour, much like how one learns an art form.

This is the lesson of Wittgenstein's well-known but often misunderstood remark: "My attitude towards him is an attitude toward a soul. I am not of the opinion that he has a soul." To talk of opinion or belief in this case would be inappropriate, for it suggests that I am mounting a hypothesis on the basis of evidence. My response to another person as a conscious being is something far more immediate and unreflective; it shows up in my natural, spontaneous ways of acting toward someone who is, say, in pain:

It is a help here to remember that it is a primitive reaction to tend, to treat the part that hurts when someone else is in pain.... But what is the word “primitive” meant to say here? Presumably that this sort of behaviour is pre-linguistic: that a language-game is based on it, that it is the prototype of a way of thinking and not a result of thought (Zettel 540-541).

When a child cries out upon, say, spilling boiling water on himself, we respond at once with first aid and comforting words – we do not pause to wonder whether the phenomenological character of his sensations matches our own, or consider them to be otherwise mysterious: “If I see someone writhing in pain with evident cause I do not think: all the same, his feelings are hidden from me” (OC, p. 223). It is in virtue of this flurry of human activity, this pre-reflective attitude that we have toward one another as minded, that we have the concepts of the soul or consciousness in the first place. It is no wonder, then, that they resist further justification in causal terms. I do not respond to another’s pain or need or simple presence because I have decided that he has a soul; rather, my assertion that he has a soul has rather put into words or rationalized my way of responding to him. If we abstract from this natural engagement with other people and ask for some deeper justification for this responsiveness to others, we cannot find it.

This, I think, is the lesson that we should take away from other-minds skepticism. Here we have looked “for an explanation where we ought to look at what happens as a ‘proto-phenomenon’. That is, where we ought to have said: *this language-game is played*” (PI 654). The evidential gap that the skeptic wants to exploit is the result of his attempt to separate the mind from its natural expression in behaviour, to make the link between them a contingent, causal one. But in everyday life it never occurs to us to question the existence of other minds, because no such gap is experienced.

2. 6. Conclusion

In PI 207, we saw, the absence of any discernible regularity between what utterance and behaviour had the effect of disqualifying the verbal activity of the group from being a language. In the private language, we saw that this was no less true of our psychological vocabulary: what makes it possible for us to have a language of mental states is that those states find a natural expression in bodily behaviour. In both cases, this served to reveal the logical connections between utterance and behaviour. This expressive relation between mind and behaviour, I think, is crucial for philosophy of mind. In my third and final chapter, I will look at the extent to which the “object and designation” model of mental concepts is still at work in contemporary philosophy of mind.

Chapter 3: The “Hard Problem” of Consciousness

At this point in our intellectual history, an attack on Cartesianism might seem redundant. Talk of conscious brains has replaced talk of conscious souls in the philosophy of mind, and naturalism has become the default position for all serious inquiry. As it turns out, however, many of the problems which plagued Descartes are still with us. For this reason, it is important to note that Wittgenstein does not resist the Cartesian first-person perspective in any of the usual ways, for example, by arguing against the likelihood that there might be causality across metaphysical gaps, or by insisting that psychological terms in fact refer to the body. Such efforts by materialists to “exorcise the ghost from the machine” leave the Cartesian conception of the body intact and with it the uncomfortable fit between the conscious mind and the universe as described by natural science. Instead, Wittgenstein’s attack on Cartesianism involves undermining the picture of the body that is at the heart of both dualism and materialism: a picture of the human form as expressively vacant, with no significant role to play in our account of mentality. I have argued that recognizing that there is an expressive relation between inner states and outer behaviour – a relationship that is neither a contingent causal relation, nor one of identity – is key to dissolving the problem of other minds, at least in so far as this can be done. In this chapter, I want to suggest that it is also key to dissolving the mind-body problem in its modern form, as it appears in contemporary philosophy of mind.

In what follows, I will examine the extent to which the “causal” conception of mental states is still at play in mainstream philosophy of mind. Again, the mind-body

problem provides an opportunity to examine how philosophical views about what it means to explain consciousness square up with our ordinary understanding of what it means to be conscious. I argue that the naturalist assumption that consciousness must be understood as a causal process can be shown to distort the grammar of this concept as it appears in our everyday life. First, I will examine the mind-body problem as it appears in its modern form in contemporary philosophy of mind. An overwhelming amount of literature has been produced on this topic, and it would go beyond the scope of this project to give it the attention it deserves. Instead, I will briefly outline the problem before moving on to argue, much as I did in chapter two, that there is a way to understand consciousness such that the mind-body problem simply does not show up. Next, I will examine two approaches to the analysis of psychological concepts – behaviourism and functionalism – which, I suggest, misrepresent our everyday understanding of mentality by overlooking the expressiveness of behaviour. In doing so, I do not mean to suggest that the appeal to how words are used carries any principled criticisms of constructive work on consciousness; rather, it throws that work into a particular light.

3. 1. The Hard Problem

The problem of other minds is just one sub-species of a broader mind-body problem. In its modern form, this problem still shows up as one of fitting the conscious mind into the universe as described by natural science, and, in particular, the body. Given that there is mind as well as body, the question goes, how are the two related? In the 1960s, an answer seemed close at hand; materialists had long objected to the

supernatural implications of dualism, but now advancements in neuroscience and computer technology gave them new hope that the mind, too, might finally yield to the treatment of science. As the cognitive revolution gained momentum, one by one the obstacles to understanding the mind naturalistically seemed to fall away, and consciousness seemed neatly transferable from the soul to the brain. It did not take long, however, for someone to point out that the mind-body problem had survived the cognitive revolution. In 1974, Thomas Nagel published the unusually-titled article “What is it Like to Be a Bat?”, in which he argued that philosophers of mind, in a fit of “reductionist euphoria,” had neglected something crucial: the subjective character of consciousness.⁸⁹ In addition to whatever physical processes are happening in one’s brain to produce a certain experience, there is “something that it is like” to have that experience, something that it is like to be that particular conscious being, and this “something” was not captured by the new materialism. So, for example, in addition to whatever science can tell us about a bat’s life – its nocturnal habits, its sonar location system, and so on – it cannot tell us “what it is like for a *bat* to be a bat.”⁹⁰ “Consciousness,” Nagel declared, “is what makes the mind-body problem really intractable.”⁹¹ Recent philosophy of mind had simply sidestepped the difficulty by omitting it from its analyses.

Though Nagel’s article was directed at a fairly crude – and by now outdated – form of identity theory, his main point has survived. In the mid-90s, David Chalmers

⁸⁹ Thomas Nagel, “What is it Like to Be a Bat?,” *The Philosophical Review* 83 (1974) 435.

⁹⁰ *ibid.*

⁹¹ *ibid.*

reformulated the complaint in terms of a distinction between what he called the easy problems and the hard problem of consciousness. The easy problems have to do with describing the way that the brain works from the third person objective viewpoint of science. From this perspective, “the brain is relatively comprehensible. When you look at this page, this is a whirl of processing: photons stirring your retina, electrical signals are passed up your optic nerve and between different areas of your brain, and eventually you might respond with a smile, a perplexed frown or a remark.”⁹² These problems are not easy in the sense that their solutions are always close at hand, but in the sense that nothing in principle stands in the way of our answering them. By contrast, the hard problem of consciousness is marked by a seemingly unbridgeable explanatory gap. This problem concerns how physical processes in the brain give rise to qualitative consciousness. We know, for example, the visual perception is associated with activity in the visual cortex, but the question of why these physical processes are accompanied by this experiential state, or indeed by any subjective experience at all, seems beyond the scope of natural science. “Remarkably,” Chalmers wrote, “subjective experience seems to emerge from a physical process.”⁹³ The question is: how and why does this happen?

What Chalmers’ article emphasized is that the hard problem is a still problem for functionalism, which had by then replaced the identity theory as the most promising candidate for explaining the nature of mentality. Many philosophers had felt that Nagel’s complaint could be avoided if we understood the relation between mental states and the

⁹² David J. Chalmers, “The Puzzle of Conscious Experience,” *Philosophical Problems*, eds. Laurence Bonjour and Ann Baker (New York: Pearson Education Inc., 2005) 405.

⁹³ *ibid.* 406.

physical brain as the relation between “functional states” and their “physical instantiation.” According to functionalism, mental states are to be characterized in terms of the tasks they executed; more specifically, the identity of a mental state is to be determined by its causal relations to sensory stimulations, other mental states, and behaviour. Consciousness, then, can be explained in functional/behavioural terms as a disposition to respond to one’s environment in a systematic way, one which can be described abstractly in computational terms and realized in neurophysiological structures. Though functionalism avoided many of the pitfalls of its predecessors, Chalmers argued that it had not circumvented the hard problem; functionalists were no more able to account for the experiential aspects of consciousness than the cruder identity theorists who had been Nagel’s original target. Granted that mental states are functional, information-processing states realized in the brain, the question remains: “Why is the performance of these functions accompanied by conscious experience?”⁹⁴ Why should physical processing give rise to qualitative experience at all?

Reactions to hard problem have been various. Most philosophers maintain that the problem, though indeed hard, will eventually be solved using the right kind of philosophical reflection and the resources already available to us in contemporary neuroscience and cognitive science. Others suggest that solving the hard problem might require a radical alteration in the way we understand reality. Chalmers, for example, wonders whether a new metaphysics – a “theory of information” – might be in order; perhaps “[t]he laws of physics might ultimately be cast in informational terms, in which

⁹⁴ *ibid.* 410.

case we would have a satisfying congruence between the constructs in both physical and psychophysical laws.”⁹⁵ Last, there are those philosophers who feel that a solution might simply be beyond human reach. Colin McGinn, for example, wonders whether we might not be faced with a kind of “cognitive closure” in regard to consciousness – while there may very well be an explanation of consciousness in physical terms, it might also be that we will never know what it is.⁹⁶ Nagel himself seems not to want to commit either way, though in *The View from Nowhere* he reiterates that the objective methods of natural science “cannot provide a complete picture of the world,” since they cannot capture the subjective facts of consciousness.⁹⁷

Most interesting for my purpose is Oswald Hanfling’s Wittgensteinian take on the hard problem. Hanfling suggests that the professed mystery is an illusion, a pseudo-problem rooted in the philosophical misuse of words. In particular, he argues that philosophers approach the subject of consciousness by asking questions appropriate to a substance or thing, and find, predictably, that they elude answer. In the opening passages of the *Blue Book*, Wittgenstein identifies this as “one of the great sources of philosophical bewilderment: a substantive makes us look for a thing that corresponds to it” (BB p. 1).⁹⁸ Philosophers, Hanfling argues, have taken the investigation into consciousness to be a causal scientific one into a *thing*; but consciousness is not a thing, nor is it a property of a thing (i.e. the brain). Of course, we know that the brain plays an important causal role in

⁹⁵ *ibid.* 413.

⁹⁶ Colin McGinn, *The Problem of Consciousness* (Oxford: Blackwell, 1991).

⁹⁷ Thomas Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1986) 6.

⁹⁸ “The questions “What is length?”, “What is meaning?”, “What is the number one?” etc., produce in us a mental cramp. We feel that we can’t point to anything in reply to them and yet ought to point to something. (We are up against one of the great sources of philosophical bewilderment: a substantive makes us look for a thing that corresponds to it.) (BB 1)

giving us conscious experience, but the concept itself does not stand in need of a causal explanation or justification. As Wittgenstein argues, it is not scientific concept at all:

Psychological concepts are just everyday concepts. They are not concepts newly fashioned by science for its own purposes, as are the concepts of physics and chemistry. Psychological concepts are related to those of the exact sciences as the concepts of the science of medicine are to those of old women who spend their time nursing the sick (RPP, vol. 2, # 62).

In what follows, I will attempt to defend this claim via a conceptual analysis of mental concepts. As Hanfling writes: “Given a scientific question, the proper approach will be that of scientific methods; but *whether* a scientific question is scientific cannot itself be a scientific question. And similarly, the assumption, made by scientists or others, that a given question or claim makes sense, is not to be confirmed by scientific methods.”⁹⁹ First, I will argue that the hard problem of consciousness is not that consciousness is inherently inexplicable, but that it resists explanation in causal-nomological terms. If, however, Wittgenstein is right that “psychological concepts are just everyday concepts”, then they do not stand in need of such an explanation. Next, I will attempt to defend this claim by arguing that there is a perfectly legitimate sense in which consciousness can be explained ostensively, by showing how we ordinarily employ the word.

3. 2. “*Explaining*” *Consciousness*

The hard problem of consciousness is neatly summed up by Chalmers as follows: “The remarkable progress of science”, he writes, “has given us good reason to believe [that] for almost every natural phenomenon above the level of microscopic physics [there

⁹⁹ (Hanfling: 2002) 99.

exists] a *reductive explanation*, that is, an explanation wholly in terms of simpler entities”, such that “an explanation of the higher-level phenomenon falls out.”¹⁰⁰ Here Chalmers is pointing to the urge to give consciousness a reductive explanation, such that what are seen as its more problematic features might go unmentioned in our final tally of things. While there are philosophers who think that this might be done,¹⁰¹ most reject such explanations on the grounds that they simply do not do justice to the facts of our mental life. For Chalmers, this indicates that there is something peculiar about consciousness which sets it apart from the physical world. His complaint, at bottom, is the same one which motivated Descartes’ separation of the mind and the body into two ontologically distinct substances: the mind, unlike physical objects, does not seem subject to the law-governed regularities of cause and effect. For Descartes, the failure of the mind to be explicable in these terms was proof of its distinctive ontological character. For many philosophers today, it shows the intractable mysteriousness of consciousness.

For Chalmers, what makes consciousness so peculiar is that the physical world is “causally closed”, and consciousness is irrelevant to the explanation of behaviour.¹⁰² This, however, is precisely the sort of assumption that Wittgenstein wants to undermine. We do not explain others’ behaviour by first observing their brute physical movements

¹⁰⁰ David Chalmers, *The Conscious Mind* (New York: Oxford University Press, 1996) 42.

¹⁰¹ David Lewis, for example, propounds a type of “supervenience thesis” according to which mental properties, being dependent – or supervenient – on physical ones are not fundamental, and therefore not essential to complete description of the world. He uses the example of a painting and the pixels of colour from which it is comprised to illustrate his point. The picture, he writes, “supervenes on the pixels: there could be no difference in the picture and its properties without some difference in the arrangement of pixels.” Thus, he concludes, the former “could go unmentioned in an inventory of what there is without rendering that inventory incomplete.” See S. D. Guttenplan (ed.) *Companion to the Philosophy of Mind* (Oxford: Blackwell) 413.

¹⁰² *ibid.* 161.

and then inferring inward to a hidden mind; we simply *see* that behaviour as expressive of mind. Accordingly, we generally think that there is nothing illegitimate about incorporating mental states into our explanations of behaviour: “‘What caused me to go indoors was that I had become conscious of an unpleasant smell.’ This would be considered a perfectly good causal explanation.”¹⁰³ The claim that this does not qualify as a good causal explanation seems to rest on the assumption that there is just one legitimate conception of causality – as a nomological, law-like relation – and this assumption is itself questionable.

In her paper, “Causality and Determination”, G. E. M. Anscombe criticizes the assumption that causality always involves a necessary connection between two events, pointing out that we are rarely so strict about the use of this word as when we are doing philosophy. Hume, she writes, actually reinforced this conception of causation (rather than undermining it) by supposing necessity to be an essential feature of relations of cause and effect: “[A]s touching the equation of causality with necessitation, Hume’s thinking did nothing against this but curiously reinforced it. For he himself assumed that NECESSARY CONNECTION is an essential part of the idea of the relation of cause and effect, and he sought for its nature.”¹⁰⁴ In response, Anscombe writes that there are many occasions on which we are justified in claiming to know the cause of an event without at the same time suggesting that there is some exceptionless law which the causal relation instantiates. It is this urge to subsume consciousness itself under some general law, for

¹⁰³ (Hanfling:2002) 118.

¹⁰⁴ G. E. M. Anscombe, “Causality and Determination”, *Causation and Conditionals*, ed. Ernest Sosa (Oxford: Oxford University Press, 1975) 65.

example, which leads Chalmers to suggest a “theory of information” to “explain” consciousness. “The ultimate goal of a theory of consciousness,” he writes, “is a simple and elegant set of fundamental laws, analogous to the fundamental laws of physics.... I suggest that the primary psychophysical laws may centrally involve the concept of information.”¹⁰⁵ If in fact there are different ways of speaking of causality, however, then there is not one correct way in which, say, a perception, can be responsive to the world; mental states are simply responsive in a different way.

From the fact that our subjective experience cannot be given a causal-nomological explanation, it does not follow that it is inherently inexplicable. Norman Malcolm has criticized this suggestion, offering up an example of a discussion in which questions about “what it is like” to have a certain experience would seem more natural: “Suppose I sat next to a long-distance lorry driver at a lunch counter and we struck up a conversation. I might say to him ‘What is it like to be a lorry driver?’”¹⁰⁶ The lorry driver might then respond by describing the long hours and frequent nights away from home, by recounting particular experiences like being caught in a snow storm, and so on. While it is true that these are not scientific explanations, they are still perfectly intelligible and we may very well feel that we now have a good grasp of what it is like to be a lorry driver. One might still insist that even if we feel that we understand what this is like, we still cannot know what it is like *for a lorry driver* to be a lorry driver. But now

¹⁰⁵ David J. Chalmers, “The Puzzle of Conscious Experience,” *Philosophical Problems*, eds. Laurence Bonjour and Ann Baker (New York: Pearson Education Inc., 2005) 413.

¹⁰⁶ D. M. Armstrong and Norman Malcolm, *Consciousness and Causality*, (Oxford: Blackwell 1984) 46-47.

the claim looks more like a truism than it does a mystery to be unraveled: description is not interchangeable with experience.

The “what it is like” argument is meant to undermine various types of physicalism. According to the physicalist, consciousness is best characterized not in terms of subjective experience, but rather as a state of the brain or as something “produced” by the brain. Thus the physicalist shares with the anti-physicalist the assumption that consciousness is something which *should* yield to a causal analysis; where they disagree is on whether this can actually be accomplished. The anti-physicalist objection is that, however detailed our account of neuronal firings, information processing, and subsequent behaviour, it cannot fully capture what goes on when someone is, say, struck by the sight of sunlight sparkling off a clear lake, or moved by the sound of an old friend’s voice, or stunned by grief at the loss of a loved one. But why should we expect science to do such a thing? Granted that mental states cannot be described as instantiating some causal-nomological law, should this be enough to qualify it as “a mystery”? Were consciousness some sort of substance – the one substance, perhaps, that had proved impenetrable by science – then it would certainly be fair to declare it one, and in this case we would do well to experiment with new theoretical frameworks or experimental methods; but consciousness is not a substance. Nor is it properly described as a property of the brain, as when Daniel Dennett speaks of “our conscious brains” making assumptions, receiving information, arriving at conclusions,

and so on.¹⁰⁷ Ordinarily, we ascribe consciousness and other mental predicates to *living people*, not to their brains.

What each of these characterizations of consciousness have in common is that they take consciousness to be something which we must *infer* on the basis of evidence. It would appear that our describing one another as minded has all along been a conjecture. Descartes had only inferred wrong: it is the brain, and not the soul, which is “the seat of consciousness.”¹⁰⁸ It is curious, though, why we should think that transferring consciousness from the soul to the brain – locating it *here* rather than *there* – should make things any less problematic. For “[i]f the picture of thought in the head can force itself upon us, then why not much more that of thought in the soul?” (PI, part II, p. 178). And if ascriptions of consciousness to other people have always been abbreviated hypotheses about what is going on inside their heads, then why might we not form similar hypotheses about other bodily organs, or about statues, or about calculators?¹⁰⁹

Herein lies the metaphysical confusion behind the theory: philosophers taking a causal approach to the study of consciousness often write as if they are *explaining* or *justifying* our ordinary use of the word, when in fact it they only suggesting a different use for it. That is to say, when theories define consciousness as a causal process, they are not improving on a rudimentary theory of the mental processes that underlies our use of

¹⁰⁷ (Hanfling: 2002) 106. As Hanfling notes, Dennett seems to hold “both that brains *are* conscious and that – as discussed above – they *produce* consciousness.”

¹⁰⁸ This term is taken from Daniel Dennett: “How could the brain be the seat of consciousness?” (*Consciousness Explained* (Penguin 1991) 43).

¹⁰⁹ Indeed, Chalmers’ suggestion at the end of his article “The Puzzle of Conscious Experience” is that perhaps we *should* postulate such things about calculators and other devices. Nagel has also experimented with a version of panpsychism.

the concept, but are rather introducing novel ways of using it by introducing new criteria for its employment. Of course, we know that the brain plays an important causal role in our cognitive capacities and their exercise – and exactly what physical conditions are necessary for conscious experience is a matter for scientific investigation – but it does not follow that when we ascribe mental predicates to others we are (tacitly) ascribing to them a certain brain state, or forming any other sort of explanatory hypothesis. Rather, as mentioned in chapter two, we ascribe consciousness to a person who acts in certain characteristic ways. As Wittgenstein writes:

Only of a living being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious (PI 281).

If this is true, then the question “What is consciousness?” is not the same as “What causes consciousness?”. In what follows, I will suggest that a perfectly good explanation of what it means to be conscious might come in the form of an ostensive definition which simply points to “*the* bearer of mental predicates”¹¹⁰ – the living, corporeal human subject.

3. 3. *Consciousness and Behaviour*

One standard motivation for denying that consciousness can be explained ostensively in this way is the feeling that consciousness itself must be distinguished from the behaviour typically associated with consciousness or psychological ascriptions. If this is indeed a condition for a legitimate explanation of the phenomena, then it seems to

¹¹⁰ D. M. Armstrong and Norman Malcolm, *Consciousness and Causality* (Oxford: Blackwell 1984) 100.

have been met by contemporary theories. It is important to note that in all of the above explanations of the word “conscious”, little mention is made of any of the behaviour or background circumstances which we would ordinarily associate with consciousness. Whether the word is envisioned to refer to a substance produced by the brain, to our introspective awareness of mental objects, or to some neurobiological state, the suggestion seems to be that we need not imagine any behaviour or context in order to illustrate what it means. Indeed, a popular characterization of consciousness seems to be just whatever is left over after all the physical facts are in. As Chalmers writes, “[o]ur grounds for belief in consciousness derive solely from our own experience of it. Even if we knew every detail about the physics of the universe ... *that* information would not lead us to postulate the existence of conscious experience.”¹¹¹ At best, the behaviour of the conscious person provides inferential grounds for ascribing to a person that state. Noting certain analogies between his behaviour and mine, I infer that his behaviour must stand in a similar causal relation to his mental states as my behaviour does to mine: “We note regularities between experience and physical or functional states in our own case, postulate simple and homogenous underlying laws to explain them, and use those laws to infer the existence of consciousness in others.”¹¹² Chalmers concedes that this may not be the reasoning normally employed in believing that others are conscious, but claims that nevertheless “it seems to provide a reasonable justification for our beliefs.”¹¹³ And indeed, given Chalmers’ strict dichotomy between first-person subjective experience and

¹¹¹ (Chalmers: 1996) 101.

¹¹² *ibid.* 246.

¹¹³ *ibid.*

third-person observation, we seem to have no other choice; either we infer the existence of consciousness from our own first-hand encounter with it, or we do not come to know it at all.

The assumption that the behaviour of the conscious person bears no logical connection to their being in a conscious state is one reason that an ostensive definition of the word “conscious” – which points to the characteristic behaviour of the living human being – is bound to appear trivial or merely heuristic. As noted above, however, in ordinary practice, psychological ascriptions are not logically distinct from behaviour. To describe someone as conscious is to say that she is able to think, perceive, feel, and so on. Now consider the question: What makes it true that this person is conscious? One sensible answer is: that is just what “conscious” means. If asked why these actions illustrate the meaning of the word “conscious”, we might point to various features of the person’s behaviour – she is listening to the Beatles, she is admiring the artwork, etc. – and of the context. In doing so, one offers reasons – justifications – for using the word in this way.

Of course, one might object here that offering examples of conscious behaviour has little to do with a conceptual analysis of the term “conscious”. At best, this has illustrated the meaning of the term in its everyday role – it has not explained it; and while such an illustration might be helpful in teaching someone the *use* of the term, it hardly constitutes a genuine *explanation*. This objection is based on the assumption that giving an explanation must be a different sort of activity than offering a rule for the use of a word; for example, one that uncovers an object or causal process to which the term refers.

And this view certainly has a distinguished past. We might reasonably enough suppose, along with Aristotle, that to explain something just means to give an account of the causes underlying its being the kind of thing it is. Of course, this idea is devoid of content until the relevant sense of “cause” is specified. In contemporary philosophy of mind, we saw, it seems to be most often a nomological, law-like relation, such that consciousness is analyzed in terms of a relation obtaining between a psychological state and its causes which instantiates a law.

This seems to be what the physicalist D. M. Armstrong has in mind, for example, when he suggests “a Causal analysis of the mental concepts.”¹¹⁴ A mental concept is “the concept of a state that is *apt to be the cause of certain effects or apt to be the effect of certain causes.*”¹¹⁵ An analogy in the non-mental world might be the concept of a poison, by which we mean “that, whatever it is, which produces certain effects.”¹¹⁶ This allows for the possibility of the scientific identification of poisons. Likewise, a person’s mental state might be characterized as “a cause within him that brings about, or tends to bring about, [a] particular line of conduct.”¹¹⁷ Armstrong uses the example of “a man’s purpose to go to the kitchen to get something to eat.” In this case, the purpose would be the cause within him – probably within his brain – which brings about his entering the kitchen. But, as Hanfling points out, this would be to posit a relation of cause and effect between having the purpose to perform a certain task and performing that task; to the question “What caused him to go to the kitchen to get something to eat?”, we would have

¹¹⁴ D. M. Armstrong, “Early Causal and Functionalist Views”, *Mind and Cognition*, ed. W. G. Lycan (Oxford: Blackwell 1990) 40.

¹¹⁵ *ibid.*

¹¹⁶ *ibid.*

¹¹⁷ *ibid.*

to give the (rather uninformative) reply: “His purpose to go to the kitchen to get something to eat.”¹¹⁸

Armstrong’s example might seem plausible at first glance because it does, in a sense, point to one use of the word “purpose”, and might be elaborated upon to explain to someone the ordinary use of the word. That is, we might describe a person as having a purpose to go to the kitchen if he behaves in certain characteristic ways, for example, his going there appears direct and intentional rather than accidental, he seems to know what it is that he wants to do there, and so on. But this does not mean that we have formed, or need to form, some sort of hypothesis about what is going on in his brain. As a causal explanation of the concept, Armstrong’s example fails, but it may qualify as a partial explanation of the ordinary use of the word. Of course, to give an explanation will require something different in the context of psychological theory, where it does take the form of a hypothesis; nevertheless, the fact remains that not all genuine explanations of psychological terms come in the form of psychological hypotheses.

My suggestion, then, is that there is a perfectly intelligible sense in which the meaning of “conscious” and other mental predicates are explained by pointing to these phenomena as expressed in behaviour. There are perhaps other intuitions at stake here, though. One is the assumption that what is pointed to when we cite examples of, say, purposive behaviour is not the purpose itself, but, at best, manifestations of it. For one, it would seem a plausible thought – even a truism – that on a strict sense of “see” that we

¹¹⁸ (Hanfling: 2002) 112. We might suppose that the example could be improved by replacing one’s “purpose” as the cause of his going into the kitchen with his “hunger”. This would be a perfectly intelligible explanation of his behaviour, but it does not seem to be the law-like relation between cause and effect that we were looking for.

cannot *see* the emotions of others; the suggestion is, for example, more than a strict empiricist would acknowledge. It is on this narrow definition of “see” that we come to the belief that there is an evidential gulf between, say, pain and pain-behaviour that can only be traversed with the help of an inference. But why should we so narrowly delineate the meaning of the concept here? In ordinary practice we tend to employ the word quite liberally – we tend to think, for example, that there is nothing illicit about saying that we see that someone is feeling ill, or that he is bluffing at poker, or that he is considering the best position in which to hang a painting – and it would seem arbitrary to insist on a specialized use of the term in this case.

There is an analogy to be made here with a piece of art. Ordinarily, we do not speak of inferring from the physical marks on a paper that a painting is of, say, a face. Unless the painting is in an abstract or mannerist style, we simply *see* it. In the other, less straightforward cases, we may need to learn to see how the drawing can be seen as a face – we might have someone point out to us, for example, where the different features are – but even then, it would seem unnatural to say that we have *inferred* it. In cases in which no such difficulties are present, however, we are more likely to speak of the material composition of a painting as expressing a human form, in the same way that we speak of a certain emotion being expressed by a melody. Analogously, we speak of someone’s physical behaviour as expressive of her mood. Unless one’s actions are difficult to read – for example, he is a notorious liar or just very reserved – we do not speak of his mental state as something behind his behaviour, as, to use Armstrong’s words, “that, whatever it

is, which produces certain effects.”¹¹⁹ Now if one’s behaviour is expressive of consciousness in this sense, then it would seem an arbitrary and misleading stipulation to insist that this consciousness cannot itself be directly perceived. And if a person’s mental state is indeed there to be seen, then an ostensive definition of “conscious” does not merely point to behaviour but to the consciousness itself, in a way which is no more problematic than an ostensive definition of, say, “blue”, which points to the sky on a clear day.

Of course, this does not mean that we see someone’s mood in the same sense that we see, say, the colour of her sweater; that is to say, we can distinguish two different uses of the word “see” here. The difference is revealed in the different ways in which the content of what is seen is under each use is expressed. When someone speaks of seeing an object, a copy or reproduction of it will usually be enough to illustrate what is seen. In the case of an angry face, however, one might have to go on to direct another to the shape of the mouth, the tension in the jaw, and so on. One intuitive reason for thinking that someone’s mood cannot be seen is just that it cannot be seen in the same sense as the colour of her sweater. If one assumes that there cannot be two genuine senses of a word, then one must conclude that the second sense – seeing that someone is, say, angry – cannot be a genuine use of the word; however, a simple difference in sense does not entail that one is illegitimate. Some further argument would be needed to establish this.

In what follows, I will examine two reasons for wanting to distinguish mental states themselves from the behaviour that typically accompanies them or from the

¹¹⁹ D. M. Armstrong, (1990: 40).

contexts within which the term is employed. First, we might think that mental predicates just refer to certain characteristic behaviours. Second, we might assume that ascriptions of these terms refer to inner or mental states which are caused by certain things (stimuli) and cause other things (behaviour). The first can be identified as a roughly behaviourist, and the latter a roughly functionalist, approach to the analysis of psychological concepts. Both make problematic the claim that pointing to one's behaviour could serve as an ostensive definition of the concept of consciousness. Using Wittgenstein's favoured example of "pain", I will outline the ways in which these two theories conceive of the relationship between pain and pain-behaviour. In the case of behaviourism, I will argue that the counter-intuitiveness of the theory has to do with the way that it renders the expressiveness of human behaviour unintelligible. This section will thus also serve as a critique of behaviourist interpretations of Wittgenstein. In the case of functionalism, I will argue that insofar as the theory conceives of our everyday psychological ascriptions as performing a primarily explanatory function, it is also a distortion of our ordinary psychological concepts.

3.4. Behaviourism

The assertion that an ostensive definition of a psychological concept which points to behaviour counts as a genuine explanation of the meaning of the concept might be taken to suggest a commitment to logical behaviourism, since it implies that there is a logical connection of sorts between, say, pain and pain-behaviour. Wittgenstein's insistence that "an inner process stands in need of outward criteria" (PI 580), and his

frequent mentions of the body in discussions of mental phenomena are typically taken to imply such a connection, but there is little agreement about how it is to be understood. To some, it is simply the logical behaviourist idea that statements which appear to be about mental states are actually about, in the sense of being logically equivalent to, statements about behaviour or dispositions to behave in certain ways.¹²⁰ If this is the case, then we can simply sidestep the whole complex of mind-body problems by abandoning all talk of the inner; however, the problems with giving behavioural definitions to mental predicates are well-known. Quite simply, it is an affront to common sense to suggest that there is no distinction between statements about pain and statements about pain-behaviour. It does not seem plausible to think that a particular mental state *entails* a particular corresponding behaviour, or even a disposition toward a certain behaviour in any way that could support a definition of that predicate. There may be cases where all the behavioural criteria for asserting that someone is in pain are met, but he is not in pain; perhaps he is acting, or angry, or moved. “Pain”, it is clear, does not just mean a kind of behaviour. But neither, I have suggested, does it refer to something hidden behind this behaviour.

The felt need of the behaviourist to articulate entailment relations between expressions referring to pain and expressions referring to pain-behaviour arises from a particular, problematic conception of behaviour. As mentioned earlier, Descartes’ conception of the body as essentially a complicated machine has been left intact; the behaviourist has only “exorcized the ghost from the machine.” Like Cartesianism, then,

¹²⁰ See, for example, C. S. Chihara and J. A. Fodor, “Operationalism and Ordinary Language: A Critique of Wittgenstein”, *American Philosophical Quarterly* 2 (1965) 292.

behaviourism overlooks the *expressiveness* of human behaviour, which it construes for the most part as mere physical movements and noises.¹²¹ If we are to avoid the route of circuitous inference from behaviour to internal state, then, the conceptual relation between pain and pain-behaviour must be one of entailment. As we have seen, however, this is plainly absurd. There is something at once more arbitrary, and at the same time, closer conceptually, about the “relation” between the sense in which someone is in pain and the way that his face shows his pain, than the suggestion that one proposition entails another analytically. A characteristic pain-behaviour – say, a facial grimace – is more than simply a particular muscular configuration. Pain is expressed *in* the face; it is not *inferred* from the face’s appearance. In that sense, pain is not something entailed by the physical arrangement of the face. The connection is closer conceptually than entailment in the sense that each of these elements contributes to the semantics of the other. This is most obvious in the case of behaviour which cannot be identified independently of its being identified as behaviour *of*, for example, of someone who is hesitant or nervous. There are many examples in which the behaviour seems impossible to specify as anything other than expressive of mentality.¹²²

At the same time, however, it is not as if the feeling of pain could not have been expressed differently, or that given different circumstances the same physical shape could not have been the expression of a different emotional state, say, one of confusion, or intense joy. The connection is looser than entailment, then, in the sense that we can

¹²¹ P. M. S. Hacker makes a similar point in his discussion of logical behaviourism in *Wittgenstein’s Place in Twentieth-Century Analytic Philosophy* (Malden: Blackwell Publishers, 1996) 253-55.

¹²² Try, for example, to explain rule-following behaviour in purely mechanical terms.

imagine that the same psychological state having different behavioural expressions in different circumstances. Wittgenstein constantly emphasizes the fact that psychological concepts are applied within a certain context – not just to a given type of behaviour but to that behaviour seen in relation to what occasions it and what follows from it. Human behaviour is not invariant: “[N]either our circumstances nor our reactions to those circumstances are so inflexible as to give point to a conceptual structure which yoked such elements together as a matter of unvarying necessity; and different cultures will shape and give prominence to certain patterns in different ways.”¹²³ The logical behaviourist goes astray in thinking that mental states can be identified with behaviour; but when a given behaviour is pointed to as a criterion for a psychological state – for example, that of being in pain – the behaviour is pointed to as being *expressive of* suffering. The suffering is not the behaviour itself, but that which the behaviour expresses.

Nevertheless, there might be some truth salvageable from the behaviourist framework: namely, the suggestion that there is a logical connection of sorts between kinds of behaviour and psychological ascriptions. That is, behaviourism has only mischaracterized the logical relation between psychological states and psychological ascriptions; it is not altogether wrong in claiming that there *is* such a relation. As I have argued, the behaviourist misreads the expressiveness of human behaviour by conceiving it in terms of entailment relations between expressions referring to psychological states and expressions referring to behaviour. I have also suggested that this expressiveness,

¹²³ Stephen Mulhall, *Inheritance and Originality: Wittgenstein, Heidegger, Kierkegaard* (Oxford: Clarendon Press, 2001) 173.

properly characterized, can hold together a concept of consciousness that need make no mention of hidden internal causes that would have to be reached by inference. If this is the case, this would support my claim that “consciousness” might be explained via an ostensive definition referring to criteria underlying ordinary psychological ascriptions.

3. 5. *Functionalism*

Functionalism has replaced behaviorism as perhaps the most popularly held position in the philosophy of mind. And with good reason: it offers a nonreductionist account of the meanings of psychological concepts that sits well with contemporary thinking in psychology and cognitive science,¹²⁴ and is able to accommodate certain ordinary intuitions about what psychological states are which behaviourism, with its verificationist commitments, cannot. In characterizing psychological states as functional states, it conceives of them in terms of serving a causal role mediating stimulus-inputs and behavioural-outputs. On the one hand, this conception supports the natural idea that psychological ascriptions refer to more than mere behaviour – they refer to the internal states of the organism – and so captures the idea that there is an important difference between a person pretending to be in pain and one behaving the same way but with the actual sensation. On the other hand, the functionalist conception supports the realist idea that psychological states have genuinely causal properties, such that we are justified in

¹²⁴ In particular, functionalism offers a way of undercutting the reductionist view of mind-brain identity theories or type physicalism (the thesis that there are causal-nomological laws connecting mental kinds with physical kinds) that is hard to reconcile with the idea of computational structure employed in cognitive and computer science. If mental states are functional states realized in some physical structure, there is no theoretical pressure to find some exact neurological characterization of that state. See Jaegwon Kim’s discussion in Chapter One of his *Mind in a Physical World* (Cambridge: MIT Press, 2000).

saying, for example, that the sight of the teacher caused the students to stop shuffling, or that the cause of her quick departure was her desire to make it to the movie theatre in time, and so on.

Despite the greater sophistication of functionalism over behaviourism, however, there are problems with the functionalist conception of behaviour insofar as it is construed as a conceptual analysis of ordinary psychological concepts.¹²⁵ As I will argue, the functionalist account of psychological concepts seems to carry with it the idea that psychological ascriptions play a fundamentally explanatory, or “folk-psychological”, role. The feeling of pain, for example, causes pain-behaviour, but there is no conceptual relationship between the two – only the only the external, contingent one that holds between a cause and its effect. This is to draw a distinction between behaviour and psychological states – in our example, between pain-behaviour and pain – that, I will suggest, distorts the grammar of our psychological concepts.

First, it must be noted that, for all of its sophistication, functionalism can itself be seen as a more advanced form of behaviourism. As Jaegwon Kim writes, “[b]oth functionalism and behaviourism speak of sensory input and behavioural output – or ‘stimulus’ and ‘response’ – as central to the concept of mentality. In this respect, functionalism is part of a broadly behavioural approach to mentality and can be considered a generalized and more sophisticated version of behaviourism.”¹²⁶ Where behaviourism translates the meaning of psychological concepts into stimuli and

¹²⁵ That is to say, I do not want to give the impression that the criticisms I offer of functionalism impinge upon its merits as an empirical theory. It is clear that as a framework for scientific inquiry, functionalism is an improvement over behaviourism.

¹²⁶ Jaegwon Kim, *Philosophy of Mind*, (Boulder: Westview Press Inc., 1996) 77.

behavioural expressions, functionalism translates their meaning into internally-mediated causal connections between stimuli and behaviour. Thus the stimulus-response connections defining the content of psychological terms in functionalism are not available for direct observation, as they are in behaviourism. We can see, then, how functionalism has a greater explanatory power; for what behaviourism takes to be more or less a brute fact – a stimulus-response connection – is for functionalism capable of being given a more detailed explanation in terms of internal states.

This way of conceiving of psychological concepts is tied up with the idea that ordinary psychological ascriptions serve an essentially *explanatory* function, as part of a folk psychology. To speak of someone being in pain, for example, is to attribute to him a particular internal state with certain causal powers that, as part of a loosely articulated background theory of human behaviour, accounts for the observable evidence – his behaviour – and yields predictions about future behaviour. To simply state that a person is in pain is to offer a highly abbreviated description of these causal conditions, insofar as the concept of pain is taken to refer to this process. As Charles Chihara and Jerry Fodor put it, “in learning a language, we develop a number of intricately interrelated ‘mental concepts’ which we use in dealing with, coming to terms with, understanding, explaining, interpreting, etc. the behaviour of other human beings (as well as our own).”¹²⁷ When we say that another person is conscious, for example, we are attempting to *explain* her behaviour. As a crude, tacit hypothesis, it goes without saying that our pre-theoretical

¹²⁷ C. S. Chihara and J. A. Fodor, “Operationalism and Ordinary Language: A Critique of Wittgenstein”, *American Philosophical Quarterly* 2 (1965) 292.

understanding of its content is superficial at best, and stands in need of a detailed reconstruction on the basis of empirical research and conceptual clarification.

While no doubt helpful as an empirical theory, I would argue that this view of psychological concepts is not a very accurate characterization of our ordinary conceptual practices. First, there are good reasons to doubt that when we ascribe a psychological term to someone, we are (tacitly) attributing to him his being in a certain psychological state. The first is that this suggestion, too, misreads the expressiveness of behaviour. If I can be said to just *see* how someone is feeling, then to speak of an explanatory hypothesis being formed here is as out of place as it would be if I were to speak of forming the hypothesis that the sky outside my window, on a clear afternoon, is blue. Moreover, it runs counter to my above suggestion that certain behavioural phenomena serve as criteria for the application of concepts. This involved the claim that behaviour typically illustrating, for example, someone's being conscious, does not provide inferential grounds for a hypothesis about the causes of her behaviour, but a logical *criterion* specifying just what counts as being conscious.

An analogy might be drawn here with Wittgenstein's example of looking at a drawing in projective geometry. He asks the question: What do we *mean* when we speak of someone being able or not being able to read a drawing three-dimensionally, that is, as being able to *see* the three-dimensional figure rather than just knowing that it is a drawing of one? His answer is that to speak of someone's being able to see the drawing three-dimensionally means in this example that they have a certain facility with the drawings, for example, that they can explain them well. In the rule-following remarks, we saw in

chapter one, Wittgenstein entertained and ultimately rejected another, common sense answer that this “knowing one’s way around” is not a criterion of understanding, but rather a symptom of it. This involved picturing understanding as a mental state that is the source of correct use, “as if the pupil’s ability to apply the rule were something he derived from his understanding in the same way we can be said to derive a series from its algebraic formula.”¹²⁸ The problem that we found with this suggestion was that a formula, like a picture, can be variously interpreted. Of course, we often do experience feelings like a sudden dawning of understanding, and their occurrence may well be involved in our claim to understand; however, there is no single, unique experience that is common to all cases of sudden understanding, and no such experience guarantees that I have actually understood in any particular case. To say that the use of “see” here *refers* to a state of the brain that is the cause of the behaviour (or, in the case of behaviourism, to the behaviour itself) is to neglect the contributions of many other factors to the meaning of the word, for example, the context, and the difference between seeing and knowing in such a context. These elements contribute in the sense that they form the setting in which the word functions, like a piece on a chess board which is moved according to the rules of the game. To say that the behaviour of the person looking at the drawing is a criterion for the concept of being able to see in three dimensions just means that the behaviour plays a role in fixing what the word “see” means on that occasion.

The second reason we might have for doubting that functionalism accurately portrays our ordinary conceptual practices recalls Wittgenstein’s discussion of the

¹²⁸ (Mulhall: 2001) 103.

primary colours in chapter two. The lesson of this discussion was that it is futile, if not nonsensical, to try to justify a concept's criteria by pointing to something which confirms it. While a hypothesis about causal conditions for consciousness might be right or wrong – with respect to the accuracy of its observations, the validity of its reasoning, the success of its predictions, and so on – criteria for the meaning of a concept cannot be true or false: they are constitutive of a concept having a truth-value in the first place. Apart from its various criteria for use, there is no concept to be independently specified, and so there is no use in pointing to those same criteria in an attempt to confirm the truth or falsity of the concept. When we learn the meanings of words like “conscious”, “pain”, and so on, we do so with reference to human beings and other animals that behave in relevant ways. To then worry whether these behaviours qualify as cases of consciousness or of pain is analogous to worrying, as I put it in chapter two, whether it is *really* Thursday, or whether the Standard Metre is *really* one metre long. Now if what is taken as data for a hypothesis is in fact a criterion for ascribing a certain mental concept to someone, in the sense that it fixes the meaning of that ascription, then the mistake lies in trying to explain nomologically a correlation between two things which is purely conventional or arbitrary.¹²⁹

Returning now to my original suggestion that our ordinary employment of psychological ascriptions does not stand in need of justification, we might now consider in more detail how this critique of functionalism ties in to my claim that is no inherent evidentiary gap between mental states and the behaviour that expresses them. To start,

¹²⁹ P. M. S. Hacker criticizes folk psychology along these lines in his “Methodology in Philosophical Psychology,” *Wittgenstein: Mind and Will* (Malden: Basil Blackwell, 2000) 119-128.

we might consider what it would mean to claim that someone is justified in her use of a word. If we construe characteristically human behaviour as playing a criterial role in fixing the semantics of the word “conscious”, then certainly pointing to those criteria is a perfectly direct way of justifying one’s use of the concept, and thus a perfectly direct way of justifying what “consciousness” means. And this means: there is no point in asking for a deeper justification. What this suggests for our present discussion is that an ascription of consciousness to another should not be construed as resting on explanatory hypotheses. Hypotheses can be true or false, more or less accurate explanations of the phenomena, but statements asserting a criterial connection cannot be true or false. They are among the conditions for a sentence’s having a truth-value in the first place. To state that the assumption that our ordinary understanding of consciousness, as embodied in our everyday talk of how we feel, begs the question as to whether we are correct in this understanding would be akin to opening the same sort of discussion about the Standard Metre, or about any other conventionally determined connection. Why, we should then ask, are we justified in calling *that* the Standard Metre, or *that* Thursday? It sounds odd to speak of being either justified or to stand in need of justification here.

Next, we might consider what it means to understand the meaning of a concept. If the ordinary English speaker’s understanding of what it means to be conscious does not take the form of a rudimentary theory of human behaviour, then there is no theoretical object – for example, no set of truth-conditions to be grasped or causal process to be explained – involved in learning what “conscious” means. This assumption – that psychological ascriptions involve an inference from behaviour to the causes of behaviour

– is, we have seen, a deep-seated assumption in the philosophy of mind, one which has set the stage for much of the drama that has unfolded since Descartes' bifurcation of mind and body. But if we look at how we actually use the concept of knowing what a word means, we see that that linguistic competence is not a *knowing that*, but a *knowing how*, or as Wittgenstein puts it, a "mastery of a technique" (PI 150), like learning how to play chess or ride a bicycle. This is what Wittgenstein's insistence that "meaning is use" really comes down to – it is not a theory about what meaning essentially is, but a reminder of the ways in which we ascribe mastery of a concept to a person. Understanding is something that is manifested in a practice, in using words appropriately within a certain context. Whether or not one has really understood something is not a question she can answer by introspection – say, by checking that she is in contact with some autonomous "meaning" – but by participating in the relevant practice. So if one can, in fact, go on to use the mental concepts appropriately, then she has a perfectly good understanding of what they mean.

Last, we might consider what it means to explain the meaning of a word. If explanation is an attempt to give someone else an understanding of something, then ordinary explanations of "conscious", as rules for the correct (i.e. grammatical) use of the word are genuinely explanatory. That is, they lead to understanding – understanding as knowing how to go on. Such explanations are akin to explanations of how a particular piece functions in a game of chess; for example, "The rook begins the game here and can move in only straight lines and only over unoccupied squares..." and so on. One reason that these kinds of explanations might appear inauthentic or superficial is that they are

context-specific. Depending on to whom I am speaking, on how their prior knowledge of the game, my explanation of how a rook works will be long or short, detailed or brief. In PI 31, for example, Wittgenstein describes three ways in which a person might learn that a certain piece in a game of chess is called the “king”. In the first scenario, a person already knows the rules of the game, including the role of each piece; the only matter that remains to be settled is which one functions as the king. In this case, saying “This is the king”, while pointing to the piece will be enough to secure the meaning of the word for the learner. This is akin to offering an explanation of “conscious” to someone who already speaks a natural language, but, say, does not know the English word for it. In the second scenario, a person who has mastered the game without explicitly learning the rules is shown pieces of unfamiliar shapes. In this case, too, saying “This is the king” will normally suffice to convey which pieces are kings, because “the place for [the word] was already prepared” (PI 31). The person does not know the rules of the game in the sense of being able to list them; however, he does know “in another sense,” since he knows how to proceed (PI 31). This example perhaps comes closest to describing our knowledge of our native language; while most of us can distinguish a grammatical sentence from an ungrammatical one, few of us can express the rules in any detail. In the last scenario, a person is told “‘This is the king,’ it can move like this, ... and so on” (PI 31). If the person “knows what a piece in a game is,” having played or watched others playing before “‘and understood’ – *and similar things*” (PI 31), this will suffice to fix the meaning of ‘king’. As in the first two scenarios, a place for the word has been cleared; the student knows enough for the new information given by his teacher to fix the

meaning of the word. Only under such conditions – when one “already knows how to do something with it” (PI 31) – can one significantly ask (or give) a name.

A theory like functionalism is almost irresistible conceptually if one assumes that psychological ascriptions are statements about inner mental entities, or causal connections between stimuli and behaviour that are mediated by inner psychological states. It might be, however, that this view rests upon a particular picture of how language works – in particular, what the meaning of a word is – which is inappropriate in this area of our lives. According to this picture, the primary function of language is to *describe* or *represent*. Functionalism seems a natural choice for someone who wants to give a nonreductive account of the mind, but who also accepts this referential picture of language. In attacking this picture throughout the *Investigations*, Wittgenstein does not suggest that it is never appropriate to think of words as designative but only that we must fight the tendency to think of descriptive uses of language as paradigmatic of language use in general. There are other uses of words which are less formal or rationalistic than the activity of describing an object.

Think of exclamations alone, with their completely different functions.

Water!
Away!
Ow!
Help!
Fine!
No!

Are you inclined still to call these words “names of objects”? (PI 27).

In the case of the concept of pain, in particular, there is an intuitive distinction between *descriptive* and *expressive* language that I hope will illuminate his general approach. In

speaking of pain – crying “Ouch!”, for example – it would be a distortion to say that I am describing some internal state or voicing some proto-hypothesis about causal connections. This is true of third-personal ascriptions as well. Depending on the occasion, uttering the expression “He is in serious pain” could mean an indefinite variety of things; for example, that we had better try some different medication, that he is not enjoying the game anymore, that we ought to send a card or gift, and so on. In the case of ascriptions of consciousness, this implies that understanding the semantics of psychological concepts requires that we look at the functions such ascriptions play in our lives. The lesson here is not that our language can never be used to describe what another is feeling, but that expression is not necessarily semantically or epistemologically derivative from description.

If it is true that our ordinary psychological ascriptions are better characterized as expressive than as descriptive, this will carry with it the meta-theoretical consequence that, in an important sense, empirical theories of consciousness do not, and in fact cannot, impugn this kind of understanding as rudimentary, because there are no hypotheses to be revised or undermined. This, I take it, is what motivates Wittgenstein’s notorious dismissal of psychology in Part II, section xiv of the *Investigations*:

The confusion and barrenness of psychology is not to be explained by calling it a “young science”; its state is not comparable with that of physics, for instance, in its beginnings. (Rather with that of certain branches of mathematics. Set theory.) For in psychology there are experimental methods and conceptual confusion. (As in the other case conceptual confusion and methods of proof.)

The existence of the experimental method makes us think we have the means of solving the problems which trouble us; though problem and method pass one another by.

The existence of an experimental method, that is, makes us think that the analysis of mental concepts is a causal one, when in fact it is conceptual. If, as I suggested in chapter two, concepts like “mind” and “soul” arise in our spontaneous, pre-reflective engagement with one another – rather than our introspective awareness of private mental objects – then a causal analysis of these concepts does not *explain* them but, at most, recommend new criteria for how to use the word. While scientists studying the physical world can and do make discoveries which shatter our pre-scientific cosmological guesswork, neurobiology (or any other sort of cognitive scientific discipline) does not shatter our ordinary understanding of consciousness in any straightforward sense. This is not to say that developments in empirical psychology *never* lead us to see the grammar of our concepts differently; they might, for example, help to undermine any superstitions that had arisen when that phenomena was shrouded in mystery. Of course, home-grown *theories* of consciousness can be, and usually are, corrected by experimental psychology; there are cases in which our beliefs about the world change, causing us to renegotiate the boundaries of our concepts. In that sense, a bit of rudimentary theory can be said to underlie a linguistic practice. But conceding that does not mean conceding that *all* linguistic practice is theory-laden.

One might object here that the functionalist conception of psychological states need not be construed as a conceptual analysis of what we *mean* in ascribing a psychological concept to someone, but rather as a scientific hypothesis. This would be one way of exonerating functionalism by distinguishing it from ordinary psychological language. However, even if one interprets functionalism as an empirical theory, the

assumption is still made that ordinary psychological ascriptions are part of a folk psychology whose business it is, if not to formulate law-based causal explanations, then to predict or rationalize human behaviour on a tacit and inchoate understanding of causal connections between stimuli and behavioural outputs. This, I have tried to show, is not an accurate representation of our everyday talk of our mental lives.

3. 6. *Conclusion*

I want to conclude this chapter by considering a thought experiment that will at the same time address the claim that inner states bear a logical connection to outer behaviour and further elucidate the critical force that Wittgenstein's claims about the arbitrariness of grammar have for the philosophy of mind. A popular argument against functionalism is that of "absent qualia" – of beings who are functionally like us, but whose behaviour is not accompanied by subjective experience. According to Frank Jackson, for example, we are supposed to be able to imagine a world in which there are "organisms exactly like us in every physical respect ... but which differ from us profoundly in that they have no conscious mental life at all."¹³⁰ If indeed mind and body bear only a contingent, causal relationship to one another, then imagining one without the other should be relatively easy; we simply give that behaviour a different cause – say, a computer program. If, on the other hand, there is a logical connection between mind and body, then imagining this should be much more difficult. Wittgenstein challenges us to try to do so in the following remark:

¹³⁰ Frank Jackson, "Epiphenomenal Qualia", *Mind and Cognition*, ed. William Lycan (Malden: Blackwell Publishing, 1990) 472.

But can't I imagine that the people around me are automata, lack consciousness, even though they behave in the same way as usual? – If I imagine it now – alone in my room – I see people with fixed looks (as in a trance) going about their business – the idea is perhaps a little uncanny. But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: 'The children over there are mere automata; all their liveliness is mere automatism.' And you will either find these words becoming quite meaningless; or you will produce in yourself some kind of uncanny feeling, or something of the sort (PI 420).

I suspect that Wittgenstein is right; at the very most, this is something we might manage to do just momentarily. For could we really doubt that beings who display the same fine shades of behaviour that we humans do – beings who play together, who learn and teach a language, who look in mirrors, who writhe and cry and call for help when injured, who write letters and save them, who beg for their lives, who mourn, who paint pictures and read philosophy – are not conscious, even if they are physiologically dissimilar? Confronted with these creatures, we might very well redraw the boundaries of our concepts to exclude them, but such a question would not be for science to decide. A choice will have been made, one that cannot simply be read off the world. In this sense, it could only be called *ethical*.

Conclusion

What contemporary, analytic philosophers of mind – be they physicalists or anti-physicalists – share in common is an eagerness to think of the scientific description of the world as metaphysically fundamental. Science, for them, reveals the world as it fundamentally is. Since anti-physicalists are – justifiably – reluctant to endorse a radical physicalism which refuses to admit the ultimate reality of conscious experience, they are left with the feeling that there is something deeply mysterious about consciousness. In chapter one, I tried to show that the scientific worldview is just one way, amongst others, of looking at the world. The idea that science is not just a useful human activity which enables us to make better sense of certain phenomena but also the last word in what is actually *real* is nothing but a modern superstition. In our ordinary encounter with the world, wholesale skeptical doubt does not occur to us. Thus, rather than thinking of skepticism as the inevitable product of deep thinking, we might take its appearance in philosophy as a sign that perhaps we have stepped *too* far back, turned the detached gaze of science onto ourselves, and thereby misrepresented some aspect of our lives. In chapter two, I suggested that Descartes did such a thing when he characterized the body as essentially a complicated machine with a hidden soul inside. This picture of the body as expressively dead, and of the mind as something “behind” behaviour, we saw in chapter three, is still standard in contemporary philosophy of mind. It should not be surprising, then, that mind-body problem that occupied Descartes – the problem of fitting our ordinary experience as embodied agents into scientific accounts of the human body –