Wittgenstein, Criteria and the Ineffability of Consciousness

A Thesis Submitted for the BPhil in Philosophy

Candidate 275898

29,452 Words

Candidate 275898

Abstract

The present discussion seeks to determine the scope and nature of Wittgenstein's opposition to some prominent approaches to consciousness in contemporary philosophy of mind and cognitive science. We begin with some general remarks about this opposition, and in Chapter 1 appraise some prominent contemporary exemplars: Hacker & Bennett's (2003, 2007) Wittgensteinian critiques of attempts to study and define consciousness, and Searle (2007) and Block's (2007) opposing views. Their dispute hinges on whether Wittgenstein's notion of criteria yields a *prima facie* opposition to the distinction between consciousness and language (the 'being'/'saying' distinction) and more generally whether his later view is opposed to the ineffability of consciousness. We will demonstrate that while Wittgenstein's notion of criteria, properly understood, does provide a *prima facie* case against the 'being'/'saying' distinction (Chapters 2 and 3), there is an alternative sense in which Wittgenstein's view may endorse a notion of 'ineffable consciousness' importantly similar to that endorsed by Block (2007) and others. As such, Wittgenstein's position concurs *neither* with attempts by Block, Searle and others to distinguish consciousness from language, *nor* with claims such as Hacker and Bennett's which suggest a wholesale opposition to consciousness or its ineffability.

Contents

Introduction

Contemporary philosophy of cognitive science is an exceptionally active and promising field within philosophy. For its proponents, it provides a framework within which philosophical discourse and contemporary scientific research can provide a mutually supporting basis for the resolution of pressing questions about the nature of the mind and its relationship to body and world. The most intransigent and puzzling of such questions concern the nature of consciousness. The strictures of materialism compel an insistence that the brain must be responsible for our mental lives; yet there is a pervasive intuition that consciousness – that richly nuanced 'phenomenal' aspect of our subjective mental lives that constitutes its being ‘like’ something to be us (Nagel 1974) – is profoundly distinct from the inanimate physical world. As Colin McGinn quips, “how can technicolour phenomenology arise from soggy grey matter?” (1989:438)

The nature of consciousness is the axis for a great bulk of discourse in contemporary cognitive science and philosophy of mind. Though rooted in Descartes and in the Platonic taxonomy he masterfully elaborated (see Hacker 2007b), this discourse found its present incarnation largely in the reaction against the stringent reductive paradigms of logical positivism and behaviourism, and to an extent against the ordinary language philosophy of the mid-20th century. This reaction may be characterised as, first, a rejection of the notion that the mysteriousness of our conscious lives is an illusion produced by the misuse, and in particular the reification, of our everyday psychological concepts; and second, an attempt to reanimate the idea that consciousness is a genuine, deep and as yet poorly understood mystery.

The aspiration to a genuine solution, rather than a dissolution achieved through conceptual analysis, implies a commitment to a distinction between language and ontology: between our concepts on the one hand, and the objects, states and processes – and facts about these – to which they refer, on the other. More specifically, the search for a solution to the problem of consciousness can be linked with an attempt to deal with consciousness the entity as *distinct* from 'consciousness' the concept; and such an enterprise assumes not just that such an entity exists, but that there exists a space between it and the ways we talk about it: between '*being*' and '*saying*'. D.M. Armstrong, for example, writes approvingly of the departure from "Wittgensteinian and even Rylean pessimism", which offered "nothing beyond an analysis of the various mental *concepts*", towards the view that "the central task of philosophy is to give an account...of the most general nature of *things*" (1980:38, my italics). We will refer to this rather diffuse viewpoint as 'cognitive philosophy', or 'cognitivism' for short[1]; and to its widely varied adherents as 'cognitivists'.

Our characterisation of this divide goes some way to explaining a curious feature of the 'cognitivist' approach: its under-appreciation, and indeed routine ignorance, of Ludwig Wittgenstein. Wittgenstein may be regarded as one of the 20th century's most innovative and influential thinkers in philosophy of mind and psychology. His writings on sensation and experience, psychological language-use and the relationship between mind and body constitute a profound and trenchant assault on the Cartesian taxonomy of mind and body, and a novel understanding of our capacity to define and describe our own and one another's conscious experience. Yet for all its crucial (if perhaps unclear) bearing on cognitive

1 Not to be confused with cognitivism as it is contrasted with non-cognitivism in other philosophical discussions.

philosophy (not least the attempt to come to grips with consciousness), Wittgenstein's voice has been largely under-represented in the rise of contemporary cognitive science.

This has not gone unnoticed by Wittgensteinians: Severin Schroeder, for instance, laments that communication between “those with a strong interest in Wittgenstein, on the one hand, and leading philosophers of mind, on the other, seems to have all but broken down” (2001:vii). Schroeder quotes Anthony Kenny's remarks on developments in philosophy of mind in the 1980s: “some of the philosophical gains we owe to Wittgenstein seem in danger of being lost...not because his work has been superseded...Rather, his contribution has been neglected because more and more philosophers...have attempted to model their studies on the pattern of a rigorously scientific discipline” (Kenny 1984:vii-viii). And similarly, Oswald Hanfling complains of the “almost total disregard” (in Schroeder 2001:58) with which Wittgenstein’s work is treated in mainstream philosophy of mind and cognitive science. If the 'cognitive turn' is indeed a product of the dissatisfaction with conceptual analysis with which Wittgenstein's name is chiefly associated, and of an aspiration to unite philosophy and empirical research to which Wittgenstein certainly seems hostile, then this attitude is, if not warranted, at least unsurprising.

Correspondingly, Wittgensteinian scholars are themselves often dismissive of contemporary trends and advances in philosophy of mind and cognitive science. For many of them, the methods and ambitions of 'cognitivists' seem so enmeshed with contemporary science as to be unphilosophical, and so fundamentally awry that they are beyond meaningful criticism or engagement. The result is that, with sporadic exceptions, there is not so much a dispute between Wittgensteinians and the ‘cognitivists’ as there is estrangement and silence.

Whatever its causes, the consequences of this estrangement have been damaging. The failure of some Wittgensteinian scholars to engage with the complexities of contemporary philosophy of mind, and their sometimes monolithic and hyperbolic dismissal of this discipline, deprives their insights of credibility and traction. On the other hand, the association of Wittgenstein's views with the antiquated paradigms of logical positivism and behaviourism means that many of the crucial and persuasive subtleties of his account are passed over by those within mind and cognitive science who, deeply engaged with the very questions Wittgenstein addressed, are best placed to learn from and implement them.

Our broad ambition with this discussion is to contribute to closing this disconnect. Wittgenstein's later work, and contemporary philosophy of mind and cognitive science, are enormous and moving targets, and there is much we cannot discuss here. It particularly needs to be stressed that our concern is not with *adjudicating* the dispute between Wittgenstein and the cognitivists, or espousing one view against the other. For as we have pointed out, there is hardly an active and accurate dispute to adjudicate here; there's mostly silence, and where it is broken, little more than wholesale disagreement or the fruitless exchange of orthogonal views. Rather, our purpose, and our thesis, is of a more fundamental and exegetical kind: we want to reappraise the bearing of Wittgenstein's position, and specifically his notion of criteria, on approaches to the definition and description of consciousness in contemporary cognitive science.

To that end, we will begin in Chapter One by looking at some recent and characteristic 'shots across the bow' between Wittgensteinians and cognitive scientists; in particular, Peter Hacker and Maxwell Bennett's (2003, 2007) critiques of attempts to study and define consciousness, and some replies from John Searle (in Hacker & Bennett, 2007)

and Ned Block (2007). As will become clear, both Block and Searle rely on the cognitivist's distinction between 'being' and 'saying'; and their dispute with the Wittgensteinians hangs on whether Wittgenstein's later view yields a notion of criteria which is intended to, and can, provide a *prima facie* opposition to this distinction.

Accordingly, Chapters Two and Three will be concerned with this question. In Chapter Two we will look at three types of approaches to Wittgenstein's notion of criteria, and argue that what we will call the 'inclusive' view is the most exegetically plausible, and makes good sense of the intuitions of the other two approaches. Chapter Three will show how the 'inclusive' view supports an opposition to the distinction between 'being' and 'saying', and raise and rebut some objections to it.

Finally, in Chapter Four, we will look at another sense in which, notwithstanding his rejection of the distinction between 'being' and 'saying', Wittgenstein endorses a kind of 'ineffable consciousness' which is significantly similar to conceptions of qualitative consciousness endorsed by Block and others. To that extent, our examination will yield the surprising conclusion that Wittgenstein's view, even – and in fact, especially – when taken at its strongest, provides not a wholesale opposition to prominent accounts of the nature of consciousness, but a nuanced support for some of their conclusions. At the very least, it is hoped we might demonstrate that the dispute between Wittgenstein and contemporary cognitive science is a rich and productive one, and not so straightforward as to warrant only a stony silence.

1 - Being and Saying:

Wittgenstein and the Cognitivists

A recent, direct and comprehensive assault on contemporary philosophy of cognitive science is offered by Peter Hacker and Maxwell Bennett in *The Philosophical Foundations of Neuroscience* (2003) and again, in dialogue with John Searle and Daniel Dennett, in *Neuroscience and Philosophy: Brain, Mind and Language* (2007). Bennett and Hacker's position is an exceptionally strident variant of the Wittgensteinian critique of cognitivism. By dint of its zealousness, it presents many aspects of the Wittgensteinian critique 'writ large', and so, though in some respects atypical, it is a fine catalyst for our discussion.

Hacker and Bennett's critique is essentially an extrapolation from the former's criticisms elsewhere (e.g. Hacker 1986, 1990, 2007b). For present purposes, we can extract two distinct objections from their critique. The first can be called the 'monolith' objection, and summarised as follows:

(i) there is a widespread contention amongst cognitive scientists that there exists a 'monolithic' category of 'consciousness';

(ii) the putative identity condition for this category is 'qualitative feel' or Nagel's 'what it is like'-ness;

(iii) many putative members of this category in fact do not possess this identity condition;

(iv) Therefore, there is no basis for the supposition of a singular, monolithic category of 'consciousness'.

The second objection, which, following Hacker and Bennett, we can call the ‘mereology objection’, may be summarised as follows:

(i) there is a widespread tendency amongst cognitive scientists to predicate psychological attributes of brains;

(ii) the identity conditions for psychological predicates can only be satisfied by entire animate organisms, not their parts; so

(iii) their predication of brains is nonsensical, or logically illicit.

A further step in their argument is, roughly, that:

(iv) The ‘mystery of consciousness’ is an illusion generated by this illicit practice.

1.1 Mereology and Monoliths

These two objections are closely related, and draw support from the same view of Wittgenstein. We will begin with the 'widespread tendencies' that Hacker and Bennett identify. First, the 'monolith' approach. It is commonplace for contemporary cognitivists to deny and denigrate Cartesian tendencies[2]; and most stop short of crude substance dualism, or insist, *pace* Descartes, that the mental is reducible in some sense to the physical. Nonetheless, Hacker and Bennett (see also Hacker 2007b) insist, they have inherited Descartes’ taxonomical framework. And this is evident, as elsewhere, in discussions of consciousness; and in particular, in the contention that ‘consciousness’ is a recognisable and identifying characteristic of a diffuse range of psychological states and processes: “not only perception, sensation and affection, but also desire, thought and belief” (Hacker & Bennett 2003:39).

2 A comprehensive and excellent discussion of this is Dennett's famed rejection of the 'Cartesian theatre': Dennett 1991)

These states and processes are bound together as aspects of our subjective mental life, characterised by their ‘phenomenal character’, what is sometimes called ‘qualia’ or ‘what-it-is-like’-ness. It is *like* something to think, as it is to want, to believe, to perceive, to be saddened or elated by something, and so on. *What* it is like is intrinsic to the identity of each of these states.

Closely related to this is a tendency to attribute psychological predicates to the brain. Hacker and Bennett offer various examples: attention is drawn, for instance, to Frisby’s (2007:28) hypothesis of the brain’s capacity to ‘map’ or ‘symbolise’ the external world, through the semantically sensitive, rule-governed manipulation of data; to the ascription of a capacity for ‘perception’, ‘belief-formation’ and ‘thought’ in the much discussed commissurotomy cases; etc. Hacker and Bennett concede that such discourse might be permissible if it is shown to be metaphorical, or to constitute an analogical extension of the existing concepts, or a circumscribed technical usage with rules of use distinct from ordinary language. But the ascription of psychological predicates to the brain is theoretically interesting and non-trivial precisely because it is intended *not* to be merely metaphorical, stipulative or analogical: what is important about these theories is the claim that neural activity constitutes actual cases of belief-formation, perception, semantically-sensitive symbol manipulation, etc. *as these concepts are normally understood*; and what is more, that there are actual cases of symbols, images and maps in the brain.

Most importantly for our purposes, the brain *itself* is widely said to be conscious: to be the bearer, and the locus, of this 'monolithic' thing, 'consciousness'. That the nature of consciousness presents a mystery is premised, in part, on this contention: “how could the brain be the seat of consciousness?” (Dennett 1991:433); and how is it that consciousness,

this distinctive and ubiquitous 'what-it-is-like-ness' of our mental lives, could just *be* matter configured and behaving according to biochemical principles (e.g. McGinn 1999)? These very propositions have a jarring incongruity; Wittgenstein writes of the “slight giddiness” that they induce (PI[3]§412).

For Hacker and Bennett, such mysteries are of our own making. Their rallying cry is a remark of Wittgenstein’s in §281 of the *Investigations*: “...only of a living human 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”. Hacker writes that this remark “epitomises the conclusions we shall reach in our investigation” (2007:19).

Hacker and Bennett take this remark to express the necessary relationship between behavioural criteria and psychological predicates; and more specifically, the view that we cannot predicate psychological properties of anything less than the entire animate organism. This is so, they claim, because the ascription of psychological predicates is determined by the exhibition and recognition of behavioural criteria which “only the whole animal is capable of exhibiting” (2007:102). For instance, in the case of toothache experienced by a human being, behavioural criteria might include clutching one's cheek, groaning, grimacing, exclaiming 'my tooth is throbbing!', and the performance of a range of other behaviours in circumstances which do not undermine our confidence in the sincerity of these exhibitions. These criteria are said to constitute the rules for usage of these concepts, as well as being “partly constitutive of [the] meaning” (ibid:83) of the concepts themselves.

This observation underpins both the mereology and the monolith objections. Regarding the 'mereology' objection: brains are not capable of exhibiting the behavioural

3 *Philosophical Investigations*

criteria necessary to warrant the ascription of these predicates. So it is not clear what it would *mean* for a brain to, e.g., be 'conscious'; and not clear how an explanation of *how* the brain could be conscious would proceed. We might say that, given that criteria determine cases of consciousness, the question 'how could a brain be conscious?' is like asking how something that is not what we call conscious could be something that we call conscious.

Of course, it *cannot*; but for Hacker and Bennett this is the expression of a *verbal* limit, not an empirically significant one: our concept of 'consciousness' does not apply to cases of brains. The incongruity between consciousness and the brain is genuine, but signals conceptual confusion rather than empirical mystery. Cognitivists are *rightly* puzzled by the ill-fitting imposition of psychological predicates onto the brain – suggestions that my brain 'loves', 'decides' or 'is me' – but fail to recognise that this mystery is induced by predicates that are ordinarily, and sensibly, ascribed to *persons* (living human beings) being instead ascribed to organs which are mere parts of those persons. Their application in this context is not bound by the criteria that constitute settled practices of language-use.

Equally, the tendency to conceive of consciousness as a distinct, 'monolithic' phenomenon or kind of thing is said to result from a failure to attend to the conditions within which we, apparently unproblematically or non-mysteriously, apply psychological predicates on the basis of behavioural criteria. Again, because their use is grounded in behavioural criteria, such predicates are applied to *human beings*, and to other living things capable of exhibiting the same criteria. Facts about consciousness and descriptions of states of consciousness are, ipso facto, facts and descriptions about human beings. The idea of consciousness *itself*, a distinct phenomenon about which there are distinct facts, constitutes an illicit reification of our language-use. Consciousness is not, say Hacker and Bennett (e.g.

2007), a distinct kind of thing, subject to its own description; rather, human beings are the subjects of consciousness.

Hacker and Bennett seek to demonstrate this by insisting that some of the putative members of this kind, ‘consciousness’, cannot be sensibly type identified by qualitative content. Evidently understanding qualia as “affective attitude[s]” (2007:44), they suggest, e.g., that “[s]eeing an ordinary table or chair does not evoke *any* emotional or attitudinal reaction...the experiences differ in so far as their objects differ” (ibid:40); “for a vast range of things that can be called ‘experiences’, there isn’t a ‘way it feels’ to have them, i.e. there is no answer to the question ‘How does it feel to...?” (ibid:41).

In fact, it must be said that this demonstration is in itself very unconvincing. It appears to conflate phenomenal, qualitative, or experiential aspects of our mental lives with attitudes *to* those states. One may indeed regard a table or a lamppost as uninteresting – it may not excite any distinctive affective reaction – but this does not suggest that it is not ‘like something’ to see a table or a lamppost. It seems clear that I have an experience of those objects, composed of various sensory-perceptual modalities, which these objects do not have of me. It is *this* distinction, arguably, that prompts talk of phenomenal experience or 'qualia'.

But this misstep conceals a more lasting and compelling point: that each distinct psychological state is possessed of distinguishing behavioural criteria, and that conglomerating them as states of 'consciousness', rather than of human beings, amounts to ignoring their particular, concrete circumstances of employment and reifying the 'mind'. It is in consequence of *this* that there is said to be no distinct thing called 'consciousness' over and above these distinct states.

So again here, the apparent mystery is of our own making: it is indeed the case that we cannot say what kind of state consciousness is, what its distinct properties are, and so on – but, again, this is a verbal constraint with the *appearance* of a genuine mystery. 'Consciousness' *itself* is not a state or entity possessed of criteria independent of human persons; to fail to find such independent criteria and then declare this to suggest the mysterious nature of consciousness is the cognitivist's mistake.

This is Hacker and Bennett's position. Though as we mentioned it is perhaps more strident in degree than other Wittgensteinian critiques, it is by no means uncharacteristic in kind. Numerous other commentators have made essentially similar criticisms of contemporary philosophy of mind and cognitive science. The best known and most obvious comparison is Anthony Kenny's 'The Homunculus Fallacy', which also begins with PI§281, but takes as its exclusive focus a problem akin to the mereological fallacy, the “reckless application of human-being predicates to insufficiently human-like objects”, (including the brain and the sensory apparatuses) leading to “conceptual and methodological confusion” (1991:155). Norman Malcolm complains in similar terms that contemporary philosophy of mind has “lost sight of *the bearer* of mental predicates”, perpetuating Descartes’ ascription of such predicates to the mind with an idea “equally absurd, namely that the human brain [is] the bearer of mental predicates”, rather than “the corporeal human being” (1984:100). Drawing on Malcolm’s comments, Oswald Hanfling suggests, contrary to Dennett’s talk of ‘conscious brains’, that “[i]t is Dennett, the human being, and not one of his bodily organs, that arrives at conclusions, replies to objections, and so on. The behaviour on which these descriptions are based is that of a ‘living human being and what resembles’ one...these activities cannot be performed by brains any more than by other bodily organs or, for that

matter, sticks and stones" (2001:39); James C. Klagge suggests, similarly, that because psychological predicates are "always answerable to, the original behavioural criteria", our *brains* can only exhibit "the physical correlates of the mental", which are distinct from "our use of mental concepts", and so "can do nothing to...resolve our philosophical perplexities about the mental" (1989:322). And Michael Proudfoot, again on the basis of PI§281, suggests that Wittgenstein's view entails that "we cannot ascribe psychological properties to...souls, brains, or disembodied computer systems", since "there are conceptual connections between psychological phenomena and contingent features of the behaviour and appearance of human beings" (1997:197-8). The monolith objection is implicit too in Malcolm's (1984) suggestion that 'what it is like' questions evince only an odd assortment of characteristic feelings and reactions to something, which do not share any distinctive identity conditions, much less anything "irreducibly subjective" or "mental" (ibid:46-7); and in Hanfling's subsequent suggestion that we should not assume the coherence of a 'general' mystery of consciousness (2001:48; cf Heil 1981).

In the next section, we will look at some replies and alternative views from John Searle and Ned Block.

1.2 Replies from Cognitivism

John Searle (in Hacker & Bennett 2007) offers some remarks in direct reply to Bennett and Hacker, while Block's (2007) claims emerge independently from a consideration of the bearing of Wittgenstein's ideas on Block's own arguments for qualia. Again, these are far from the only responses of their kind (see, e.g., Rey 2003, Budd 1991, Dennett in Hacker

& Bennett 2007); but Block and Searle's positions aptly exhibit the cognitivist's separation between language and ontology that we earlier raised. So it is with them that we will complete our initial picture of the dissonance between Wittgenstein and cognitivist approaches.

Reacting to Bennett and Hacker's first objection, Searle contests what he sees as an inference from the criterial role of behaviour to the rejection of consciousness:

> "just as the old time behaviourists confused the behavioural evidence for mental states with the existence of the mental states themselves, so the Wittgensteinians make a more subtle, but still fundamentally similar, mistake when they confuse the criterial basis for the application of the mental concepts with the mental states themselves...the behavioural criteria for the *ascription* of psychological predicates with the *facts described* by those psychological predicates" (2007:103).

Importantly, Searle thinks this conflation is a misinterpretation of Wittgenstein's own view (2007:103). Wittgenstein, Searle claims, did not wish to deny the distinction between, for instance, pain and pain-behaviour (ibid:102). Rather, Wittgenstein's remarks on criteria stress only the dependence of the former on the latter so far as the development of the language game is concerned: if there were *never* any behavioural criteria of one's inner mental life, we should never in the first place have developed a public language-game concerning our psychological or experiential states. Behaviour, insofar as it is a publicly observable manifestation of our inner states, is what allow our psychological terms to refer to inner states. But although it establishes the circumstances of ascription for our psychological concepts, it does not wholly constitute the meanings of these concepts. It is therefore possible, in individual cases and *against the background* of an ordinary correspondence between experience and expression, to distinguish the experience from the expression. So, he writes that, "even if the Wittgensteinian approach is 100 percent correct as a philosophical

analysis of the operation of the vocabulary", nonetheless we may "carve off, in any individual case, the existence of the inner, qualitative, subjective feeling from its manifestation in external behaviour" (ibid:103).

And indeed, it is Searle's point that we *must* do so, if we want to distinguish between mere conceptual analysis and an understanding of the real phenomena of conscious states and processes, which are not grammatical fictions, but exist in space and time (2007:103), distinct from the language-games by which we refer to and describe them. Searle's reply therefore expresses the cognitivist's distinction between 'being' (states, processes, objects and facts about them) and 'saying' (language).

Introducing a distinction between *what* we mean and *how* we mean it goes some way towards vindicating both the practice of talking about consciousness as a distinct kind of thing, and the practice of attributing this thing to *brains*: the two tendencies to which Bennett and Hacker object. If we can distinguish between the conditions of ascription and the objects of ascription, we can concede that the conditions that frame our language-use rest on behavioural criteria while insisting that what is referred to *by* our words *in virtue* of this framework is independent of any such criteria. From here, we can go on to claim that this referent is intimately connected – both spatio-temporally, and causally – with the brain, and that understanding its nature is a genuine task which is currently incomplete. So, we might say, when we attribute consciousness to the brain, we are merely 'carving off' the real, neurally-located phenomenon expressed in our behaviours from those behaviours. And when we speak of consciousness as distinct from unconsciousness, we are speaking of the generation of this distinct phenomenon by the brain.

It is important that Searle imputes this distinction to Wittgenstein himself. And indeed, there appear to be at least reasonable *prima facie* grounds for his doing so. For instance, when the imagined interlocutor of the *Investigations* demands that Wittgenstein acknowledge the distinction between pain-behaviour accompanied by the sensation of pain, and pain-behaviour without it, Wittgenstein readily concedes, “what greater difference could there be?” (PI§304). When the interlocutor imputes to Wittgenstein a denial of the ‘inner process’ accompanying remembering, he replies, “What gives you the impression that we want to deny anything?...What we deny is that the picture of the inner process gives us the correct idea of the use of the word ‘to remember’” (PI§305). And later in the *Investigations*, Wittgenstein famously writes, “an “inner process” stands in need of outward criteria” (PI§580). One way of taking these remarks (though, we will ultimately suggest in Chapter Four (**4.2**), the wrong way) is as expressing an endorsement of the distinction between how we use psychological predicates, and the conditions which make such use possible, on the one hand, and 'inner states' on the other (cf Hunter 1977).

The 'being'/'saying' distinction is equally instrumental to Ned Block's argument in ‘Wittgenstein and Qualia’ (2007). Block relates his familiar ‘inverted spectrum’ argument to some remarks in Wittgenstein's 'Notes for Lectures on Private Experience and Sense Data' (1968). In 'Notes' Wittgenstein raises a possibility that Block refers to as the 'innocent' scenario. The 'innocent' scenario is one in which, Wittgenstein suggests, we might say of a colour-observer that his colour-experience has shifted or inverted:

> "[S]omeone says, "it's queer/ I can't understand it/, I see everything red blue today and vice versa"...I think we should under these or similar circum. be incl. to say that he saw red what we saw [blue]. And again we should say that we know that he means by the words 'blue' and 'red' what we do as he has always used them as we do". (Wittgenstein 1968:284)

In this case, the distinction between ordinary colour experience, and this aberrant colour experience is capable of articulation in virtue of what we assume is a shared colour-vocabulary; we assume that concepts like 'looking green' and 'looking blue' have a common meaning for us and the abnormal observer, and that the perceptual distinction between our experience and theirs is thereby expressible.

Taking this as his starting point, Block envisages the conditions by which the shift in experience might become inexpressible (a case that he calls the 'dangerous' scenario). Suppose, he tells us, one were to undergo colour-inversion surgery, and subsequently, as in the 'innocent' case, report a shift in their perception. Over time, they accommodate their use of colour terms to match that of normal perceivers - saying of apples that they 'look red', and so on - though of course they remain aware that the way things look has changed for them. Later in life, though, they lose their memory of the colour-inversion surgery and their colour experience prior to the surgery. So, they use colour-terms as we do, and have no recollection of things ever having looked different. In such a case, it would in a certain sense be wrong, Block concedes (2007:85-87), to say that 'red things look green' to them - this is neither something they would accept, nor something we would ascribe given their behavioural indistinguishableness from other normal perceivers. But it remains true, no less than it was before their vocabulary-shift and amnesia made it undetectable, that they *do* see things differently from ourselves.

That one cannot express this difference by ordinary talk of 'how things look' to them suggests that the reach of our ordinary language is incomplete; 'looks red' “does not fully capture the content of the state” (Block 2007:67) of looking red. Block suggests, following Shoemaker (1982) that concepts like 'looking red' refer to the *intentional* content of our

experiences but fail to refer to their *qualitative* content. We may have been given “rigorous training in the application of accepted colour terminology” (Block 2007:82), such that we all agree that grass, peas and malachite 'look green'. But what 'looking green' is *like* for individual perceivers, the *way* 'green' things look, may dramatically differ. Such a case would be one of “same use of public colour terminology, different phenomenology” (ibid:87). The existence of shared public language-games concerning colour experience may not reflect or guarantee shared colour experience.

Ultimately, Block suggests on this basis that it is both logically and empirically possible that there are *ways* things look to us which are independently variable of our linguistic descriptions of our experiences; that are, therefore, *ineffable*. He insists that while we can *refer* to qualitative states “in public language, for example as the quale I get when I see green things” (2007:67), such references, again, do not 'fully capture' the content, the “individuating particularity” (ibid:62), of our experience. By contrast, statements like 'Napoleon is buried in Paris' *do* 'fully capture' the sense expressed in them. And for Block, this intrinsic quality of our experience which cannot be fully defined in terms of representational, functional, cognitive or intentional language is precisely what is meant by *qualia* (62; see also Block 1980, 1995).

So both Block and Searle rely on a distinction between objects, states or processes and facts about them on the one hand, and our criteria-governed language on the other. Searle exploits this distinction to suggest that consciousness is a real entity, and that the project of locating it within and studying its relationship to the brain is a legitimate one. Block goes further, suggesting that the ontology *outstrips* public language; that the 'ways' we experience things, and the differences between these ways, are linguistically inexpressible (ineffable).

Critically, Block and Searle share the view that Wittgenstein's position does not (whether Wittgenstein wanted it to or not) oppose such a distinction or its consequences for the nature, and study, of consciousness: Searle, remember, says that his distinction would hold even if Wittgenstein's view were '100 percent right'; and Block that the 'innocent' scenario Wittgenstein already permits inexorably lays the groundwork for the 'dangerous' scenario. So, what is at issue is, at least in large part, not whether Wittgenstein's position is *successful* in rebutting the cognitivist position, but whether it even provides an opposition to it or *could* oppose it.

The view expressed by Hacker and Bennett (among others), of course, presumes that it can and does provide such an opposition. This dispute therefore turns on how the implications of Wittgenstein's notion of criteria are understood, and so how we ought to understand that notion itself. For Hacker and Bennett, there is a necessary link between behavioural criteria and psychological predicates. Grasping their position is a matter of elucidating the link between two apparently distinct roles of criteria. In replying to Searle, they write that “only the whole animal is capable of exhibiting the *behaviour that is partly constitutive of the conditions of application of the concept in question*” (2007:102). Earlier, though, they suggest that the criterial grounds for the application of a psychological predicate “are partly constitutive of” the predicate's “meaning” (ibid:83). Searle, as we've said, understands these kinds of statements to evince an equivocation between a psychological notion's meaning, and the conditions of ascription of that notion; for him and (albeit more implicitly) Block, criterial constraints apply to the conditions of ascription, not the psychological states or processes thereby ascribed.

If Wittgenstein's view is that one and the same criterion or range of criteria can provide *both* our warrant for the ascription of a certain predicate to a subject, *and* the *definition* of the predicate thus ascribed, then we require an account of how this might work, and how it provides an opposition to the aspirations of Block and Searle. And of course, this account must be a genuinely 'Wittgensteinian' view – one plausibly creditable to Wittgenstein himself. In the next chapter, we will attempt to discern such an account, using as our guide both Wittgenstein's own works and the complex body of literature around his notion of criteria.

2 - Consciousness and Criteria

Our intention in the present chapter is to examine if, and how, Wittgenstein's model of criteria may be deployed to rebut, first, the putative distinction between ontology and language ('being' and 'saying') that Searle and Block, amongst others, employ with respect to consciousness; and second, the purported ineffability of consciousness which Block contends follows from that distinction. We will look at three distinct but interrelated understandings of how Wittgenstein's notion of criteria should be understood: what we will call the 'epistemic', 'reference-fixing' and 'inclusive' views. Roughly speaking, epistemic theories suggest that criteria determine our ways of judging and justifying instantiations of psychological phenomena; 'reference-fixing' theories that by fixing the reference, criteria define psychological phenomena; and 'inclusive' theories that criteria, in part through epistemic means, define psychological phenomena (though not simply by 'reference-fixing'). Our objective is to discern an approach to criteria which is both (i) genuinely 'Wittgensteinian'; that is, credibly imputable to Wittgenstein given the substance of his later works; and (ii)

capable of offering at least a *prima facie* rebuttal of the 'being'/'saying' distinction and its consequences raised by Block and Searle. We will demonstrate that a version of the 'inclusive' view fulfils these requirements.

But first, some caveats. We have already stated that what we are *not* doing here is providing a Wittgensteinian case against the cognitivist position espoused by Block and Searle that is more than a *prima facie* case - more, that is, than something which *could* rebut their position. Our interest is in seeing whether the Wittgenstein's view should be taken as an attempt to rebut Block and Searle's position *at all*, not whether it is ultimately successful in doing so. Secondly, the literature on criteria is complex and nuanced. The tripartite taxonomy deployed here is intended to provide a sound representation of major features of this literature, not to capture each of its subtle features. As will become apparent, we have selected particular exemplars of each approach. These are influential, plausible and typical expositions of the approaches they represent. Given the confines of the present discussion, appraising typical exemplars of broadly-drawn positions seems the best way to proceed.

It will clarify our inquiry if we begin with some general remarks on criteria, and on the three approaches to criteria under discussion here. It is non-controversial to suggest that a sustained preoccupation of Wittgenstein's work in the philosophy of psychology is the repudiation of the Cartesian view of the mind; a view of the mind that is manifest, in part, in the apparent distinctness of the mind and its incongruity with matter which we alluded to in the first chapter. With its roots in the Platonic distinction between mind and matter, the Cartesian picture asserts, or presumes, that there exists no logical or conceptual relation between behaviour on the one hand, and inner mental states on the other (cf Chihara & Fodor 1965:281; Hacker 2007b). Therefore, the ascription of psychological states to others on the

basis of such behaviours requires an inferential leap from the behavioural to the psychological.

Very briefly: it is this which gives rise to the well-known problem of other minds. Because the presence of the relevant inner state is never logically *entailed* by the presence of some behaviour/s, and nor is its absence entailed by the absence of those behaviours, the ascription of psychological states to others is always on unstable epistemic grounds. Moreover, because what we *mean* by psychological concepts - a sensation, or a qualitative experience - is necessarily something only *we* have, it is unclear that its ascription to others is not nonsensical: if what we mean by some term is just what *we* have, what we feel or observe via introspection, then how is it possible to ascribe that same state in cases where it is necessarily *un*felt, *un*experienced? Taken seriously, Cartesian scepticism thereby festers into solipsism (see e.g. Hacker 1986, 1990; Moran 2001; Overgaard 2004).

We can see how some of these presumptions are implicit in the putative distinction between 'being' and 'saying', and in particular the distinction between the kinds of behaviours that serve to 'anchor' our language-use and the mental phenomena for which they stand as evidence. It is in part this approach which, for instance, ground's Searle's insistence that behaviours are contingently, and externally (not necessarily or conceptually) associated with consciousness *itself*, notwithstanding his concession that the behaviours serve as the basis for the ascription of psychological predicates.

The notion of criteria is a centrepiece of Wittgenstein's repudiation of the Cartesian view. An obvious means of resisting these contentions and obviating the difficulties to which they lead is to claim that there *is* a conceptual relationship between psychological and

physical concepts. For Wittgenstein, criteria is the mechanism of this relationship. Precisely what kind of relationship it is is a matter of some controversy, to which we will now turn.

The most direct way of rendering a conceptual link is simply to posit a constitutive or identity relation between psychological phenomena and behavioural criteria. In the spirit of logical behaviourism, Wittgenstein might seek to suggest that a certain cluster of behaviours *are* a given psychological state, process, condition etc.

There is a wide consensus in the criteriological literature, upon which each of the three views under consideration here converge, that this is precisely the position manifest in Wittgenstein's early treatment of criteria. So, for instance, Rogers Albritton (1959), whose influential article influenced the development of the 'epistemic' view, draws on a series of remarks in the *Blue and Brown Books* (1991) to demonstrate the existence of an early 'defining' view of criteria. In one such remark, using the example of angina, Wittgenstein contrasts criteria with 'symptoms'. If medical science *calls* angina 'the presence of a particular bacillus in the blood', then to claim that we know that someone has angina by saying that we have found this bacillus in his blood is to give the *criterion* of angina; it is to state, “in a loose way...the definition of angina” (in Albritton:847). If on the other hand we point to the inflammation of the throat correlated with this bacillus, we are pointing to a symptom of angina: something which experience and observation have taught us may indicate the presence of this criterially-defined state. Symptoms thereby presuppose criteria and are related contingently rather than necessarily to a defined phenomenon (cf Wolgast 1964).

Wittgenstein relates this explicitly to psychological states and associated behaviours using the example of ‘toothache’ (Albritton 1959:847). The criteria for toothache may, he

suggests, indicate one's holding one's cheek, and exclaiming “I have a toothache”. These “first criteria” (ibid) for having a toothache may be supplemented by further observations, such as that, whenever one complains of having a toothache, their cheek appears to be inflamed. We say of such an inflammation that it always coincides with a toothache. But, says Wittgenstein, if we were asked why, in response to someone's holding his cheek and exclaiming, “I have a toothache”, we should say that he has a toothache, our answer would ultimately lead us back to *conventions*: we should be compelled to say simply that what we *call* 'a toothache', what we *mean by* 'toothache', includes these behaviours.

Albritton suggests that Wittgenstein's meaning in these passages is that:

> “[A] criterion for this or that's being so is, among other things, a logically sufficient condition of its being so. That is: If I find in a particular case that the criterion for a thing's being so is satisfied, what entitles me to claim that I thereby know the thing to be so is that the satisfaction of the criterion *entails* that it is so, in the technical sense of the word “entails” in which if a man owns two suitcases, that entails that he owns some luggage.” (1959:852)

In other words, “to be a criterion of X is just to *be* (what is called) X” (1959:852).[4] Albritton sees no reason to suggest that Wittgenstein “is using the expressions “call”, “refer to” and “describe” in abnormal senses”; Wittgenstein simply means that, for instance, “a man's preparing tea for two, say, may be part of what is properly called his “expecting someone to tea”” (ibid). Given appropriate circumstances – i.e., ones which do not contradict or induce doubt as to this being a case of some phenomenon X) – X “may *consist* in a phenomenon that satisfies a criterion of X” (ibid).

4 We have not made mention of the distinction Albritton draws between cases where some criterion is the *only* criterion of some phenomenon, and cases where it is one of a range of criteria, no one of which is necessary to our saying that X is so. In the latter case, he suggests, to be a criterion of X is to be X only *under certain circumstances* (1959:852). For our purposes, though, what is important is that on Wittgenstein's early view a criterion might be considered to *be* some phenomenon, whether in any *or* some circumstances.

In an espousal of the 'reference-fixing' view, John Koethe (1977) adopts much the same reading of Wittgenstein's early approach to criteria. Like Albritton, he relies on Wittgenstein's suggestion that the criteria for angina constitute a 'tautology' or 'loose definition' of angina to suggest that Wittgenstein's early view was that statements of criteria and their associated phenomena (that for which they stand as criteria) were synonymous: angina *means* 'an inflammation caused by bacillus B' (ibid:604). And likewise, John Canfield (1974), in positing a version of the 'inclusive' view of criteria, agrees, again by reference to Wittgenstein's remarks on angina, that Wittgenstein's early position was a 'defining' one: that "the presence of a criterion is linked to the thing of which the criterion is a criterion via a convention, definition or rule of language" (ibid:71).

Where these three positions come apart is in their response to Wittgenstein's initial view. Both Koethe and Albritton find the initial view manifestly implausible, on two related counts: its failure to register the 'non-entailment' relation between criteria and subject (that is, fact that the presence of some sufficient criterion or range of criteria typically does *not* entail the presence of some phenomenon) and its failure to register our sense that, in speaking of these subjects, we take ourselves to be referring to something 'beyond' the behavioural criteria exhibited. But both detect a shift in Wittgenstein's later view of criteria towards a more credible position; where they disagree is in diagnosing precisely what this more credible position amounts to. Canfield, with whom we will to a large extent agree, understands Wittgenstein's later work as expanding upon, rather than shifting from, this initial position.

The present discussion will demonstrate, first, that the 'reference-fixing' and 'epistemic' views of criteria fail to provide a *prima facie* rebuttal of Block and Searle's claims.

Second, we will show that both positions neglect an important thread in Wittgenstein's later thought, such that neither can be credibly imputed to Wittgenstein. So they fulfil neither of the objectives we initially stated. The 'inclusive' view *does*, we will argue, respond to this thread in Wittgenstein's work, and in Chapter Three, we will argue that, in doing so, it provides a *prima facie* case against the 'being'/'saying' distinction; it therefore satisfies both of our stated requirements. Finally, we will defend the 'inclusive' approach from some objections.

2.1 The 'Epistemic' View

With our framework established, we can look at each of these positions in greater detail. Let us return, first, to the 'epistemic' view. Albritton concludes, as we discussed above, that Wittgenstein's remarks in the *Blue and Brown Books* evince the view that criteria may simply be identified with the phenomenon for which they stand. He is sharply critical of this view of criteria, particularly its consequences for psychological predicates and their criterial behavioural manifestations. He denies, for instance, that what we call someone's 'having a toothache' could be anything that a man “says or does”; rather, “there is something that I have *called* having a toothache, when I had a toothache, namely *having* one” (1959:853). Similarly, regarding the psychological state of 'expecting', he dismisses Wittgenstein's proposal in *The Blue Book* that 'expecting' could be any one, or combination, of the behaviours which, in combination with the right circumstances, we might diagnose as indicating that someone is 'expecting' someone or something: “What is expecting? If it is an empty idea that what we call “expecting” is a queer, incorporeal something hidden away in that remarkable medium, the mind, what *do* we call “expecting”?...Unfortunately, the *Blue Book* is ready with a wrong answer...[That] [u]nder certain circumstances “all this [the

behavioural processes] is called 'expecting B from 4 to 4:30.'" But it isn't." (ibid). As such, criteria *fails* to supplant the Cartesian picture of the mind with a persuasive alternative; the sense of a "hidden something or other that is uniquely called" (ibid) by some psychological predicate is not refuted.

However, Albritton detects a significant and redeeming shift in Wittgenstein's treatment of criteria between the *Blue and Brown Books* and subsequent works. The shift detected is a movement from a constitutive view of criteria to an *epistemic* one. On the latter view, some phenomenon (X) does not consist in criteria (and therefore, the presence of those criteria does not entail X), but criteria *do* constitute the epistemic conditions that determine whether we say that something is shown to be, and shown not to be, an instance of X. They may, therefore, entail that we are *justified in saying* that something is or is not an instance of X. What is more, they may entail this is a case where X is *almost certainly* present. This is so because it is a corollary of *what we mean* by 'having a toothache' that, under normal circumstances, somebody with a toothache will manifest a certain behavioural criterion or range of criteria (ibid:855-6).

This last point has an important consequence[5]. Albritton insists that although criterial propositions such as, "a man who behaves in this manner, under normal circumstances, always or almost always does have a toothache" (1959:856) have the appearance of factual/empirical claims, they are understood by Wittgenstein, on this later view, to be *necessarily* true. So the shift towards the 'epistemic' view is not a shift from a defining view

5 We should be clear that Albritton only briefly raises this consequence in his initial (1959) article; and whilst the idea of 'necessary evidence' is essential to the 'epistemic' view, it would perhaps be overly strident to impute the 'epistemic' view to Albritton's rather cautious remarks. Instead, we are using these remarks to motivate and develop our portrayal of the 'epistemic' position subsequently developed by Lycan (1971) and others (see below).

to mere premises for inductive hypotheses about the presence or absence of a certain, e.g. psychological, phenomena. If it were, Wittgenstein may as well have conceded in his later view that there is no necessary link between behaviour and psychological states. Rather, Wittgenstein's later view of criteria, on this view, is that they serve as *necessary* evidence for something's being or not being the case, and sufficient grounds for one's making the justified claim that something is or is not the case. This is the 'epistemic' view.

This view gained considerable traction in the literature subsequent to Albritton's article. In his thorough survey of the criteriological literature, William Lycan speaks favourably of the role of, as he calls it, 'noninductive evidence' in Wittgenstein's notion of criteria (1971:110). He agrees with Albritton that Wittgenstein, between and even within his earlier and later works, speaks in at least two different ways of criteria. But like Albritton, he takes the better view to be that whilst criteria do not entail some psychological state or process, and therefore are "*not*...defining characteristic[s]" (ibid, my italics), it is (and here he is quoting with approval Chihara & Fodor 1965) nonetheless "necessarily true that instances of [some phenomenon] Y accompany instances of [a criterion] X in *most* cases, or in all 'normal' ones...the very meaning or definition of Y...justif[ies] the claim that one can recognise, see, detect or determine the applicability of Y on the basis of X in normal situations" (in Lycan 1971:110). Unlike 'symptoms' (items of evidence that are merely contingently linked to instances of a given concept), then, criteria are conceptually linked to a given phenomenon, *not* for Wittgenstein's initial reason that the phenomenon *consists* in criteria but in that the criteria constitute dimensions of our epistemic engagement with that concept.

2.1.1 The 'Epistemic' View and the Case Against Cognitivism

We will now consider whether the 'epistemic' view can provide *prima facie* opposition to Searle and Block's remarks. We will suggest it cannot. First, we need to deal with something of a red herring. It is important not to underemphasise the sense of 'necessity' at play in the 'epistemic' view. It is easy to do so because, as Lycan's survey demonstrates, there is an association in the literature between the epistemic view and 'genetic' characterisations of the role of criteria. For instance, it might be claimed that the 'necessity' of some piece of evidence is derived from its "playing an essential role in the way certain concepts are formed, and in the way certain words are learned"; or because the difference between criteria and symptoms is expressed in terms of symptoms being things that we have "found to be" evidence for a concept, rather than things we have "learned to call" evidence (ibid). Such accounts seem to suggest that the fact that criteria serve as necessary evidence for some phenomenon is a matter of convention, and is indicated by the role that criteria play in our concept-*acquisition*.

A reading of the 'epistemic' approach which places this kind of emphasis on the 'genetic' role of criteria would likely fail to oppose the cognitivist position espoused by Block and Searle. Recall Searle's point that we may 'carve off', in individual cases, consciousness from its criteria. If Wittgenstein's view is only that 'what is sometimes the case could not always be the case', then we might put Searle's point as the complimentary thesis that 'what must sometimes be the case need not always be the case'. That is to say: we might grant that behavioural criteria serve an indispensable role in establishing our existing psychological language-games – that we would not have the present psychological concepts, including consciousness, that we do, were it not for certain behaviours that formed part of our concept-acquisition. In the case of pain and pain-behaviour, for instance, we might agree with

Wittgenstein that the natural relationship between experience and behavioural expression is what facilitates the capacity for our words to “refer to sensations” (PI§244). We may further grant that the linguistic patterns of our psychological terms reflect their origins: the ways we explain and justify our ascriptions of psychological concepts, the ways we teach their use, retain strong associations with behavioural criteria. But we may nonetheless argue, with Searle, that we are entitled, *given* the background of the language-game established by criteria, to 'carve off' the thing ascribed from the conditions on which ascriptions are grounded; carve off the referent, that is, from the conditions of reference (see also Putnam 1957; Gibbs 1969; Hunter 1977).

However, such an interpretation largely undervalues the force of 'necessity' in the 'epistemic' view espoused by Lycan, Albritton and others. John Pollock's (1967) discussion of criteria is useful in bringing the sense of 'necessity' in this view to light. Pollock contends that violations of relations between criteria and some phenomenon are conceivable only against a background of their correspondence in normal cases. In Pollock's terms, violation is intelligible only against the background of 'general beliefs', beliefs which are acquired on the basis of the correspondence between criteria and the relevant phenomena in ordinary cases. So, for example, diagnosing a mental disorder which causes a person to make false pain-avowals can only sensibly be done against a stable background of ordinary cases in which pain-avowals are held to be genuine (cf Kelly 1991; Rorty 1972). Lycan (1971), amongst others, links this position to a remark of Wittgenstein's:

> “The fluctuation in grammar between criteria and symptoms makes it look as if there were nothing at all but symptoms. We say, for example: “experience teaches that there is rain when the barometer falls, but it also teaches that there is rain when we have certain sensations of wet and cold, or such-and-such visual impressions.” In defence of this one says that these

sense-impressions may deceive us. But here one fails to reflect that the fact that the false appearance is precisely one of rain is founded on a definition" (PI§354).

This suggests that the fact that rain has these epistemic criteria is *necessary* insofar as there exists a logico-grammatical connection between it and these epistemic criteria; it is a corollary of what we call 'rain' that it has these epistemic criteria.

More broadly, criteria constitute these epistemic features of associated phenomena, and in this way 'punctuate' our epistemic language-games. The circumstances in which we speak of being 'certain', of 'knowing' that, e.g., some psychological concept is applicable, and equally the circumstances in which it makes sense to doubt its applicability, *are fixed by criteria*, determined by their occurrence or non-occurrence.

The logical relationship between criteria and epistemic language-games in such instances means that where relevant criteria are satisfied, one does not need and *cannot have* any further justification for claiming that some phenomena is, or is not, instantiated in a given case. The resolute sceptic may continue to express doubt that some phenomenon is instantiated – and legitimately so, given that the satisfaction of some criterion or set of epistemic criteria does not *entail* an instantiation of the associated phenomenon. But their complaint cannot be expressed in terms of a demand for further or better evidence; the criteria define, exhaustively, the range of possibilities for verification, certainty and so on (the 'epistemic possibilities', we might say). As Lycan puts it, "criteria are the ultimate (logically possible) court of appeal in deciding the questions to which they are relevant - even though they neither provide deductive certainty nor exhaust the meanings of the terms whose use they govern" (1971:112). On this somewhat stronger rendering of the 'necessary evidence'

view, we can see that our 'genetic' emphasis in some respects misses the profundity of the 'epistemic' approach's claim.

Nonetheless, even put in these stronger terms, the 'epistemic' approach remains vulnerable to the ontology/language distinction. Searle and Block may be prevented by it from seeking more direct *ways of knowing* that some psychological phenomenon is instantiated. But the restriction of this approach to epistemic grounds still means that there remains a distinction – precisely the distinction Searle insisted upon - between the criterial grounds for saying that some thing is instantiated and the thing said to be instantiated. It may be a corollary or aspect of what some phenomenon *is* that it has certain epistemic connections. But there is, of course, more to how it is defined, 'what it is' than how we judge it to be instantiated. As we've noted, proponents of this view stress that epistemic criteria are *not* "defining characteristics" (e.g. Lycan 1971:110); rather, we might say ascribe to them the view that *in virtue of* the definition of something, in virtue of 'what it is' (which is a distinct matter) it necessarily has these kinds of epistemic criteria. Therefore one may not, without violating the bounds of sense, seek to 'know' something more directly. But one *may* sensibly seek to *define* or *refer to* it more directly, to know better 'what kind of thing it is' as opposed to knowing that it is instantiated in some case. And to that extent, one might legitimately seek to 'carve off' the thing itself from its epistemic grounds.

For proponents of this view, that this gap between the thing itself and criteria opens up is essential to the redeeming shift away from what is considered an unacceptable 'defining' view in the *Blue and Brown Books*. But so long as there is some distinction between *what* something is and how we know *that* it is, and so long as it is only the *latter* which are

necessarily linked to that thing via criteria, then it appears that the 'being'/'saying' distinction is not completely opposed.

If the ‘epistemic’ view of criteria is correct, then, it appears that criteria cannot provide a *prima facie* case against Block and Searle's arguments; and we face a choice between an implausible early view and a more modest one which fails to rebut cognitivism.

2.2 The 'Reference-Fixing' View

Given the failure of the 'epistemic' account to provide a *prima facie* case against cognitivist aspirations, and given in particular that this failure was a consequence of its inability to negate the distinction between referent and conditions of reference, we might turn instead to a view which identifies criteria with the referent more directly: the 'reference-fixing' view. Perhaps the clearest support for this position comes from John Koethe's (1977) discussion.

Koethe's position is framed as a corrective to more popular 'epistemic' approaches. He begins by identifying a widespread conflation between the 'epistemic problem' - “how do we tell which mental state a person is in on a particular occasion” - and the 'semantic problem': “how do we manage to talk meaningfully about mental states” (1977:602). Where the epistemic view characterises the shift in Wittgenstein's later position as a shift towards a concern with the epistemic problem, Koethe insists that the transition that takes place between the *Blue and Brown Books* and Wittgenstein's later work is from one kind of response to the semantic problem to another: as such, "Wittgenstein's notion remains one of "defining criteria" throughout, but...his views on definition evolve between the *Blue Book* and the later works" (ibid:603).

The early 'defining' view Koethe detects is one we have already canvassed, through the example of angina: a simple 'synonymy' relationship between a statement of criteria and a statement of associated phenomena. Wittgenstein's later works, Koethe suggests, retain the defining role of criteria, but understand definition in terms of 'reference-fixing' rather than synonymy. Criteria provide the framework for telling us what phenomenon or kind of phenomenon some predicate refers to in both cases. But in the later view, rather than criteria having a synonymous or 'meaning-giving' relationship with their associated phenomena, such that, e.g., the phenomena is necessarily present if, and only if, the criteria are present (which Koethe suggests leads to entailment), criteria provide only contingent reference-fixing 'links' to the target phenomena; they enumerate contingent characteristics by which the phenomena may be identified.

The distance from the early defining view to the later one is considerable. First, the 'reference-fixing' view is considered a 'defining' view only under Koethe's rather broad notion of 'defining'. Koethe proposes as conditions of definition that “first...definitions are meant to *establish* semantic relations, rather than describe them or state that they hold”; “second...the definition of a referring term should *make clear* to the linguistic community what the referent is; not...must state what it is, or provide a true description of it”; “third...nearly *anything* serving these purposes might be regarded as a definition” (1977:610).

And, in stark contrast to definition by 'synonymy', Koethe proposes that criteria in Wittgenstein's later view are related to their *definiendum* (e.g. a psychological phenomenon) only contingently. He uses this broad notion of definition to suggest that definitions need not be necessarily, analytically, or *a priori* true, or true by definition/convention/rule of language. He takes more or less as established Kripke's well-known suggestion based on the metre rule

in Paris[6] that a definition need not possess analytic or necessary truth. He argues, by means of a hypothetical, that it need not possess *a priori* or definitional/conventional/linguistic rule truth either. Briefly: we are asked to imagine a primitive society who devise a small unit of measurement of approximate length – a 'minch'. They fix the reference for this definition using the full moon, which, owing to deficits in their astronomical knowledge, they take to be the size it appears to the naked eye from earth (and so a small object). So they determine that "one minch = the diameter of the moon" (1977:612). Suppose that this definition by reference becomes the determinate measurement used to construct other facts and assertions ("most members of the community are 60-70 minches tall", "huts shall be at least 100 minches from..." etc. [612]).

Now, suppose the tribe subsequently discovered that the moon was in fact much larger than previously thought. Since their definition refers to a specific item's length, we may say either that (i) their definition continues nonetheless to be true, but that what they thought they were referring to was not 'the moon' at all but something else, or (ii) that their vast range of associated facts and assertions is simply wrong; *or*, what is for Koethe the most sensible alternative, that (iii) this definition *did* serve to define (in the broad sense stipulated

6 The details of this account are well-known, but to briefly summarise: Kripke famously suggested that a statement that serves as a definition, such as "one metre = the length of B [the metre bar in Paris] at [time] t", need not be necessarily true nor analytically true. Regarding necessary truth: suppose that the quoted definition does fix the reference of 'one metre' – one metre is the actual length of B at t. It is counterfactually possible that B might have had a different length at t. But if its actual length is 'one metre', then its counterfactual length is other than 'one metre'. So this definition does not hold true in all possible worlds; it is not necessarily true. Following Koethe, we may suppose that, whatever else the concept of analyticity incorporates, analytic statements are necessarily true; so by dint of the above argument this definition is not analytically true either.

by Koethe) a 'minch', and has been proven empirically to be false; so it was not known *a priori*.

Finally, regarding 'truth by definition, convention or rule of language', Koethe insists that, where Kripke's example of the metre rule in Paris already shows that something can be true without serving as a definition, the 'minch' example further shows that something can serve as a definition without being true; so:

> "If [some definition] (M) could have been true without serving as a definition, and it could have served as a definition without being true, I cannot see how (M) could be said to be true by definition except in the *trivial* sense in which the truth of *any* statement depends on definitions and conventions-namely, that had the terms in the statement been given meanings and references different from those they in fact have, the statement would not have been true (or meaningful)." (1977:614)

With these and Kripke's arguments Koethe purports to completely sever anything but a contingent connection between defining criteria and phenomena. It is in this contingent sense that, he claims, criterial statements should be taken as defining. So, for instance, criterial definitions have the form of "Pain = the mental state persons typically exhibit by C" (ibid) where C is some behavioural criterion or set of criteria. In this sense Koethe's view is complimentary to the epistemic position: criteria provide, rather than the epistemic framework, the *ontological* framework, the definitive characteristics of the referent phenomenon: they identify the thing or kind to which the term applies. Of course, the major difference is that here the link between criteria and phenomena is contingent. That links between criteria and phenomena hold true is a function of:

> "the beliefs, practices, institutions, and the psychological and physiological character of the members of the community in which it is adduced; as well as 'certain very general facts of nature'. That these background conditions...should be propitious is a contingent matter." (1977:616)

So, for instance, that someone in pain always or almost always exhibits C is a purely empirical matter. Its being almost always true follows not of necessity from its function as a definition but in virtue of fortuitous and wholly contingent features of our biological condition: that, broadly speaking, we express pain in certain ways, and that we are able to recognise certain behaviours *as* expressive of pain (ibid:615).

2.2.1 The 'Reference-Fixing' View and the Case Against Cognitivism

It is perhaps already obvious that (whatever else is wrong with it) Koethe's argument, because in jettisoning necessity altogether it places an even *greater* distance between criteria and phenomena than the 'epistemic' view, does not provide a *prima facie* case against the 'being'/'saying' distinction. In fact, it is much more plausibly an exemplar of the kind of 'genetic' approach we have considered than was the 'epistemic' view. Koethe's suggestion seems to be that the capacity for certain criteria to serve a reference-fixing role is dependent upon there being, in 'nature', a regular correspondence between the criteria and the phenomenon. This relationship includes, in the case of psychological predicates, our natural human capacity and tendency to *express* psychological states through certain behaviour, and correlatively, the capacity and tendency of others to *recognise* certain behaviours as expressive of certain states (ibid:615).

Perhaps if these connections did not exist, we could not, or would not, have acquired the concept of this phenomenon. But if this relationship is a contingent, empirical matter, and if criteria merely *establish* ('fix') the reference at first instance, then there is nothing preventing the same phenomenon from being defined by other means - nothing to stop us augmenting, substituting or disposing of our present reference-fixing criteria - and certainly

nothing to stop us 'carving off' the referent phenomena from the criteria that made reference to it possible. It therefore gives considerable leverage to the 'being'/'saying' distinction Searle and Block wish to make.

So it appears that neither the 'reference-fixing' view nor the 'epistemic' view provide an adequate *prima facie* opposition to the cognitivist claims explored in Chapter One. But does either position represent a genuinely 'Wittgensteinian' view; that is, one which could be responsibly imputed to Wittgenstein? In the next section, we will turn to this question.

2.3 The 'Epistemic' View, the 'Reference-Fixing' View and Wittgenstein: Eight Scattered Remarks

Our appraisal of the exegetical soundness of the 'epistemic' and 'reference-fixing' views will begin with eight remarks in the *Investigations*: §145, §177, §183, §444, PI§541, §573, PI§586, and Part II, section ii, lines 10-14. These remarks were selected not by us, but in fact by Albritton. In judging a shift to have occurred towards 'epistemic' criteria in Wittgenstein's later view, Albritton concedes that at least “eight scattered remarks in the *Investigations* still suggest that a criterion is something that may be described as X under certain circumstances, something in which X may consist” (1959:854): that is to say, eight scattered remarks suggest that Wittgenstein has *retained* his earlier view of criteria. Albritton nonetheless imputes a shift to the 'epistemic' approach to the later Wittgenstein, remarking that these eight remarks simply suggest that Wittgenstein “may have been unconscious of this change, striking as it is” (ibid).

Albritton's suggestion that Wittgenstein was unconscious of a profound shift in his own view is strikingly glib, and suspiciously self-serving. It is also unconvincing given the

substance of the eight remarks. For it is our contention that, far from being 'scattered', these remarks evince a distinct and unified viewpoint; and do so, by Wittgensteinian standards, quite unambiguously. In fact, these remarks give expression to a recurring and central preoccupation of Wittgenstein's later thought, one which *neither* the 'epistemic' view nor the 'reference-fixing' view adequately registers, and by dint of which the apparent dichotomy between these two views is shown to be misleading and a better view revealed.

We will begin by looking at each of these remarks in turn; following this we will give some account of *Koethe's* response to these and similar remarks in the *Investigations*, which is more detailed than Albritton's.

Firstly, in §145, we see Wittgenstein, in the midst of the rule-following discussion, introducing a behaviour-based conception of 'understanding' or 'learning' a rule:

> "Suppose the pupil now writes the series 0 to 9 to our satisfaction...Now I continue the series and draw his attention to recurrence of the first series in the units...let us suppose that after some efforts on the teacher's part he continues the series correctly...now we can say that he has mastered the system.-But how far need he continue the series for us to have the right to say that? Clearly you cannot state a limit here."

What is important here is what the pupil *does*, not what psychological states he is experiencing.

Similarly, in §573, Wittgenstein is undermining the idea that 'having an opinion' possesses some essential 'statehood' independent of particular cases in which we speak of it. He directs us instead to these cases: "What, in particular cases, do we regard as criteria for someone's being of such-and-such an opinion? When do we say he reached this opinion at this time?...The picture which the answers to these questions give us shews *what* gets treated grammatically as a *state* here".

And again in §541, Wittgenstein counsels us to resist the temptation to suggest that there is a "quite particular kind" of psychological feeling that constitutes 'meaning' or 'understanding' something by some particular combination of words. Instead, we are directed, again, to *expressions* of these feelings, to behaviours (including, of course, speaking). Wittgenstein undermines his interlocutor's suggestion that "the point is, the words *felt to him* like the words of a language he knew well" (my italics) by suggesting that even something's 'feeling a certain way' to someone is a notion dependent upon behavioural criteria, for instance that "he later said just *that*" (that the words felt a certain way).

In §177, in similar terms, Wittgenstein once again denies a canonical state of 'being guided', any "experience [of] the because". But here he consolidates this with a denial that there is any particular *behaviour* either which, on its own, constitutes being guided: "it is correct to say I drew a line under the influence of the original: this, however, does not consist simply in my feelings as I drew the line...under certain circumstances, it may consist in my drawing it parallel to the other-even though this in turn is not essential to being guided". So *what is called* being guided in particular cases depends on the impression produced in the observer by a confluence of behaviour and circumstance; neither some action nor some inner state sufficiently constitutes this phenomenon in its own right.

§183 takes us further still, suggesting that this confluence of behavioural criteria and appropriate circumstances is not a determinate calculus resulting in *entailment* of some phenomena, but is defeasible and open-ended; in speaking of the various senses of 'walking', Wittgenstein suggests we must "be on our guard against thinking that there is some *totality* of conditions corresponding to the nature of each case (e.g. for a person's walking), so that, as it were, he *could not but* walk if they were all fulfilled".

Part II, section ii, 10-14, §586 and §444 brings similar considerations to bear on the sense in which the same form of words (like the same behaviour – for speaking, of course, is a kind of behaving for Wittgenstein) may have different meanings in different circumstances. So, in Part II section ii, he considers the word 'till' – in one case a verb, in another a conjunction – and suggests that conceiving its different meanings is a matter of conceiving different *circumstances* of use: one is "not asked to *conceive* the word one way or another out of any context, or to report how he has conceived it". The meaning of the word is not, as it were, something static, remote, determinate, but something that comes with these 'contexts' – there *is* no understanding the meaning without understanding the use of the term in particular cases. In the same way, in §586 he suggests that the phrase "I'm longing to see him!" may or may not be an expression of expectation or anticipation; it might just as easily be the result of self-observation. The important thing, "the point", is, he tells us, "what led up to these words?" And in §444, he emphasises the difference in meaning of the phrase "he is coming" when used on its own, and in the sentence "I expect he is coming". Were the meaning of this phrase fixed independently of circumstances of use, this flexibility could not be accounted for; so quite like in PI§541, 'what we mean' is not a question of the words themselves but of the circumstances in which they are used. So, what some phenomenon *is*, how it is defined, is a matter of this non-entailing confluence of behaviour and circumstance.

We will venture a more explicit analysis of these eight remarks in the section to follow presently. Suffice to say that Albritton seems quite right to have detected a sense of the early, 'defining' view of the relationship between criteria and phenomena. These remarks seem to express a sense that understanding what some phenomenon – 'understanding' itself for example - is, is a matter of observing the circumstances in which one would be prepared

to say that, e.g., someone ‘understands’ something; and to suggest that this preparedness to apply some predicate is in turn hinged upon the exhibition of behavioural criteria in appropriate circumstances. We will argue that this line of thinking is not some unconscious vestige of Wittgenstein's early view, but the key to understanding his later view.

Before we make this argument, though, it is worth considering Koethe's response to the appearance of this line of thought in Wittgenstein's later work. Where Albritton merely disavows these passages as a hang-over, unrepresentative of the shift in Wittgenstein's view, Koethe makes some attempt to come to grips with several of these and similar remarks. In particular, his discussion refers to Wittgenstein's remarks on opinion and expectation, as expressed in §573, one of the eight passages above, and PI§452, which is not one of the eight remarks but deals with 'expectation' in a very similar fashion to §444 and §586, warning us away from isolating “the mental process of an expectation” and redirecting us to its expression (“if you see the expression of an expectation, you see what is being expected. And in what other way...would it be possible to see it?”).

Koethe rightly understands the first of these remarks, §573, to indicate that criteria are not understood by Wittgenstein as providing merely the conditions for epistemic warrant: these are not the conditions for “determining when a particular person is, in fact, in [state] S” (1977:606), but instead play an ontological role, telling us what kind of state some phenomenon is. But in contrast to Albritton's sense that this remark indicates definition in the sense of the *Blue and Brown Books*, Koethe concludes that we ought to understand the phrase 'what gets treated grammatically as a state here' in line with the 'reference-fixing' view of criteria; as indicating “what [some state] S refers to” (ibid).

The second remark (§452) is also taken to support the 'reference-fixing' view. Koethe responds to Wittgenstein's suggestion that there is no other sense but via behaviour in which we can speak of seeing a state to mean that behavioural criteria, in light of our (contingent) capacity to recognise the behaviours of other animate beings as expressive of inner states, serve as our means of perceiving these states, though criteria are not identified with them (cf Schulte 1993, Ch 3).

2.4 A Third Way: Meaning as Use and the 'Inclusive' View

However, we will argue here that there is a different and exegetically more persuasive approach to these remarks. Briefly put, they are best understood in terms of a pervasive theme in Wittgenstein's work which neither the 'reference-fixing' nor the 'epistemic' views adequately register: the connection between meaning and use.

In the eight remarks above we find the recurring suggestion that psychological predicates - understanding, opinion, rule-guidance, meaning something by one's words, and so on - cannot be defined by looking at the meanings of the predicates independently of their circumstances; *nor* by identifying some exhaustive, determinate calculus of circumstances and behaviour within which these predicates take on a certain meaning; *nor yet* by isolating some kind of object or entity – a 'state', a 'feeling', etc. - by reference to which they are ostensively defined. So, for instance, Wittgenstein suggests that the same behaviour may in the right circumstances stand as a criterion for 'expecting', and in others for self-observation; that drawing a line, or counting from zero to nine, *may* or may not constitute the criteria for understanding; that pointing to the sky might or might not constitute meaning a certain thing by certain words - but that in none of these cases should these be taken as essential, final, or

exhaustive, and in none should we look for such a discernible *essence*, least of all in some concealed state or entity.

Wittgenstein's tone in these passages is rhetorical, deconstructive. His concern, quite clearly, is not with formulating a strong positive account of meaning so much as undermining common inclinations and presuppositions that accompany a search for meaning; and his ambitions are therapeutic. By guiding us towards the circumstances of use in which we comfortably, and unproblematically, apply psychological predicates to given cases, he seeks to demonstrate that in fact, we *already* understand what, and how our words mean, and this is shown by our comfortable and clear application of them in various circumstances. It is only our fixation upon the idea that there must be some*thing* behind and beyond this linguistic employment, our desire to locate some concrete object or essence to explain and ground our language use, that gives rise to a sense of dissatisfaction with our language-use (cf Baker & Hacker 1993).

These suggestions are of a kind with his edict in PI§143 that “for a large class of cases – though not for all – in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language”. Wittgenstein's ambition in the selected passages is to answer the question 'what does X mean?' by directing our attention to cases where, because we confidently and consistently apply a given term, the answer to that question is already known; and to suggest, moreover, that there is nothing more to what that word means than what is manifest in these instances of use where its meaning is already known (cf Hacker 1990, 1990b, McGinn 1997).

Within this picture, criteria have a promiscuous, inclusive function: they are the phenomena or states of affairs which, in the right combination with each other and in

combination with the right circumstances, form an indeterminate, defeasible basis for our usage of concepts. In other words, they constitute the phenomena (in the case of psychological predicates, the *behaviours*) which, in virtue of the rules of our language, *we call* X in certain circumstances. Critically, this formulation intends to cover *both* defining *and* epistemic uses of a certain word. The distinction between these two uses, moreover, becomes blurred, and, indeed, less important. For our 'calling something X' in given circumstances, our ascription of that word to a given case, is both an epistemic use *and* a defining use. The usage reflects both by what, *and* in what circumstances, we take some phenomenon to be instantiated, and so not just when it is instantiated but what kind of thing it is. Meaning is not, on this view, ot by some concealed object or essence, nor is it determinable by some fixed calculus of criteria and circumstance. Rather, it is constituted by this shifting network of uses determined by the presence of criteria. We can call this the 'inclusive' view.

Though he does not emphasise the centrality of 'meaning as use', John Canfield's (1974) discussion offers a view of criteria that resonates quite closely with the above claims. Essentially, Canfield is *denying* that any shift towards an 'epistemic' or 'reference-fixing' view of criteria occurs in Wittgenstein's later thought[7]. Drawing again on the angina example, Canfield suggests that Wittgenstein's view of criteria remains the 'defining' view, whereby the “presence of a criterion is linked to the thing of which the criterion is a criterion via a convention, definition or rule of language" (71).

7 In fact, Canfield is taking himself to be defending a version of the view that Albritton, in the 1966 postscript to his 1959 discussion, expressed: that if a behaviour is a criterion for, e.g., 'toothache', it is part of the grammar of our language that one who exhibits it, in at least most cases, has toothache.

To clarify Canfield's slightly awkward formulation: let us suppose that it is a rule of language that angina is defined as the presence of bacillus B in the blood. And suppose we discover bacillus B in my blood. This discovery entails, in virtue of the rule of language that determines *what we call* angina, that I have angina; because according to that rule, the presence of bacillus B *is* angina. We could compare this to our analysis of any of Wittgenstein's eight remarks: the student's independently counting from 0 to 100 is a criterion for his having learnt the number sequence because, in virtue simply of the conventions of our language, this behaviour (given appropriate circumstances) determines the use, and so the meaning, of 'learning the number sequence'; my peering out the window, saying “I hope he comes soon”, and the like serve as criteria for my 'expecting' someone because we would be prepared to use 'expecting' here, such that this is part of what expecting means. Criteria, in this instance, 'track', 'reflect' the rules of language; in virtue of the conventions that govern our language, their presence or absence in given circumstances determines our patterns of ascription[8]. Accordingly, Canfield grants criteria a wide, 'inclusive' role: he suggests that, *inter alia*, they can be used to claim something is the case, to form or justify a judgement, and to identify something (ibid:74).

8Importantly, Canfield takes his sense of 'rule' or 'convention' from PI §54, which identifies three senses in which we say a game being played according to a rule: "The rule may be an aid in teaching the game...Or it is an instrument of the game itself.-Or a rule is employed neither in the teaching nor in the game itself; nor is it set down in a list of rules...But we say that it is played according to such-and-such rules because an observer can read these rules off from the practice of the game." So either, in a sense analogous to the 'historical' function of criteria discussed earlier, a criterion operates to teach us the meaning of a psychological concept, but once the concept is learned, the rule is no longer used; or it is used as part of the language-game itself; or we do not use the criterion consciously as a rule, but our language-game accords with its being described as a rule. To say that X is a criterion of Y is to say that 'if X, then Y' is true in virtue of these ways of using language.

Explaining criteria in this way not only makes good sense of the remarks of Wittgenstein's which we surveyed above, and others like them. If we think of criteria as determining 'what we call' something, we obviate the need to distinguish between the epistemic and reference-fixing views of criteria; and we can make good sense of the intuitions that led to both the 'epistemic' and the 'reference-fixing' view whilst explaining how each of them miss their mark, exalting one emphasis at the expense of another, and presuming too wide a distinction between meaning and use.

The 'epistemic' view rightly proposes that there is a necessary connection between some phenomenon and its epistemic criteria, such that it is a corollary of the definition of, say, rain that it has particular epistemic connections. But in attempting to distinguish this view from Wittgenstein's early position, proponents of this view seek to restrict the necessary relationship between criteria and phenomena to an *evidential* role, explicitly distinct from a defining role. It is suggested that it is in consequence of what something is that certain things serve as evidence for it; but *what* it *is*, however, is a distinct question. Indeed it is explicitly stressed that criteria are not “defining characteristics”, but rather *in virtue of* those characteristics, they serve as evidence for some phenomena (1971:110). This distinction is quite deliberate; as we noted in **2.2**, it is the essence of the shift perceived from Wittgenstein's earlier view to his later view. And it is a distinction which licences the 'being/'saying' distinction on which Block and Searle depend.

Albritton's eschewal of the 'eight scattered remarks' as a mere vestige of Wittgenstein's old view is indicative of this limitation in the 'epistemic' view. For in the 'eight scattered remarks' we find the suggestion that criteria, within certain circumstances, define 'what we call' some phenomena, and do not just define what stands as evidence for it; and this

is conceived by Albritton as a return to Wittgenstein's early position. We can make sense of the persistence of these eight remarks by suggesting that in fact, no shift away from a 'defining' view has occurred. Moreover, we can say that those remarks in the *Investigations* that suggest criteria determine the 'epistemic conditions' of our phenomena are not, in fact, at odds with the eight scattered remarks. Rather, epistemic uses *are* among the defining uses of a predicate. If we take proper account of the maxim of meaning as use, it is clear that, for Wittgenstein, the conditions under which we ascribe some concept (its 'epistemic dimensions') partly constitute that concept's meaning. Criteria are therefore *both* partly constitutive of the conditions of ascription of some concept and, by the same token, partly constitutive of the concept thereby ascribed; for the concept is defined by its circumstances of its employment. The idea of a shift towards 'epistemic' criteria from 'defining' criteria fails to recognise that epistemic uses are, already, defining uses.

Koethe's account, on the other hand, rightly recognises that the 'epistemic' view stops short of proposing an association between criteria and 'what something is'. But he too is concerned to avoid imputing to the later Wittgenstein the implausible view of the *Blue and Brown Books*; and so only claims a *defining* rather than *epistemic* association between criteria and phenomena on the condition that a logico-grammatical connection between criteria and phenomena is jettisoned. Koethe is right to suggest that formulations such as “Pain = a state typically exhibited by C” tell us what kind of thing something is, rather than merely indicating its epistemic criteria. Like Canfield, Koethe rightly argues that there is no shift to the *exclusively* epistemic and away from the 'defining' view of criteria.

But he fails to note two things. First, the persistence of a 'definitional' view *does not rule out* Wittgenstein's identifying criteria, in part, with the epistemic grounds of our

assertions, because these assertions are part of the use, and so the meaning, of psychological predicates. And second, his suggestion that criteria merely fix the reference of some phenomena, that they "establish semantic relations" *with* some phenomena, "make clear...what the referent is" (cf Schulte 1993, Ch 3), falls short in the same way as the 'epistemic' view of picking up on the sense in which, in these 'eight scattered remarks', there exists a logico-grammatical link between criteria and phenomena, rather than a contingent relationship between the criteria and our 'conditions of reference' to that phenomena. On the present reading of Wittgenstein's view, criteria do not 'tell us what kind of thing something is' just by referring to it, or by setting up linguistic links or correspondences with it. They tell us what something is in virtue of the fact that what it is is a product of our use of it in our language-game. Use does not *facilitate* meaning by reference, but *constitutes* meaning; our uses are not just reference-fixing but meaning-giving. This explains, in a way which does not rule out the 'epistemic' view, the two remarks that Koethe remarked upon in the previous section; that "grammar tells us what kind of object anything is" (PI§371); and, in PI§452, that there is nothing to what some state is but behavioural criteria in appropriate circumstances. It is so because there is nothing more to what some concept means than what is found in these manifold grammatical employments of it.

We have suggested that an accurate exegetical appraisal of these remarks in the *Investigations* does indeed, as Albritton suggests, imply the continuation of a 'defining' view of criteria; that this view can be reconciled, in a way the alternative views cannot, with a major theme in Wittgenstein's later work, namely: 'meaning as use'; and that viewing criteria in this way allows us to incorporate the strengths, and appealing intuitions, of the 'epistemic' and 'reference-fixing' views, whilst demonstrating that they are not mutually exclusive. We

can claim, on this basis, that this reading of Wittgenstein appears at least to succeed on exegetical grounds, as a genuinely 'Wittgensteinian' view of criteria.

In the next chapter, we will begin by considering whether the 'inclusive' view, in addition to being exegetically plausible, might support a sound *prima facie* case against the cognitivist claims of Block and Searle. Following this, we will consider several objections to this view, each of which, perhaps unsurprisingly, stems from a sense that it suffers from the same implausibility for which Wittgenstein's earlier view was abandoned by the 'epistemic' and 'reference-fixing' theories.

3 – Testing the 'Inclusive' View

In the first section of this chapter we will show how Wittgenstein's view opposes the 'being'/'saying' distinction, and 'qualia inversion' scenario that is premised upon it. In the second section, we will raise and reject the possibility that the 'inclusive' view inherits from his early view the implausible possibility that criteria *entail* phenomena. And in the third section, we will raise and reject the related suggestions that the 'inclusive' view entails crude forms of behaviourism or verificationism, and idealism about truth.

3.1 The Cognitivist Threat Opposed

If the view we have outlined in the previous chapter is correct, then Hacker and Bennett's view appears to be vindicated. They are right to speak in the same breath of the criteria for the conditions of ascription of some predicate and the criteria for the predicate itself. For on the 'inclusive view' the conditions of ascription are part of the concept itself; they constitute uses which are constitutive of the meaning of that concept. As such, the distinction Searle attempts to draw between the conditions of ascription and the thing

ascribed, a distinction Block exploits to provide for the possibility of qualia, may equally be explained as the misguided separation of meaning and use. Searle's presumption is that Wittgenstein's position is *consistent* with isolating the thing talked about from the various ways in which we talk about it. But, like the 'reference-fixing' and 'epistemic' views, Searle has underestimated the necessary connection between these two elements. The excerpts above demonstrate that grasping the meaning of words is a matter of attending to the often diffuse circumstances in which words are meaningfully used. These aggregates of uses might build up a certain picture, but it is misguided, Wittgenstein insists, to take this aggregative picture as a distinct or distinguishable object.

The same logic condemns the suggestion that there might exist a something about which nothing could be said. For it is in virtue of the rules of grammar that we know 'what we call' something – when and by what it is instantiated, and what it is. If something is bereft of a criterial framework, if we do not know 'what we call' it, then on what basis might we say that it is a something or a nothing, true or false, exists or does not? If we cannot say anything about 'something' – cannot use it or call anything it - then *it cannot be a something*: it is in virtue of criteria, on this view, that we know what kind of thing it is and the circumstances in which it is and is not extant. It is by *language-use*, in other words, that its ontological and epistemic dimensions are constituted.

On this view, the qualia inversion scenarios proposed by Block and others present a violation of the bounds of sense. The breadth of the 'inclusive' view, its suggestion that criteria are exhaustive of *both* our ways of judging whether something is or is not instantiated and our knowing what kind of thing it is, provides a firm basis to oppose Block's talk of

differences between *ways* of seeing something which do not, *cannot* express themselves via criteria – because here, an inexpressible difference is simply no difference at all.

As we mentioned, a view closely resonant with the 'inclusive' view is that espoused by John Canfield. The implications we have raised are clearly not lost on him: he has devoted a much more recent article (2009) to considering and rejecting Block's (2007) Wittgensteinian qualia inversion scenario; and his approach sheds light on the opposition the 'inclusive' view provides to Block's position. In Block's 'innocent' scenario, remember, two individuals have different colour-experiences, and, in virtue of shared colour terminology, this difference is detectable and expressible (see also Kiverstein 2009). Block uses three further steps - colour-inversion surgery, vocabulary adjustment, and memory loss – to get to a 'dangerous' scenario; a scenario where it is incorrect to say that 'red things look green' to that person, but where in a certain sense they of course *do*. In such a case we are bound to accept the limits of language and to speak instead of ineffable *ways* things look to us.

The problem with this scenario, Canfield insists, is where it begins. Block imagines the 'innocent' scenario as a case where our experiences differ from others, and we subsequently express this fact. He thereby envisages a relation between observers, their experiences, and their expressions of these experiences. From here, the slide to his 'dangerous' scenario seems difficult to avoid. But according to Canfield, Wittgenstein envisaged the 'innocent' scenario in the 'Notes for Lectures' as concerning only expressions and our corresponding responses to them:

"[S]omeone says, "it's queer/ I can't understand it/, I see everything red blue today and vice versa"...I think we should under these or similar circum. be incl. to say that he saw red what we saw [blue]. And again we should say that we know that he means by the words 'blue' and 'red' what we do as he has always used them as we do".(Wittgenstein 1968:284)

We conclude, on the *basis* of certain (verbal) criteria, that he sees things differently.

For Canfield (2009) – and, we suggest, for Wittgenstein - Block's understanding 'jumps ahead', so to speak, of the language game, taking this criterially-based conclusion as a starting premise. Block takes the 'picture' that the 'innocent scenario' has built - a picture of someone's experiencing something different, and moreover, a picture of an 'inner state' – and, so to speak, 'excises' it from the language-use from which it emerged. The stages that lead to the 'dangerous' scenario are based on the idea that this picture can remain fixed while our uses change around it. So he proceeds, so to speak, with picture in hand, to a case where because of memory loss and linguistic retraining, one does not behave differently or report their experiences as different, and where other party/ies in consequence do not ascribe to them a different experience. But nonetheless, he insists, the *picture*, the state of this person's inner experience, is unchanged. Only our uses have, through each stage, lost touch with this state; it has become *ineffable*.

But on a Wittgensteinian view, Block has here committed the cardinal error of separating meaning from use. If we put use and picture in the right order, we find that *we can only say* of someone that he 'experiences something different' on the basis of certain criteria. It is the distinguishing behavioural criteria which *make it the case* that this is an instance of (what we call) someone's 'seeing green as red'. In the complete absence of these criteria, we cannot, so to speak, 'get to the stage' of saying this. The assertion itself, its truth or falsity, is governed by criteria. If someone, for instance, readily picks out green objects when asked to, reports their colour experience as normal, and so on, then this is simply not a case where red things look green to him – as the criteria are 'the highest court of appeal'; there is no other

basis on which we might judge the situation to be otherwise. So there *has been* no qualia inversion.

At the heart of Canfield's critique of Block is the sense that Block is purporting, *per impossible*, to speak of our language-game from beyond it. Canfield draws on Wittgenstein's example, which he takes to be a *reductio*, of the possibility of pseudo-blindness: "A blind man sees everything just as we do but he acts as a blind man does" (in Canfield 2009:697). For Wittgenstein, though such a case *seems* to make sense to us, it is a mere 'image'; it is not in fact something we can put to use in our language games. Canfield insists that "if we imagine a blind man frantically struggling to feel his way out of a smoke filled burning house, it becomes absurd to suppose that he sees but only acts as if he doesn't" (ibid:698). He is not simply pulling on our heartstrings here, nor making some kind of appeal, Humean in spirit, to the ridiculousness of philosophical concerns in light of the everyday. He is suggesting that this possibility can only be *suggested* in virtue of criteria which at the same time make it *impossible*, or rather, *nonsensical*. More precisely: the very concept of blindness, and our assessments of cases thereof, has as its defining criteria certain behavioural manifestations[9]. One cannot imagine the complete absence of these criteria without at the same time imagining the absence of the concept itself. In purporting to do so, one is conducting a kind of logical sleight of hand - in Canfield's words, "speaking from

9 There are quite remarkable parallels, though beyond the direct purview of our discussion, with the hypothetical notion of 'super-blindsight' discussed by Block and others (e.g. Block 1995) – a notion which extrapolates from actual cases where some responsiveness to visual stimuli is retained despite 'phenomenal' blindness to a case where *complete* responsiveness is retained. Dennett (1995) suggests that such a case would “stretch our credulity beyond the limit; we would not and should not take somebody's word that they were "just guessing" in the absence of all consciousness...in such a case”(252). In other words, we would simply *call* such a case consciousness, on the basis of consistent appropriate behaviours; as we would simply call a case of someone acting blind 'blind'.

within the framework of ordinary language", from "inside the language-games concerning 'sees' and 'blind'", yet judging "contrary to their rules" (ibid:698)[10].

Block is drawing on common concepts, and so taking advantage of existing language-games, but subsequently excusing his own use of these concepts from these language games; presuming, therefore, that his own use of a concept is somehow not part of the totality of uses of that concept, but sits beyond or behind them. His view is therefore characteristic of the 'cognitivism' mentioned in Chapter One, insofar as it expresses a dissatisfaction with the totality of our language-games, and assumes that these games are outstripped by a reality which it is their purpose to mirror. Cognisant of Wittgensteinian objections, he introduces a distinction between something merely 'looking red', which he admits cannot express the distinction he wants to make without falling into logical error, and *ways* of its looking red. But for his Wittgensteinian critic this merely sets the problem back a step: "'the way red things look is green' and 'red things look green' are, in ordinary language, governed by the same criterion. Therefore it changes nothing to substitute talk of *ways* for simple talk of looks" (ibid:710).

In taking this view of Block, Canfield is giving voice to his earlier (1974) point that, again in Wittgenstein's words, criteria are the "highest court of appeal" (PI§56; Canfield 1974:75) in the formation of judgements. Beyond words set against their proper, criteria-based grammatical framework, there is merely nonsense. Canfield's strategy provides support for the 'inclusive' view. We can see how, by understanding criteria as governing 'what we call' something, the 'inclusive' view places these criterion-less applications of our terms –

10 P.F. Strawson puts a similar point well in a passage from *Individuals*, when speaking against sceptics: "He pretends to accept a conceptual scheme, but at the same time quietly rejects one of the conditions of its employment" (p.35; Cf Rorty, 1971).

and so, the distinction between 'being' and 'saying' - beyond the realm of sense. Once some psychological notion is *inexpressible*, or undetectable, in behaviour, it is, in virtue of this inexpressibility, nonsensical; once a distinction does not register in our language, there is not an inexpressible distinction, but no distinction at all[11].

It appears, then, that the 'inclusive' view of criteria, as well as being a well-founded representation of Wittgenstein's view, sustains a strong *prima facie* case against Searle and Block's aspirations. It therefore fulfils the objectives we set ourselves at the beginning of Chapter Two. However, in what remains of this chapter, we will consider some objections to the 'inclusive' view.

3.2 The Problem of Entailment Revisited

Thus far we have ignored an important objection to the 'inclusive' view of criteria. We have expressed the view, in agreement with Canfield, that the best interpretation of Wittgenstein on criteria is that this view remained, throughout his career, tethered to a definitional link between criteria and associated phenomena; a link of which 'epistemic' uses of criteria are merely a part, and which provides not just 'reference-fixing' conditions but *meaning*. But a large part of the appeal of the 'reference-fixing' and 'epistemic' views is that they allow for a redeeming shift in Wittgenstein's view away from his early remarks on criteria. It is widely maintained that the remarks in the *Blue and Brown Books* express a view of criteria that has the implausible consequence that the presence of some criterion or set of criteria *entail* the presence of a given phenomenon.

11 Canfield's claim, like Pollock's (1967), at times trades on something like a 'transcendental' approach to Wittgenstein's notion of criteria; see, for more general discussion, Rorty (1971), Schwyzer (1973), Hacker (1972), Westphal (2005).

Of course, our objective is to locate a genuinely Wittgensteinian view that provides a *prima facie* case against Block and Searle's claims; it does not extend to *espousing* or *defending* Wittgenstein's view. Nonetheless, Wittgenstein's view cannot provide a *prima facie* case against Block and Searle's view if it is manifestly weak and implausible; and nor would it be exegetically credible to impute a manifestly weak and implausible view to the later Wittgenstein. So in the present section, we will show how the 'inclusive' view might avoid, so to speak, entailing entailment. We will first look at a defence offered by Canfield (1974) – but this, we will conclude, only succeeds at the expense of conceding defeat to Block and Searle's position and contradicting the Wittgensteinian opposition to it we established in the previous section. We will then look at some remarks from Wittgenstein himself which can be understood to provide a more credible response to the charge that this view 'entails entailment'.

We will begin, then, by explaining Canfield's position. Canfield suggests that the common perception of a shift in Wittgenstein's thought away from his early view is prompted by a sense that a link *by definition* between criteria and their subject implies entailment. If, for instance, the criteria for a psychological phenomenon are linked by definition to criteria – if this, indeed, is what make criteria *criteria*, rather than merely symptoms – then it would not appear logically open to us to deny that some phenomenon was present when its criteria are present.

Clearly, though, it *is* ordinarily understood to be open to us to make such denials. For instance, FIFA's criteria for 'football' are 'an air-filled sphere with a circumference of 68–70cm, a weight of 410–450g, inflated to a pressure of 0.6–1.1 at sea level, and covered in leather or other suitable material'. However, there are seemingly open-ended range of

defeating conditions which would prevent the satisfaction of these criteria from entailing the existence of a football - suppose that in addition to FIFA's criteria the ball had the magical capacity to yell offensive epithets at players when they kicked it, or was forever changing location without a discernible physical cause. It seems difficult, perhaps impossible to exhaustively enumerate and formulate provisos to cover each possible defeating condition. And much more importantly, our use of concepts is clearly not conditional upon such exhaustive knowledge of or ability to articulate these conditions.

Canfield (1974) concedes this point. But he makes use of an important distinction between linguistic rules and unstated background conditions to argue that absolute, defining criteria are compatible with non-entailment between criteria and their subject. Our use of criteria operate, Canfield claims, against an overwhelmingly stable background of conditions which facilitate the correspondence of these criteria with their subjects. To use an example salient for present purposes: our practice of using colour words makes sense against a background of our sharing more or less the same neurological and perceptual equipment, of that equipment operating in a more or less consistent fashion, and of our living in a world where the molecular properties of objects are such as to give rise to more or less stable colour properties in interaction with our biological equipment (cf Shoemaker 1982). Were these background conditions to alter, our present criteria for the usage of colour-concepts would be ineffective: that this chair appeared blue at one time would not suggest that it *was* blue if its appearance was in constant flux or our perceptual abilities were consistently unreliable. Of course, our rule-governed use of colour concepts does not entail that we are aware of all or any of these background conditions that make such rules for use possible; Canfield (1974) writes:

"[I]n general, in a criterial rule used to teach a language or used in the practice of one, there will be no statement given of the general facts of nature and normally existing circumstances that form the background against which the criterion is used...one teaches [and uses] against this background, one does not teach that it obtains" (1974:80).

So, it may be a rule of our language that if criterion X occurs, state Y obtains. As a rule of language, this is true *by definition* – and it is exceptionless and absolute. However, the linguistic rule does not incorporate the background conditions that facilitate its employment: conditions which make it the case that, where X occurs, Y *is* present. These background conditions may change in such a way that X could occur without Y.

It is this possibility which precludes a strict entailment relation between X and Y. However, because these background conditions, and *a fortiori* the possibility of their variation, is not *part* of the language-game, nor part of how it is taught, but is an unstated background against which teaching and participating in language-games operates, the absence of strict entailment on the one hand is consistent with exceptionless, fully defining criterial rules on the other.

That, then, is Canfield's (1974) view. We want now to consider an objection to it, one of our own making. The upshot of it is that Canfield's position fails to exclude, and in fact implies, the kinds of ineffable states or entities for which Block argues. Its form is quite simple. Canfield suggests that though our judgement that, e.g., a particular psychological phenomenon is instantiated based on the presence of appropriate criteria is exceptionless and absolute, the ineliminable possibility of a breakdown in the natural background conditions that facilitate the correspondence between criteria and their subject means that such criteria-governed ascriptions do not entail the presence of the phenomenon thereby ascribed. The example of colour-concepts, and perceptual equipment/stable colour properties, has already been offered.

It is important to grasp Canfield's (1974) distinction between our rule-based uses of language and these background conditions. These background conditions are not just conditions of which we are implicitly aware when we use language or form linguistic rules; nor things about which "everybody concerned knows, or can be easily informed" (1974:80). We do not normally "possess knowledge of the limits of criterially governed concepts" (81). Canfield is not, of course, suggesting here that these background conditions are *unknowable*. But what he *is* suggesting is that there are states of affairs, beyond our language use, which give make these language-uses either entailing or non-entailing.

Compare this now to Canfield's more recent argument against Block (2009), which hinges on the idea that beyond our criteria-governed language-uses, there is merely nonsense. This later argument trades on the idea that criteria set the limits of what is said to be 'true' and 'false', and of the existence of objects or states; so there is no qualia, nor qualia inversion, beyond what is expressed within our language-games.

The combination of these arguments is problematic. For on the earlier view, the epistemic dimensions of our language-games are determined by states of affairs that are 'extra-conceptual' or 'extra-linguistic'. The non-entailment which characterises our psychological language-use is determined by these extra-conceptual states of affairs, and their being extra-conceptual is what gives rise to the characteristic of non-entailment. However, on the later view states of affairs which are extra-conceptual are condemned as nonsense.

If we were to adopt *both* views, the result is that there exist states of affairs which do not show up in our language-games, and that such states of affairs are 'nonsense' because they exist beyond our language-games. But it is not straightforward, therapeutic, later-

Wittgenstein nonsense – these are states of affairs which are *there*, beyond our language. It looks worryingly similar to *Tractarian* non-sense: unspoken, extra-conceptual 'somethings about which we cannot speak'.

What's worse, it is worryingly similar to *Blockian* 'nonsense'. Moreover, it is arguably achieved by just the same means, condemned out of Canfield's (2009) own mouth (see **3.1**). For in speaking of states of affairs, beyond our language, which affect the epistemic dimensions of our concepts, he is essentially stretching our concepts beyond what he later takes to be their own bounds. Specifically, he is borrowing notions of 'states of affairs' which are *themselves* dictated by criteria, and applying them to cases which, because they are extra-conceptual, *have* no such criteria. He is employing elements of our language-game to purport to speak of what is apparently beyond it. As Block himself says, "the view...that leads to the epistemic problems that exercise Wittgensteinians is that there are determinative facts...independently of our cognitive access to them" (2007:80). Canfield quotes this passage, to illustrate Block's error, in 2009; but it is just such states of affairs which his earlier discussion seems to entertain.

One might say in Canfield's defence that his position is different from Block's in that he is suggesting the existence merely of things of which we *do* not speak, not of which we *cannot* speak. The background conditions which he speaks of are not, again, things we cannot know; they are just not part of our present use of our concepts. But they could be, were we to thoroughly investigate the 'conceptual limits' of our language-games. Background conditions are not, therefore, said to be beyond the realm of sense in the same way as qualitative colour experiences.

However, such a response masks exegetical problems in Canfield's position. It is far from clear that Wittgenstein would countenance Canfield's speculation that we *could*, if we chose, improve our concepts so as to make the relationship between criteria and their subject one of strict entailment. To suggest that our present criteria may be improved or altered, that they are therefore contingent and separable from their subject, runs directly against the very point that we have, with Canfield's help, been making: that it is Wittgenstein's view that criteria are connected not contingently but by the rules and conventions of our language with their subjects, and that these 'subjects' are individuated by, and identified with, their criteria. On this view, to change criteria would just be to change concepts altogether.

What is more, to suggest that our language-games are, so to speak, 'beholden' to something beyond them runs against another pervasive theme in Wittgenstein's later work: what Hacker (e.g. 1986) has famously dubbed the 'autonomy of grammar'. The 'autonomy of grammar' constitutes Wittgenstein's rejection of the view expressed in the *Tractatus* of a direct isomorphism between language and reality. In his later view, a concept's 'grammar', its logico-grammatical dimensions, expressed in the variety of ways that we use it, include the epistemic language-games we associate with it. Concepts of entailment and non-entailment are part of the 'grammar' of our concepts. Suggesting that the distinction between entailment and non-entailment is reliant on empirical background conditions amounts to the suggestion that the forms of our grammar are determined by correspondence with an extra-grammatical reality. But Wittgenstein himself suggests that such a view is a misunderstanding. In *Philosophical Grammar*, he explains:

"*As long as* we remain in the province of true-false games a change in grammar can only lead us from *one* such game to another, and never from something true to something false. On the other hand if we go outside the province of these games, we don't any longer call it 'language'

and 'grammar', and once again we don't come into contradiction with reality." (Wittgenstein 1974:68)

Implicit in this remark is a sense that the forms of our grammar are autonomous from 'reality', and it is wrong to see their features as a product of *conflict* or *accordance* with reality. As Hacker puts it, "facts of nature do not make concepts *correct* or *true to the facts*" (1986:190). Hacker goes on to suggest that facts of nature *do* relate to our language-games in a different sense: it is only in light of certain stable regularities of nature that our language games are *purposeful* and *useful*. We would have no *need* for colour-concepts if we did not see, or (perhaps) if we did not see with sufficient regularity. Canfield's argument seems to incorporate a recognition of this point, but he takes matters too far - he allows non-entailment, a *grammatical* feature, to be dependently variable upon correspondence with 'very general facts of nature' (cf Kiverstein 2009).

So, it appears that the position espoused by Canfield is not as straightforwardly effective as first thought. The difficulty, it seems, lies in grounding a feature of grammar - non-entailment, in this case - in certain extra-conceptual features of the world. As long as we do this, the possibility of a gap between language and reality, underpinned by the presumption of a semantic *correspondence* between language and reality, is possible; and with it, the possibility of features of reality which lie beyond expression, 'somethings' about which nothing is said. Indeed, Canfield's position seems not just to permit but to lead to that very possibility.

We have a dilemma. Though the prospect of implausible entailment marred the viability of our initial view, relying on Canfield's strategy for avoiding entailment brings us further from an exegetically plausible version of Wittgenstein's position, *and* brings us into

conflict with Canfield's own later (2009) view. If we attempt to remove the inconsistency retracting his earlier (1974) view, we continue to face the entailment problem. If we remove the inconsistency by retracting his later (2009) view, this deprives us of the grounds for opposing Block's position – and by association, the 'being'/'saying' distinction - that we relied on in **3.1**. But keeping both gives us a position that is neither exegetically plausible *nor* succeeds in opposing Block and Searle's claims.

However, we will demonstrate now that an account of non-entailment which is exegetically plausible and consistent with Canfield's – *and Wittgenstein's* - later view may be derivable from Wittgenstein's own works; in particular, from remarks on criteria in his *Last Writings on the Philosophy of Psychology (Volume 2)* (1992).

Here, Wittgenstein writes of the "endless multiplicity" of our behavioural expressions of the mental; of, for instance, the "countless configurations of smiling...And smiling that is smiling, and smiling that is not" (1992:81). He connects this multiplicity with what he regards as an essential feature of the 'mental' - the "unforeseeability" of others' behaviour, and, what is closely related, the uncertainty of the evidence for "what is experienced" (ibid:65). Elsewhere, he remarks on this relationship between indeterminacy and 'the mental' as follows: "we don't need the concept 'mental' (etc.) to justify that some of our conclusions are undetermined, etc. rather, this indeterminacy, etc., explains the use of the word 'mental' to us" (ibid:63); and again, in similar fashion: "it is not the relationship of the inner to the outer that explains the uncertainty of the evidence, but rather the other way round - the relationship is only a picture-like representation of this uncertainty" (ibid:84). Finally, he suggests that "if one says that one never knows whether someone really felt this way or that, then that is not because perhaps after all he really felt differently, but because even God so to speak cannot

know that the person felt that way...in a game in which the rules are indeterminate one *cannot* know who has won and who has lost" (ibid:85-6).

We can make sense of these remarks in the following way. Those psychological states and processes which we might categorise as 'mental' are distinguished by there being an indeterminate range of behavioural criteria (and indeterminate combinations thereof, and indeterminate combinations of criteria and circumstance too) which we take as sufficient for their ascription in a particular case. Some combination of criteria and circumstance may be taken to provide sufficient evidence of some mental state; but there is no definite, isomorphic correspondence such that the absence or presence of some criteria or other is necessary and sufficient for the ascription of a psychological concept. It is this peculiar indeterminacy, a feature of the *logic*, or *grammar* of our psychological language-game - and not nature – which secures the non-entailment relation between criteria and that for which they operate as criteria. Moreover, it is this grammatical feature that distinguishes the concept of the 'mental' or the 'inner', and which makes the distinction between the 'inner' and the 'outer' intelligible and useful: “the inner differs from the outer in *its logic*. And that logic does indeed explain the expression “the inner” (ibid:62).

The multiplicity of criteria for psychological concepts is a logico-grammatical basis of the non-entailment relation. Our concept of 'the mental' is a *picture* of this logical form. It is *not* the case that some feature of 'the mental', understood, say, as a biological feature of human beings and so a fact about the natural world, determines the non-entailment relation between criteria and their subject. It *may* of course, be the case that biological facts about human beings make this kind of game *useful*. But the game itself is, nonetheless, autonomous, or 'self-sufficient'. Explaining its features does not take us outside grammar into

the extra-conceptual world. And it is for this reason that 'not even God' could know whether somebody was in some state or other. There *is* no knowing this. 'Knowing' is nonsense here; it is logically excluded from our language-game by the way criteria operate in this context. The uncertainty that manifests in non-entailment is not, as Canfield would have it, some contingent feature of our language-games correctable by a closer examination of the 'mental' itself. For Wittgenstein this puts things backwards: the 'mental' is an aggregative product of our psychological language-games, distinguished in part by 'non-entailment'. Non-entailment is therefore a *necessary* feature of 'the mental'.

By locating 'non-entailment' *within* grammar, rather than in extra-grammatical facts of nature, the 'inclusive' position is put on safer ground. According to the 'inclusive' view, criteria play a constitutive, defining role in their subject. So, for instance, certain behaviours determine 'what we call' certain psychological phenomena. En masse, the logical features of these criteria – their multiplicity, their vague conceptual boundaries ('smiling'), in short the fact that they do not entail the instantiation of associated phenomena – determine 'what we cal' the 'mental', a more general grouping of psychological phenomena. In other words, in Wittgenstein's picture, 'non-entailment' can be understood as one particular outcome of the fact that criteria 'define' their associated phenomena – and therefore as thoroughly consistent with the 'inclusive' view.

Further, by extrapolating the full implications of the 'inclusive' view of criteria, and conceiving of *criteria*, rather than nature, as playing a determinative role in the entailment relations between criteria and subject, we arrive at a solution which avoids the inconsistency of Canfield's earlier (1974) view and his later (2009) view. This view explicitly departs from the notion of our grammar being determined by an 'extra-conceptual' world, and so retains an

opposition to Searle's distinction between ontology and language and Block's subsequent ineffability claims. Moreover, that we have found this solution in Wittgenstein's own works suggests it stands on firm exegetical footing.

Having dealt with the first of our objections, we will, in the final section of this chapter, consider two further objections to the 'inclusive' view of criteria.

3.3 The Problem of Behaviourism Revisited; and the Problem of Idealism

Though our final two objections are quite different, we can respond to them in much the same way as each other, and for that reason can deal with them in the same section. The first worry is most clear in Albritton's rejection of Wittgenstein's early view, discussed in 2.1; it is expressed by Albritton's statement that “there is something that I have *called* having a toothache, when I had a toothache, namely *having* one” (1959:853). Simply put, it seems discordant with our understanding of our language to speak of criteria as constituting their subject; not in virtue of the entailment problems already discussed, but because it is simply wrong to suggest that, e.g., an expectation *is* someone preparing tea for two. What we mean by 'expectation' is not 'someone preparing tea for two', nor 'someone looking out the window', but, frankly, *an expectation*, which these behaviours persuade us is present. We can call this the 'behaviourist' objection, the implication being that Wittgenstein, like the behaviourists, is reducing psychological phenomena to behavioural criteria, with implausible results.

The second objection is closely related to the entailment problem already considered. A fine expression of it can be found in an article by Norman Malcolm (1982). Here, Malcolm considers the possibility that by, so to speak, locating 'truth-conditions' within grammar – by suggesting that “the word 'know', like any other word, has its place in the language game”

(1982:262) - Wittgenstein is committing to an implausible *idealism* about truth. Malcolm's discussion is important, because it demonstrates *why* someone might adopt a view that the non-entailment of criteria/phenomena relations is a feature of nature, and why they might be uncomfortable seeing it as an autonomous feature of grammar. Given the importance of the grammatical autonomy of non-entailment to our rejection of Canfield's problematic explanation of non-entailment (**3.2**), it is worth considering Malcolm's position.

Malcolm suggests that the philosophical use of the word 'know' is commonly understood to be governed by the conditional, “if I know that *p* then *p* is true” (ibid:262). But if we suppose that our use of the word 'know', like any other aspect of our language-game, is dictated by defining criteria the satisfaction of which warrant our assertion that we 'know' something, then it may appear in virtue of the above conditional that they also warrant the assertion that something is *true*. But according to Malcolm, the idea that *language*, rather than *'reality*', could shape *what is true* in this way appears to constitute an unattractive kind of linguistic idealism (ibid:263). It is this unattractive possibility that drives Malcolm to adopt a position akin to Canfield's (1974): that a statement's being true depends on a correspondence, in the natural world, between the criteria and the instantiation of some phenomenon.

In seeking to impute this view to Wittgenstein, Malcolm points to a number of remarks, primarily from *On Certainty*. For instance, he draws on the following remark: “it would be wrong to say that I can only say 'I know that there is a chair there' when there is a chair there. Of course, it isn't *true* unless there is, but I have a right to say this if I am *sure*...even if I am wrong” (Wittgenstein 1979:§549). He takes this and similar statements to support, first, that we can be fully warranted, by the exhaustion of our epistemic criteria, in

saying, e.g., “S knows that P”, without that statement being true; and second, he infers from this that such a statement's truth is determined by extra-conceptual states of affairs.

Obviously, there are similarities here to Canfield's strategy of making entailment conditions dependent upon nature. What is *new* for us about Malcolm's already familiar strategy is that he offers it as an antidote not to entailment but to 'linguistic idealism'; to the idea that “reality is *created* by language, thought, judgment” (Malcolm 1982:266). If we locate epistemic characteristics of our language-games, such as non-entailment, in grammar rather than extra-grammatical 'nature', and so depart from the position espoused by Malcolm and Canfield, are we forced into a crude linguistic idealism? That we *are* is the substance of the second objection. Once again, our response to these objections need only be sufficient to sustain an exegetically plausible, and *prima facie* reasonable case against Block and Searle's claims.

The behaviourist and idealist objections are prompted by the same concern: that, quite clearly, we do not *mean* by our terms what a robust notion of criteria such as the 'inclusive' view appears to suggest. We understand truth and falsity to be determined by states of affairs in the world, not by mere words. If we understood truth to be a mere matter of words then the satisfaction of our epistemic standards would perhaps entail the instantiation of the associated phenomenon. But it does not. And similarly, when we ascribe a psychological state, such as 'expecting', we do not *mean* that some individual is engaging in a particular behaviour (making tea, looking out the window, checking his watch) – we *mean* that the person is expecting someone. The same is true, for example, of 'desiring', 'learning', 'understanding' – it seems to us that we mean *just these things*, and not some behaviour or other that we might take to indicate their instantiation in a given case. Again, if it were not so, this would rule out

what is clearly a sensible possibility: that these behaviours might be engaged in without 'expectation', 'understanding' etc. being instantiated.

Indeed so. But there is a sense in which, considered from the Wittgensteinian position we have developed, these assertions are *themselves* statements about the rules and conventions of our language-game. Beginning with Malcolm's worry: it is the case that we understand truth as dependent on the world and not upon language. But it remains open to Wittgenstein to suggest that *that* we understand it this way is a feature of our language-game[12]. To suggest that truth is dependent upon 'life' and 'reality' where these terms are understood as somehow entirely distinct from our language-games may be diagnosed as a repetition of Canfield's error of borrowing certain characteristics from *within* our language-games – namely, the dependence of truth upon facts in the world, and the possibility of a distinction between these facts and our beliefs about them – and attempting to use them in a way which excludes itself from that language-game, which purports to speak from a position outside language.

Some explanation is needed here. First, from the perfectly acceptable proposition that our *knowing* something does not entail its being true, Malcolm draws the conclusion that language and reality are independent. But the possibility of being warranted in saying something without its being true is one that is allowed for *by* our language-games, and so is *determined by criteria*.

12 It may be objected that this is not so much a refutation of Malcolm as the presumption of opposing premises. Indeed so; but our ambition here is merely to show that, and why, *on* the Wittgensteinian view, concerns about linguistic idealism are misguided. This is sufficient to at least make out a *prima facie* coherent Wittgensteinian position.

For Malcolm to imagine a case where, according to our epistemic criteria, we know that, for instance, a chair is present in a room, even when 'in fact', though we never find out, it is not present, is from a Wittgensteinian view similar to imagining a case of 'pseudo-blindness'. It is certainly imaginable, but not a situation that can verifiably arise. In imagining it one is adopting a kind of 'god's eye view', imagining both our knowing that something is true, and alongside it the state of affairs of its not being true. Such a view ignores the fact that the state of affairs of its not being true, like our knowing that it is true, is governed by epistemic criteria. Like Block, Malcolm is, on a Wittgensteinian view, 'jumping ahead' of the language game. Just as Block excised the picture of the state of affairs that emerged from the 'innocent' scenario, and assumed that it held constant across different language-uses, so here Malcolm is taking advantage of a state of affairs that we take to have arisen following the application of epistemic criteria, but envisaging a case where that state of affairs has arisen in the absence of criteria.

There is no *actual* (rather than imagined) case where we can speak of, e.g., a chair's being there when we know it isn't; for our saying that 'it is there' is governed by criteria in the same way as our knowing that it *is* there. Certainly we may *discover* the presence of a chair in a room even when we were justifiably certain of its absence, even when we were entitled to claim we *knew* it was absent. But only in virtue of the further application of criteria: we say that we know a chair is absent on the basis of some epistemic criteria, and we subsequently say that this was false on the grounds of further epistemic criteria which indicate the chair's presence.

Similarly, Malcolm's suggestion that there are facts in an extra-conceptual reality that may or may not correspond with our language-games, and that the correspondence between

language and reality determines 'truth', misrepresents Wittgenstein's position. From what 'higher court of appeal', is he making the assertion of some 'real truth' outside of our 'linguistic' truth, if not from within the language-game and so on the basis of criteria? On Wittgenstein's view, there is no higher court, no non-conceptual basis for talk of 'truth'.

But that Wittgenstein takes this view does not commit him to 'linguistic idealism'. The suggestion that the autonomy of our grammar entails linguistic idealism depends on presuppositions Wittgenstein need not accept. When Malcolm questions how 'reality' could be dependent upon language, he is misrepresenting the level at which this statement makes sense. Of course, *within* the conventions, practices or rules of our language-games, the notion of 'reality' is distinct from what we *say*. So within our language-games, the dependence of truth upon language is indeed illicit. But this is consistent with saying that these statements about licit and illicit uses are descriptions of *features of our language-games*[13]. On this view, Malcolm's assertion that reality is distinct from language itself takes place within language and is a reflection of grammatical rules. As Canfield's later (2009) response to Block makes clear (see **3.1**), the suggestion, no less from Malcolm as from Block, that there is a reality beyond language, is curiously self-defeating, because it is a suggestion which necessarily occurs within the realm of concepts. The kind of truth idealism Malcolm imagines is, on Wittgenstein's view, incoherent[14]; and so Wittgenstein need not oppose it by proposing an extra-conceptual reality that determines aspects of our grammar.

13 But cf Williams (1973) – Williams applies something close to our solution, but points out the possibility that *it too* may be considered a commentary on our language-games *from without*. Unfortunately we cannot entertain this possible further difficulty here; suffice to say our reply to Malcolm gives us enough to sustain a *prima facie* plausible Wittgensteinian position.

14 See note 12 above.

The form of this response can be applied, *mutatis mutandis*, to the behaviourism objection. It is true that Wittgenstein's early view, in the *Blue and Brown Books*, likely evinces a kind of crude behaviourism which he eventually developed beyond. But his development was not in the form of a shift to a 'reference-fixing' or 'epistemic' view. Rather, what developed in his later view was that, rather than meaning being fully defined *as* some behaviour or other, it was defined by a term's "use in the language" (PI§143), where *that use* was determined by criteria.

Moreover, it is indeed the case that what we mean by 'expecting', 'understanding', &c. is *just these things*, and not just some characteristic behaviour. But that we conceive of, e.g., expecting as a state distinct from some behaviour, and of that behaviour as therefore having a non-entailment relation with that state – these are themselves features of our *language-game*, and not of the world (cf e.g. Wittgenstein 1967 §357; Williams 1973). By subscribing to the 'inclusive' view of criteria we do not mean to suggest that an expectation is 'nothing but' some behaviour – for this would imply an entailment relation between the behaviour and the state. It is, certainly, a much richer confluence of behaviour, circumstance, and our dispositions to justify, explain, doubt, to 'go on from' our ascriptions in certain ways and not others, that underpins what we mean by ascribing an expectation to some behaviour. But all of this, for Wittgenstein, is nonetheless occurring *within* the circumstances of use that constitute our language-game.

Suggesting that Wittgenstein's view of criteria implies behaviourism may, again, be countered with the claim that this is conflating different levels of description – descriptions *of* linguistic conventions and descriptions *about* these conventions. The suggestion that 'meaning is use', and that, say, use of psychological predicates is determined by criterial

behaviours, is a suggestion *about* our language-games. It is, as the eight remarks discussed in **2.4** indicate, a suggestion intended to draw us away from the 'reification' of our language-use: to remind us that what we mean is determined by the complex circumstances surrounding our employment of certain words, and not by the existence of some *thing*, 'expecting', that exists independently of our uses. It is not meant to suggest that, *within* these language-uses, what we mean by, say, 'expecting', or 'desiring', or whatever, is just *some behaviour* that plays a determinative role in our ascription of these terms. For the same behaviour *may*, or *may not*, in different cases, be 'what we call', say, a desire or an expectation: and whether it is depends not on the behaviour itself but on what surrounds the behaviour - our prior and subsequent knowledge of the person concerned, the circumstances in which the behaviour takes place, and various other bases that would figure in explanations and justifications of our judgement in that case. The problem with the behaviourist objection is that it takes too simplistic a view of 'use'. Wittgenstein might suggest that it labours under the same tendencies and preconceptions that lead to reification: the idea that there must be *some single, fixed, distinct thing* – an object, a calculus, etc. - that determines what we mean and when we mean it. Such a view approaches Wittgenstein's notion of criteria as though the criteria were these fixed, distinct things that constituted meaning. But Wittgenstein's whole notion of criteria and use is intended, on our view, to suggest that this approach to meaning is misguided.

In this and the previous chapter, we have arrived at and defended a view of criteria, the 'inclusive view', which can be credibly imputed to Wittgenstein and which provides a *prima facie* opposition to the attempt, exemplified by Searle's claims, to distinguish between the ontology of consciousness and its linguistic conditions; and to the further attempt, exemplified by Block, to suggest that consciousness may not be captured by our language. In

short, the 'inclusive' view of criteria provides a model for the contention that an object is defined – that is, its epistemic and ontological characteristics are determined – by our language-use, such that language subsumes ontology, and the idea of states of our facts about consciousness existing outside our language-use is opposed. In Chapter Four, however, we will look at a second sense of ineffability that might arise, on Wittgenstein's view, *within*, rather than *beyond* our language-use.

4 - Seeing the Light: A New Ineffability

The remainder of our discussion will be concerned with identifying a line of thought in Wittgenstein's later corpus which suggests that despite his opposition to the separation between 'being' and 'saying' and, *a fortiori*, the possibility of things (entities, objects, states, processes) or facts about those things existing beyond our language-games, he was *not* straightforwardly opposed to the idea of consciousness, nor to its ineffability. What is more, we will suggest that there are important affinities between his notion of ineffable consciousness and conceptions of qualitative consciousness asserted by Block and others. In identifying this line of thought in Wittgenstein's works, we will draw primarily on the discussion of 'sight' in *Remarks on Colour*, and connect this to David Bell's (1996) discussion of the conscious subject in the later Wittgenstein.

A recurring subject in the *Remarks on Colour* (Wittgenstein 1977) is the nature of 'sight' or 'seeing'. For instance, in §164, §165 and §319, Wittgenstein remarks on the distinction between psychology's capacity to describe the phenomena of colour-blindness - a particular *case* of 'seeing' - and its capacity to describe “vision in general” (§165).

According to Wittgenstein, describing colour-blindness is a matter of demonstrating, e.g., “ways in which the colour-blind person *deviates* from the normal” (RC[15]§165), “what someone who is...colour-blind *cannot* learn” (RC§164), and “the reactions of the colour-blind person which differentiate him from the normal person”; though “not *all* of the colour-blind person's reactions, for example, not those that distinguish him from a blind person” (RC§319). Describing these deviations provides a useful definition of colour-blindness: it is a case of seeing with *these* differences. By the same token, the suggestion (e.g. by psychologists) that there are instances of colour-blindness (that 'there are human beings who are colour blind') is a meaningful and informative one; we learn from it that there are some people who see, with these differences.

However, attempts to define “vision in general” (RC§165) by the same means, and to employ such definitions in language, such as in the assertion that there are 'human beings who see' (RC§328), do not have the same success. Whereas Wittgenstein suggests that “in order to describe the phenomenon of...colour-blindness, I need only say what someone who is...colour-blind *cannot* learn”, he goes on to say that “in order to describe the 'phenomenon of normal vision' I would have to enumerate the things we *can* do” (RC§164). That is to say, to define 'seeing' in this way one need not just delineate a subset of things that a certain group of us cannot do, cannot learn, or do differently, but must enumerate all of the behavioural criteria that distinguish our case from the case of somebody who cannot see.

But where the subset of criteria for 'colour-blindness' is limited, all 'the things we can do' because of vision – that is, the behavioural criteria of vision, the cases in which we react differently from the blind in virtue of our ability to see – comprise an immense, potentially

15 *Remarks on Colour*

indeterminate category of judgments, reactions, dispositions, explanations, justifications and so on. We want to suggest that it is part of Wittgenstein's view that the contents of this category are *so* wide, diffuse and disjunctive that simply enumerating them fails to pick out an informative conception of 'seeing' over and above these particular cases of seeing. Thus, in RC§316, Wittgenstein suggests that when we try to pinpoint the nature of our conscious visual experience of the world, the answer is unacceptably broad; "'Well, all that' accompanied by a sweeping gesture." Attempting to define 'seeing' gets us no further than this gesticulating towards a vast array of particular visual experiences, or visual experiences of particular kinds; we cannot excise from this myriad array any informative notion of 'seeing' itself. So, in RC§165, Wittgenstein implies that defining 'seeing' by contrast with 'not seeing' (the way we contrasted colour-blindness with seeing) leads to an overly diffuse definition: "But couldn't [one] also describe the ways in which normal vision deviates from total blindness? We might ask: who would learn from this?" (RC§165).

Of course, Wittgenstein is not wont to deny that we may talk about seeing in useful ways. It is a large part of his point elsewhere that we can identify, with certainty, *cases* of 'seeing' on the basis of behavioural criteria. We can make *some* statements about 'seeing', such as that it is an ability that allows us to distinguish ripe fruits from unripe fruits (RC§324), to navigate our environments successfully, and so on. Even the congenitally blind can use these kinds of statements (RC§319).

What Wittgenstein *is* concerned to suggest, though, is that employments of the predicate 'seeing' in our language fail to fully capture a sense of seeing. He repeatedly stresses in RC§165 and elsewhere (RC§319, RC§328, RC§331) that employments, like for example, 'there are human beings who see', are uninformative - "to whom would th[ey]

communicate anything?" (RC§331) - and that neither the blind, nor indeed the sighted, can 'learn' what seeing is as a result of mastering the uses of the notion of 'seeing' in our language-games. This, perhaps, is what is meant by the final point in the *Remarks*: "If we introduce the concept of knowing into this investigation, it will be of no help" (RC§350). We can certainly know *that* somebody sees, but coming to know *what* seeing is like through description of cases is something that neither the blind nor the sighted can do. In RC§291, he points to a contrast with other concepts, such as higher mathematics, which we *do* learn via description; concepts whose description fully captures their definition. He expresses this contrast here in terms of cases of "knowledge by description" versus "knowledge by acquaintance". It is only by *seeing* that we grasp the full sense of uses descriptions of 'seeing' in our language-games; these descriptions *presume* an acquaintance with 'seeing' (as does, for instance, statements about 'colour-blindness') but they do not serve to convey, capture or define 'seeing'.

Some further examples are instructive. Wittgenstein draws on the distinction between a game of chess in contrast to 'games in general' (RC§282). The example follows immediately from the contrast between "colour-blindness...an inability", and "seeing...the ability" (RC§281). If we say to someone who understands chess that 'B can't play chess', he evidently understands that; but it is less clearly understood by someone "absolutely unable to learn any game" (RC§282). In the remark immediately following, Wittgenstein begins by asking, "Does everything I want to say here come down to the fact that the utterance "I see a red circle" and "I see, I'm not blind" are logically different?" (RC§283).

These remarks are not unambiguous; but what we might plausibly take away from them is a sense that a *particular* case (e.g., colour-blindness, chess) are picked out and

defined against the background of a broader category (seeing, games) by contrastive comparison with other cases in that category. And notions like “seeing a red circle”, which unlike “I see, I'm not blind”, describe '*intrinsic'* rather than contrastive notions, are not able to be picked out in this way[16]. More specifically, derivative *cases* of sight, such as colour-blindness, exploit, and make sense in terms of, the wider background of 'normal vision'; they can be individuated in virtue of criterial contrasts with ordinary cases of sight – in virtue of particular deviations, things we cannot do or cannot learn (RC§164, RC§165, RC§319): colour-blindness is thus defined as a normal case of seeing with a subset of distinguishing (criterial) differences.

Seeing itself, though, is an 'intrinsic' notion, and has no wider background against which it can define itself. Attempting to contrast it merely against the complete absence of sight results in an unhelpfully diffuse attempt at definition which fails to pick out the kind of thing 'seeing' is; and our employment of notions of seeing, correlatively, do not capture its sense either. It is in this sense that 'seeing' cannot be fully defined nor conveyed by description that we might consider it *ineffable*.

As an aside, it is worthwhile considering the sense in which our interpretation here may be corroborated by an otherwise rather unaccountable feature of the *Remarks on Colour*. In the midst of the passages quoted, in RC§317, we find a lengthy observation which concerns not sight but faith in God. Here, Wittgenstein suggests that the question, “Where did everything that I see come from?”, when asked by a religious person, though it has the form of a request for a causal explanation, is in fact the expression of an “attitude towards all

16 Of course 'seeing a red circle' may have criteria, but these would not help us discern *what it is like* to 'see a red circle'; they would fail to convey the sense of this phrase.

explanations". He goes on to suggest, of this disconnect between form and meaning, that what matters is not "the words one uses" but the difference they make at various points in one's life. These remarks echo some of Wittgenstein's conceptions of faith elsewhere, notably in *Lectures on Religious Belief.* Here he is recorded as suggesting that the difference between a person of faith and an agnostic person "might not show up at all in any explanations of the meaning" (1966:53) of their terms, because it is not a question of particular beliefs, but of "regulating for all in [the religious believer's] life" (ibid:54).

Without knowing more about the circumstances in which the *Remarks on Colour* were composed, it would be presumptuous to venture any strident suggestions about Wittgenstein's motives in interposing faith and sight. But it is worth at least pointing out some broad, but compelling, analogies between these two notions. In both cases there is a sense that something may, because it occupies a kind of 'global' position in our engagement with the world – because it 'regulates for *all* in our lives' – fail to register in more particular cases of our language-use. 'Seeing', as we have said, cuts across our entire spectrum of cases of visual perception, and so is not distinguishable in terms of a subset of behavioural criteria, as are particular kinds or cases of seeing. Likewise, for Wittgenstein, 'believing' in God, as a kind of 'attitude' or 'Weltanschauung', supervenes across, and is *presumed* rather than picked out by, more particular beliefs, more specific explanations of the world. Indeed, it is useful, perhaps, to think of 'sight' as a kind of perceptual 'attitude' – a broad, all-encompassing orientation to the world which supervenes across particular perceptual engagements with it.

In any event, seeing is not consciousness, and it is the ineffability of consciousness with which we are concerned. In the following section we will consider some remarks from

David Bell on the place of consciousness in Wittgenstein's later work which connect with and clarify our remarks on 'sight'.

4.1 Ineffable Consciousness

Through our appraisal of Malcolm's (1982) discussion in **3.4.3**, we have already obliquely referred to the notion that idealism is present in Wittgenstein's later thought. The purpose of raising Malcolm's article was of course to deny that Wittgenstein's view of criteria entails idealism about truth, and we do not wish to depart from that position. But it is nonetheless true that some points made in the course of arguments for idealism in Wittgenstein's later view, and particularly the arguments put by David Bell (1996)[17], demonstrate the connection between sight and consciousness, and explain in greater detail kind of ineffability we wish to attribute to Wittgenstein's view of consciousness.

Bell agrees with our suggestion in Chapter Two, that an enduring concern for Wittgenstein was undermining the Cartesian picture of the mind, and obviating the concerns to which it leads. He identifies Wittgenstein's discussion of the place of the conscious subject in nature as a corollary of this rejection of the Cartesian picture. In that very familiar picture, Bell suggests, the conscious mind is construed as an "object for description like any other" (1996:155), as something amenable to observation, knowledge and definition. But our knowledge of our inner states of mind is for Descartes both incorrigible and non-shareable. It is 'given' to us in introspection and, as such, is certain; and as it is not so given to others, "another person cannot have *that* knowledge of [our inner states] which is enjoyed by the subject whose contents of consciousness they are" (ibid:156). The world is for each of us divided into the 'inner' and the 'outer', what William James called the "great splitting of the

17 But see also Williams, 'Wittgenstein and Idealism' (1973).

universe into two halves" (ibid:156) – those facts and states which we observe directly, in our own minds, and in the world, which we observe only 'mediately', *through* our minds (see also Overgaard 2007, Moran 2001).

But Bell suggests that, for Wittgenstein, our own minds, and so our experience of 'being conscious' and 'having consciousness', cannot be picked out, individuated, as a distinct object or entity. His argument runs along precisely the same lines as our suggestion that 'sight' cannot be usefully picked out. Because seemingly *every*, explanation and judgment of, every reaction and response to, the world is made possible by consciousness, enumerating the criteria for consciousness (*per impossible*) gives us no more than an unacceptably broad picture of the world itself. So, Bell suggests that, when asked to consider what my own consciousness of the world consists in, I can only say, "whatever it is that is conscious of *all this*" (1996: 167), gesturing out to the world and everything in it. Consciousness is whatever it is that makes 'all this' possible. Equally, though, when asked what this totality, 'the world', consists in, I can only repeat the same gesture – 'all this'. The two are indistinguishable. Though Bell is not focused on *Remarks on Colour* here, it is striking that his words echo Wittgenstein's in RC§316: "what...am I now seeing?..."Well, *all that*" accompanied by a *sweeping gesture*" (my italics).

Moreover, says Bell, "neither the world as a whole, nor any locus of genuine subjectivity is successfully individuated by such hand-waving" (1996:167). Here, he points to Wittgenstein's enduring allegiance to Fregean ideas of definition, whereby successfully defining something is made possible by *individuating it*, distinguishing it from what Bell calls its "neighbours": according to Frege, "if we are to use the symbol *a* to signify an object, we

must have a criterion for deciding in all cases whether *b* is the same as *a*" (ibid). To be definable, there must be criteria contrasting from 'neighbouring' notions.

But 'Consciousness', like 'seeing', is 'neighbour-less'. Though particular kinds or cases of conscious experience may be contrasted with each other in virtue of criterial distinctions, the subjective consciousness that underpins the entirety of these experiences, like the 'sight' that underpins the entirety of visual experiences, cannot be contrasted with anything. Consciousness is not one experience of the world among others; it is the grounds on which all of these experiences are made possible. It is, therefore, not identifiable with any one case or kind of experience but with the "endless multiplicity" (Wittgenstein 1992:81) of conscious experiences; it is, so to speak, identifiable not with *some*thing but only with *every*thing. It 'regulates for all' of our criteria-governed conscious experiences; it is in virtue of being conscious that we have these language-games at all; and for that very reason, by dint of its ubiquity, it cannot itself be picked out and made subject to a specific subset of criteria that tell us 'what kind of thing' it is. It is, therefore, unclear *what* it is, and what we can know (or 'learn') about it. Bell boldly – but persuasively – extrapolates that, for the later Wittgenstein, the conscious subject:

> "[I]s not something that, as it were, *has* a place in nature; my own consciousness is not something I ever come across *in* the world; it is not the sort of thing that I can refer to, identify, describe, have acquaintance with, or knowledge of. And its elements are not any kind of 'things' – whether events, processes, objects, properties, or facts...Genuine subjectivity...is that to which the entire conceptual machinery of objectivity is ultimately inapplicable" (1996:156).

In some respects this puts things too strongly for us; we do not want to deny that *reference* to consciousness, any more than sight, is impossible. But an *informative* reference to consciousness – one which tells us, in a non-trivial way which does not 'presume'

acquaintance with the subject, what consciousness 'is like' – is not obviously possible. Descriptions of consciousness, as Nagelian 'what-it-is-like'-ness (e.g. 1974) famously exemplifies, presume, and cannot teach us, what consciousness is like.

We need not agree with Bell's intimations that Wittgenstein's later arguments are a retention of the solipsism of his early view; exploring or justifying this claim is beyond our purposes. But the striking similarities between his exposition and ours suggest that, independent of his conclusions, he has struck upon the same vein in Wittgenstein's view, and has persuasively shown it to be a wider vein than merely 'seeing'. In doing so he makes explicit what is already implicit in the *Remarks on Colour;* in RC§314, for instance, Wittgenstein suggests that a simplified way of talking of the 'world of consciousness' might be 'what I am now seeing'; and in RC§319, he writes, “Do the blind know what it is like to see? But do the sighted know? *Do they also know what it's like to have consciousness*?”. Bell's approach, in combination with our own investigations, stand as good grounds to impute to the later Wittgenstein the view that consciousness is the psychological analogue of sight – it is to 'experience' what 'seeing' is to visual perception; and the view that both are, in important ways, not amenable to the ordinary boundaries of criterial definition, and in that sense ineffable. In the next section we will bring this notion of consciousness to bear on the debate between Wittgensteinians and cognitivists.

4.2 Implications: A Profound Ambivalence

We can now explore some of the implications of discerning this second kind of ineffability for our delineation of Wittgenstein's opposition to the cognitivists. We will begin by situating this 'new ineffability' between the Wittgensteinian and cognitivist claims of

Bennett, Hacker, Block and Searle. We will conclude that this new position finds its place *between* the idea of extra-conceptual states or facts of consciousness as espoused by Block and others, and the denial or dissolution of the ineffability of consciousness espoused by Hacker and Bennett among others. We will then make some more detailed claims about the relationship between Wittgenstein's view and Block's position in particular; point to some of the broader implications of our reading of Wittgenstein for the relationship between his work and cognitive science generally; and finally, consider the implications of our reading for a general understanding of Wittgenstein's understanding of the mind.

It has been some time since Chapter One, so let us begin by very briefly reconsidering Hacker and Bennett's Wittgensteinian criticisms, and the retorts offered to them. Their two relevant objections were what we called the 'monolith' objection and the 'mereology' objections. The 'monolith' objection (notwithstanding its being associated with a rather wayward conception of 'qualia') claims, at base, that each distinct psychological state is represented by a range of behavioural (including, of course, verbal) criteria[18], and that conglomerating them under a kind of state, 'consciousness', constituted an illicit departure from criteria-governed language and a reification of the 'mind' in the manner of Descartes. In the parlance of cognitive science, Hacker and Bennett's approach can be associated with a rejection of *intransitive* consciousness, consciousness as a 'thing' or 'state', and a privileging of 'transitive' consciousness, consciousness *of* particular things *in the world*, and consciousness *that* particular states of affairs in the world obtain. Their claims suggest that it is not possible, indeed not grammatically sensible to 'peel away' the intransitive phenomenon/a of consciousness from particular experiences of the world.

18 I do not mean to imply, of course, that two states could not share some of the same criteria.

The 'mereology' objection is a further outcome of Hacker and Bennett's emphasis on criteria. Conceiving of consciousness as a property *of brains* – and conceiving it, therefore, as a kind of discrete phenomenon which could be, theoretically at least, excised and studied – constituted an ignorance of conditions of use of the notion of 'consciousness', and, in turn, the connection between those conditions of use and what consciousness *is*. The result of both of these objections, Hacker and Bennett suggested, was that the 'mysteriousness' of consciousness was a mere illusion; that when reunited with their proper circumstances of use, notions of consciousness were non-mysterious and unproblematic.

Our appraisal of these objections, and the cognitivist replies to them led us to Wittgenstein's notion of criteria. In Chapters Two and Three, we staked Wittgenstein's opposition to Block and Searle's 'being'/'saying' distinction and the subsequent ineffability arguments made by Block on the 'inclusive' view of criteria. We suggested that because the 'inclusive' view suggested a relationship between use (including epistemic use) of a predicate and that predicate's meaning, it provided a *prima facie* basis on which to dismiss the 'being'/'saying' distinction and with it, the suggestion that there could be states of or facts about consciousness that could be 'extra-grammatical' (beyond our language). So, Block's suggestion that there were states of or facts about consciousness *beyond* our criteria-governed language-games was, in Wittgenstein's view, illicit. Particularly in Chapter Three, we suggested that criteria constitute the epistemic and ontological dimensions of their associated phenomena; so suggesting that there were epistemic or ontological states of affairs *beyond* our criteria was, in Wittgenstein's view, incoherent.

However, in this chapter we have explored the possibility that Wittgenstein's view was consistent with, and perhaps implied, ineffability of a different kind. On this view, our

inability to describe and define something arises not because it exists *beyond* our criterial language-games, but because it, so to speak, exists *throughout* them. Because consciousness underpins, or supervenes across, the whole breadth of the innumerable and perhaps unlimited range of behavioural criteria that characterise our engagements with the world, it is not possible to distinguish it by enumerating some particular subset of criteria with which it may be usefully defined and described (and correlatively, uses of the notion of 'consciousness' in our language do not convey, but *presume acquaintance with* the sense of this term). To put it impressionistically, because it shows up *every*where, it does not show up *any*where in particular; the problem is not the *absence* of criteria so much as their 'over-abundance'.

This position occupies an interesting middle-ground between the claims made by Bennett and Hacker, and the opposing claims made by Searle and, in particular, Block. On the one hand, Wittgenstein may rightly be taken as suggesting, and using the concept of criteria to demonstrate, that we cannot sensibly postulate a state of affairs about which there are determinate facts outside of our language-use. For it is criteria which constitute the 'kind of thing' something is and what and how we know about its nature. Accordingly, it is quite so for Wittgenstein that we should, in seeking to grasp 'consciousness' *itself*, get no further than a wholesale description of the totality of our diffuse psychological engagements with the world, our myriad states of transitive consciousness recognisable by their own subset of behavioural criteria. So the idea of ineffable consciousness founded on the separation of language and ontology which Block appears to propose, is indeed anti-Wittgensteinian.

However, we have suggested in the present chapter that we cannot infer, from the denial of *this* path to ineffable consciousness, that Wittgenstein's view provides a *tout court* dismissal of the idea of qualitative consciousness, *nor* of the idea that conscious experience

was in some sense ineffable; and nor yet can we take him as suggesting that, in its right conditions of use, consciousness was entirely unproblematic and non-mysterious. Indeed, Wittgenstein's view seems to endorse, or at the very least be consistent with, the notion of a second kind of ineffability for which consciousness is a candidate; and the notion that, moreover, the inability to excise a notion of consciousness from more particular cases is not a rebuttal of consciousness' mysterious ineffability - it *is* that ineffability. To this extent, Hacker and Bennett's reading of Wittgenstein as straightforwardly opposing the cognitivist's position is mistaken.

Consciousness is not, for Wittgenstein, an extra-criterial 'thing', a distinct 'state' or 'entity' about which there are determinate facts. But this does not mean that it is *ipso facto* 'nothing at all'. Where both the Wittgensteinians *and* the cognitivists were wrong is in supposing that it had to be one or the other. In Wittgenstein's view, it is neither: it is, we might say, a kind of 'everything' about which we cannot speak. In seeking to capture the profound ambivalence of this 'new ineffability', we might avail ourselves of Wittgenstein's own phrase: consciousness is "not a something, but not a nothing either!"[19](PI§304).

The implications of Wittgenstein's position for the cognitivist's aspirations are compelling. Wittgenstein's view offers the possibility of espousing the ineffability of consciousness in a way which, unlike Block's view, is not premised on the separation of language and reality. Recall Canfield's (2009) critique of Block's (2007) article on Wittgenstein and qualia. The essence of this critique is that given an 'inclusive' view of criteria, Block's suggestion that there are ways of seeing something which are inexpressible

19 I am not suggesting this is precisely what Wittgenstein *meant* in the context of this remark in the *Investigations*; only that this phrase is fortuitously apt for the point I am making here.

via criteria is nonsense: an inexpressible difference is no difference at all. To put it another way, once a difference in the ways we perceive something, to speak, 'sinks below' our uses of words – once it no longer registers in the way we report our perceptual experiences – then it no longer makes sense to speak of there *being* a difference at all. Therefore, there *is* no qualia inversion in Block's dangerous scenario.

If the reading of Wittgenstein that has come to light in the present chapter is correct, the absoluteness of Canfield's conclusion must be augmented. Insofar as inverted qualia presupposes *extra*-grammatical facts or states, it is indeed, in Wittgenstein's view, nonsensical, by dint of the thoroughgoing role of criteria. But it would be oversimplifying Wittgenstein's view to conclude, on this basis, that *any* notion of qualia, and even of inverse qualia experiences, is completely denied by Wittgenstein – what is denied is their possession of the criterial framework necessary to ascribe to them epistemic or ontological characteristics (cf Moore & Sullivan 2003).

And indeed, Wittgenstein's view, insofar as it suggests the ineffability of an intrinsic sense of consciousness, has compelling affinities with Block's notion of qualia (see Block 1980, 1990, 2007). We might compare Block's suggestion (e.g. 2007) that the qualitative content of our experience cannot be 'fully captured' by language with Wittgenstein's sense that the ineffability of consciousness prevents our giving informative descriptions or definitions of it. In fact their positions seem very close. Neither Wittgenstein nor Block present views so strong as to deny our capacity to *refer* to consciousness. Just as Block admits that we can use phrases like 'the quale I get when I see green things' (2007:62), Wittgenstein admits that we might say, for instance, 'there has taken place in me the mental process of remembering' (PI§306); but for neither of them are such statements genuinely

informative about the nature of what is being described. And just as Block compares this to statements like 'Napoleon is buried in Paris', whose meaning may be 'fully captured' because they do not invoke our own “ways of thinking of Napoleon and Paris”(2007:62), Wittgenstein, in our view here, suggests that, for instance, higher mathematics may be fully captured by descriptions of it (RC§291).

More generally, we can only raise, and unfortunately not discuss at length here what are some obvious and tantalising parallels between our view of Wittgenstein as suggesting that we might fail to provide informative descriptions of consciousness, and that we cannot 'learn' what consciousness is by description, and well-known suggestions to this effect from Frank Jackson (e.g., 1982, 1986) and Thomas Nagel (e.g., 1974; 1986). Both Jackson's sense that we cannot convey in ordinary factual terms the experience of consciousness, and Nagel's suggestion that the identification of objective characteristics that is instrumental to the scientific study of phenomena is problematic in the case of consciousness, are resonant with the view that we have imputed to Wittgenstein in this chapter. Though it is perhaps worth stressing a *dis*analogy with Jackson's 'Mary' arguments, which mirrors Wittgenstein's putative response to qualia inversion cases: Wittgenstein may reject the idea that there are determinate 'facts' about consciousness that cannot be grasped by ordinary, objective criteria. If criteria exhaustively determine our epistemic language-games, this would perhaps be nonsensical. But Wittgenstein's view *does* seem very much consistent with the suggestion that our ordinary epistemic criteria cannot apply to consciousness. Recall, again, the concluding statement of *Remarks on Colour* that “[i]f we introduce the concept of knowing into this investigation, it will be of no help” (RC§350).

It is worth noting that insofar as it seems to go *further* than Block and, perhaps, Jackson, Wittgenstein's conception of ineffability, on our reading, is a deep one, and one perhaps as inimical to some of the cognitivist's aspirations as it is to the wholesale denial of the cognitivist enterprise suggested by some Wittgensteinians. Certainly it seems that Searle is mistaken in suggesting that Wittgenstein's view of consciousness is orthogonal, or agnostic with respect to the project of excising and studying consciousness as a distinct, neurally-located phenomenon. Wittgenstein's view, expressed *both* in his model of criteria (what we have called the 'inclusive' view) and his conceptions of sight and consciousness here, is that consciousness is not the kind of thing that can be defined, described, 'taxonomised' in the ordinary ways. Even if his suggestion that we cannot separate consciousness from its criterial manifestations should not be taken as an outright denial of consciousness, it seems that it must be taken as a condemnation of attempts to describe or define consciousness in objective terms - and so as deeply opposed, for instance, to attempts in contemporary cognitive science to identify consciousness with neural correlates, or as a non-physical 'substance' with distinct identifiable properties (see also McGinn 1991; Hopkins 1974; Klagge 1989; Papineau, forthcoming).

In closing, let us consider the broader implications of our conclusions for *Wittgenstein's* view. Recall in Chapter One (pp.15-16) our suggestion that there were reasonable exegetical grounds to believe Searle's suggestion that the distinction between ontology and language was compatible with Wittgenstein's position. There we quoted, for instance, PI§305, Wittgenstein's response to his interlocutor's accusation of behaviourism: “what gives you the impression that we want to deny anything?...What we deny is that the picture of the inner process gives us the correct idea of the use of the word 'to remember'”.

We might just as easily have quoted a neighbouring remark, PI§308, wherein Wittgenstein suggests that a primary error made in our attempts to understand consciousness is that we presume to "talk of processes and states and leave their nature undecided...But that is just what commits us to a particular way of looking at the matter. For we have a definition concept of what it means to know a process better".

The meaning of these remarks, and the profundity of the position they express, can now be made out. Clearly, Wittgenstein does not mean to deny that there *is* a conscious, subjective, inner life. What he means to deny is the Cartesian sense that we can speak about it as a distinct phenomenon, subject to the same ontological categories as the rest of nature; and moreover, as something amenable to the ordinary ontological categories of state and process (cf Overgaard 2007). Such characteristics can only be ascribed because of settled uses of criteria, and this is precisely what consciousness lacks. Wittgenstein is suggesting that, in using psychological predicates, we are not defining or describing our inner lives; and that our inner lives cannot be talked about in the same manner as ordinary objects in the world. That inner processes are spoken of in PI§305 as *pictures* is significant. For 'the inner' is a picture in the genuinely Wittgensteinian sense – though we might dimly intuit or envisage it, we cannot sensibly employ it in our language-game, cannot apply to it the criterial structures necessary for such employment. The view we have imputed to Wittgenstein in this chapter is of a kind with this broader outlook, and provides support for these contentions. What is more, it vindicates our suggestion that criteria can be understood as part of an enduring preoccupation with the refutation not of consciousness but the Cartesian approach to it.

Conclusion

Our discussion began with an overview of the dissonance between Wittgensteinians and cognitive scientists, grounded in part on disparate interpretations of Wittgenstein's own view of criteria. In Chapters Two and Three, our subsequent investigation of Wittgenstein's view vindicated the Wittgensteinian reading; the 'inclusive' view suggested that Wittgenstein was opposed to the 'being'/'saying' distinction and to that extent opposed to a notion of consciousness as some entity or object beyond the reach of our language. However, in Chapter Four, we have suggested that there is more to this story: that a notion of 'ineffable consciousness' could be imputed to Wittgenstein which was consistent with his robust notion of criteria but importantly similar to the suggestions about qualitative consciousness made by Block and others. The chief implication of our interpretation may be summarised thus: even when Wittgenstein's view is taken at its strongest – even in light of the exegetical accuracy and *prima facie* plausibility of his critique of the separation of 'being' and 'saying' – his view may not be understood as providing an outright opposition to the notion of ineffable consciousness, but as being consistent with, and evidently endorsing, a version of that notion which is not premised on the 'being'/'saying' distinction.

What seems clear, at any rate, is that the pattern of hostility and estrangement between Wittgensteinians and cognitive scientists is unwarranted, and is born of a simplistic understanding of Wittgenstein's view from both groups. Insofar as it can be made to shed light on and to some extent endorse a novel notion of 'ineffable consciousness', Wittgenstein's position should be of considerable interest to cognitive scientists and philosophers of mind engaged in contemporary debates about consciousness. And equally, insofar as it contributes to debates about consciousness, rather than merely obviating them, Wittgenstein's view

obliges *Wittgensteinians* to engage with these debates and look beyond the licence to abstain which criterial arguments are often thought to provide.

References

• Albritton, R. (1959) 'On Wittgenstein's Use of the Term 'Criterion' *Journal of Philosophy* 56 (22):845-857

•Arrington, R. 'Thought and Its Expression' in Schroeder, S. (2001) *Wittgenstein and Contemporary Philosophy of Mind*, New York: Palgrave MacMillan

•Bell, D. (1996) 'Solipsism and Subjectivity' *European Journal of Philosophy* 4 (2):155-174

•Block, N. (2007) 'Wittgenstein and Qualia', *Philosophical Perspectives* 21 (1):73-115

•Block, N. (1980) 'Troubles with Functionalism' in *Readings in the Philosophy of Psychology*, Volume 1, ed: Ned Block, Cambridge, MA: Harvard University Press, 268-305

•Block, N. (1990) 'Inverted Earth' in *Philosophical Perspectives*, 4, ed: J. Tomberlin, Atascadero, CA: Ridgeview

•Block, N. (1995) 'On a Confusion About the Function of Consciousness' *Behavioral and Brain Sciences* 18:227-47

•Budd, M. (1989) *Wittgenstein's Philosophy of Psychology* London: Routledge

•Canfield, J. (1974) 'Criteria and Rules of Language' *Philosophical Review* 83(1):70-87

•Canfield, J. (2009) 'Ned Block, Wittgenstein and the Inverted Spectrum' *Philosophia* 37 (4):691-712

•Dennett, D. (1991) *Consciousness Explained,* London: Penguin

•Fodor, J. & Chihara, C. (1965) 'Operationalism and Ordinary Language: A Critique of Wittgenstein' *American Philosophical Quarterly* 2(4):281-295

•Gibbs, B. (1969) 'Putnam on Brains and Behaviour' *Analysis* 30:53-55

•Glock, H. J. 'Wittgenstein and Quine: Mind, Language and Behaviour' in Schroeder, S. (2001) *Wittgenstein and Contemporary Philosophy of Mind*, New York: Palgrave MacMillan

•Hacker, P. (1972) 'Are Transcendental Arguments a Version of Verificationism?' *American Philosophy Quarterly* 9 (1):75-85

•Hacker, P.M.S. (1986) *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, Revised edition, Oxford: Clarendon Press

•Hacker, P.M.S. (1990) *Wittgenstein: Meaning and Mind, Volume 3 of an Analytical Commentary on the Philosophical Investigations*, Oxford: Blackwell

•Hacker, P.M.S. (2007b) *Human Nature: The Categorial Framework* Oxford: Blackwell

•Hacker, P.M.S. & Bennett, M. (2003) *The Philosophical Foundations of Neuroscience*, Oxford: Blackwell

•Hacker, P.M.S. & Bennett, M. (2007) *Neuroscience and Philosophy – Brain, Mind, and Language,* New York: Columbia University Press

•Hanfling, O. 'Consciousness: The Last Mystery' in Schroeder, S. (2001) *Wittgenstein and Contemporary Philosophy of Mind*, New York: Palgrave MacMillan

•Heil, J. (1981) 'Does Cognitive Psychology Rest on a Mistake?' XC:321-342

•Hopkins, J. (1974) 'Wittgenstein and Physicalism' *Proceedings of the Aristotelian Society* 75:121-146

•Hunter, J. (1977) 'Wittgenstein on Inner Processes and Outward Criteria' *Canadian Journal of Philosophy* 7 (4):805-817

•Jackson, F. (1982) 'Epiphenomenal Qualia' *Philosophical Quarterly*, 32:127-136

•Jackson, F. (1986) 'What Mary didn't know'. *Journal of Philosophy*, 83: 291-5; repr. in O'Connor, T. & Robb, D. *Philosophy of Mind: Contemporary Readings*, London: Routledge

•Kelly, M. (1991) 'Wittgenstein and Mad Pain' *Synthese* 87(2):285-294

•Kenny, A. 'The Homunculus Fallacy', repr. in Hyman, J. (1993) *Investigating Psychology: Sciences of the Mind After Wittgenstein,* Londong: Routledge

•Kiverstein, J. (Forthcoming) 'Wittgenstein, Qualia and the Autonomy of Grammar' (available at http://www.philosophy.ed.ac.uk/contact/publications.html; last accessed 15/06/10)

•Klagge, J. (1989) 'Wittgenstein and Neuroscience' *Synthese* 78:319-43

•Koethe, J. (1977) 'The Role of Criteria in Wittgenstein's Later Philosophy' *Canadian Journal of Philosophy* 7 (3):601-622

•Lycan, W. (1971) 'Noninductive Evidence: Recent Work on Wittgenstein's "Criteria"' *American Philosophical Quarterly* 8(22):109-125

•Malcolm, N. (1982) 'Wittgenstein and Idealism' *Royal Institute of Philosophy Lectures* 13:249-267

•McGinn, C. (1991) *The Problem of Consciousness,* Oxford: Basil Blackwell

•McGinn, C. (1989) 'Can We Solve The Mind-Body Problem?' *Mind* 98(391):349-66

•McGinn, C. (1999) *The Mysterious Flame: Conscious Minds in a Material World*, London: Basic Books.

•McGinn, M. (1997) *Wittgenstein and the Philosophical Investigations,* London: Routledge

•Moore, A.W. & Sullivan, P. (2003) 'Ineffability and Nonsense' *Aristotelian Society Supplementary Volume* 77 (1):169–193

•Moran, R. (2001) *Authority and Estrangement: An Essay On Self-Knowledge,* Princeton: Princeton University Press

•Nagel, T. (1974) 'What Is it Like to Be a Bat?', *Philosophical Review* 83:435-56.

•Overgaard, S. (2007) *Wittgenstein and Other Minds: Rethinking Subjectivity and Intersubjectivity with Wittgenstein, Levinas and Husserl*, London: Routledge

•Papineau, D. (forthcoming) 'Private Language and Phenomenal Concepts' (available at http://www.kcl.ac.uk/schools/humanities/depts/philosophy/people/academic/papineaud/articles.html; last accessed 15/06/10)

•Pollock, J. (1967) "Criteria and Our Knowledge of the Material World," *Philosophical Review*, 76:28-60

•Proudfoot, D. (1997) 'On Wittgenstein On Cognitive Science' *Philosophy* 72:189-217

•Putnam, H. (1957) 'Psychological Concepts, Explication and Ordinary Language' *Journal of Philosophy* 54:94-99

•Rey, Georges (2003) 'Why Wittgenstein Ought to Have Been a Computationalist (and What a Computationalist Can Gain from Wittgenstein)', *Croatian Journal of Philosophy* 3 (9):231-264

•Rorty, R. (1971) 'Verificationism and Transcendental Arguments' *Nous* 5(1):3-14

•Searle, J. (2002) *Consciousness and Language,* Cambridge: Cambridge University Press

•Schroder, S. (2001) *Wittgenstein and Contemporary Philosophy of Mind*, New York: Palgrave MacMillan, 'Preface'

•Schulte, J. (1993) *Experience and Expression: Wittgenstein's Philosophy of Psychology* Oxford: Oxford University Press

•Schwyzer, H. (1973) 'Thought and Reality: The Metaphysics of Kant and Wittgenstein' *Philosophical Quarterly* 21:193-206

•Shoemaker, S. (1982) 'The Inverted Spectrum' *Journal of Philosophy* 79:357-381

•Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen

•Westphal, K. (2005) 'Kant, Wittgenstein, Transcendental Chaos' *Philosophical Investigations* 28 (4):303-323

•Williams, B. (1973) 'Wittgenstein and Idealism', *Royal Institute of Philosophy Lectures* 7:76-95

•Wittgenstein, L. (1966) *Lectures and Conversations on Aesthetics, Psychology and Religious Belief,* ed: Cyril Barrett, Oxford: Basil Blackwell

•Wittgenstein, L. (1967) *Zettel*, eds: G.E.M. Anscombe & G.H. Von Wright, trans: G.E.M. Anscombe, Oxford: Basil Blackwell

•Wittgenstein, L. (1968) 'Notes for Lectures on Private Experience and Sense-Data', trans: Rush Rhees, *The Philosophical Review*, 77 (3):275-320

•Wittgenstein, L. (1969) *On Certainty*, eds: G.E.M. Anscombe & G.H. Von Wright; trans: G.E.M. Anscombe & D. Paul, Oxford: Blackwell

•Wittgenstein, L. (1974) *Philosophical Grammar*, trans: Anthony Kenny; ed: Rush Rhees, Oxford: Blackwell

•Wittgenstein, L. (1977) *Remarks On Colour*, ed: G.E.M. Anscombe; trans: Linda McAlister & Margarete Schattle, Oxford: Basil Blackwell

•Wittgenstein, L. (1980) *Culture and Value,* eds: G.H. Von Wright & Heikki Nyman; trans: Peter Winch, Oxford: Basil Blackwell

•Wittgenstein, L. (1980) *Remarks on the Philosophy of Psychology* Vol. 1, eds: G.E.M. Anscombe & G.H. von Wright, trans: G.E.M. Anscombe; Vol. 2, eds: G.H. von Wright and H. Nyman, trans: C.G. Luckhardt and M.A.E. Aue, Oxford: Blackwell.

•Wittgenstein, L. (1982) *Last Writings on the Philosophy of Psychology*, Vol. 1, eds: G.H. von Wright & H. Nyman, trans: C.G. Luckhardt and M.A.E. Aue, Oxford: Blackwell

•Wittgenstein, L. (1991) *The Blue and Brown Books*, Oxford: Wiley-Blackwell

•Wittgenstein, L. (2001) *Philosophical Investigations,* trans: G.E.M. Anscombe, Oxford: Basil Blackwell

•Wittgenstein, L. (2001) *Tractatus Logico-Philosophicus*, trans: David Pairs & Brian McGuinness, London: Routledge

•Wittgenstien, L. (1992) *Last Writings on the Philosophy of Psychology*, Vol. 2, eds: G.H. von Wright & H. Nyman, trans: C.G. Luckhardt and M.A.E. Aue, Oxford: Blackwell

•Wolgast, E. (1964) 'Wittgenstein and Criteria' *Inquiry* 7(1-4):348-366.