Fall 2010

Expressivist theories of first-person privilege

Nathanial Shannon Blower
University of Iowa

EXPRESSIVIST THEORIES OF FIRST-PERSON PRIVILEGE

by

Nathanial Shannon Blower

An Abstract

Of a thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Philosophy in the Graduate College of The University of Iowa

December 2010

Thesis Supervisor: Professor David Stern

ABSTRACT

This dissertation scrutinizes expressivist theories of first-person privilege with the aim of arriving at, first, a handful of suggestions about how a 'best version' of expressivism about privilege will have to look, and second, a critical understanding of what such an approach's strengths and weaknesses will be. Roughly, expressivist approaches to the problem of privilege are characterized, first, by their emphasis on the likenesses between privileged mental state self-ascriptions and natural behavioral expressions of mentality, and second, by their insistence that an acknowledgment of these likenesses is required in order properly to understand the characteristically singular privilege with which one speaks of one's own mental states. The dissertation proceeds in five chapters whose individual tasks are as follows:

The first chapter sets out the definition of the phenomena of "first-person privilege" in use throughout the dissertation and defends the claim that those phenomena are indeed real and so the philosophical problem of accounting for them is indeed serious. However, there is no presupposition against the possibility of an expressivist account of the phenomena of first-person privilege.

The second chapter sets out the basic motivations informing expressivist approaches to the problem of first-person privilege. Four immediate and significant questions for the expressivist approach are set out. The chapter also considers one 'simple' way of responding to those questions and sets out the most pressing difficulties for a 'simple expressivism'.

The third chapter sets out my view of Wittgenstein as a methodically non-theorizing philosopher, criticizes rival views and, finally, sets out my view of the

Wittgensteinian responses to the four questions set out in chapter two, given my view of him as a philosophical non-theorizer. Many of the later suggestions about a 'best version' of expressivism draw directly on my best understanding of Wittgenstein's own approach to the problem of first-person privilege.

The fourth chapter sets out David Finkelstein's, Peter Hacker's and Dorit Bar-On's responses to the quartet of questions for expressivists about first-person privilege, while flagging a number concerns for each author's approach.

The final chapter condenses and reviews the concerns already raised for the expressivist approaches already canvassed and makes a number of suggestions about the most viable expressivist options for dealing with them. With that in place, the last chapter proceeds to comment on the overall plausibility of the sketch of a 'best-version' of expressivism that emerges. Also, concerns to do with the relationship between expressivism about first-person privilege, epistemological foundationalism, content externalism and the mind-body problem are discussed.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

EXPRESSIVIST THEORIES OF FIRST-PERSON PRIVILEGE

by

Nathanial Shannon Blower

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Philosophy in the Graduate College of The University of Iowa

December 2010

Thesis Supervisor: Professor David Stern

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph. D. thesis of

Nathanial Shannon Blower

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Philosophy at the December 2010 graduation.

Thesis Committee: _____
David Stern, Thesis Supervisor

_____
Richard Fumerton

_____
Ali Hasan

_____
Gregory Landini

_____
David Rudrauf

TABLE OF CONTENTS

INTRODUCTION

I'd like to take just a little space at the outset to say a few words, first, about bow I arrived at the specific project on expressivism that is this dissertation, and second, about the role of the exegetical work on Wittgenstein's later writings that figures in the carrying out of this project.

Expressivist theories of first-person privilege will be most attractive to those who would not only account for a serious kind of first-person privilege but who would do so without appealing to any kind of private/inner access to private/inner objects. Since Wittgenstein's later work on the area, I believe, there's been a strong suspicion that, one way or another, likening privileged mental state self-ascriptions to natural, behavioral expressions of mentality should play an integral role in the banishment of the delusion of private access to private objects, and that while paying due respect to the very real phenomena of first-person privilege.

However, it's often been thought that the price of going expressivist, as it were, about privileged mental state self-ascriptions was that those self-ascriptions would have to turn out looking only superficially like ascriptions, so that the expressivist analysis would show that they weren't, after all, even in the running for truth or falsity at all. Recently, though, Dorit Bar-On and David Finkelstein have challenged this presumption, arguing that viable expressivist treatments of the problem of first-person privilege can be given without taking on the usual technical headaches associates with the presumed truth-valueless analysis. This recent surge in the interest being paid to expressivist accounts of privilege has set the stage for the

project undertaken in the following pages: to sketch a 'best version' of expressivism and to say how difficult it will be to defend.

As will be seen, it consistently turned out that, on my view, the best version of expressivism should end up taking on certain (more or less basic) Wittgensteinian commitments in the way of dealing with certain inevitable worries attending expressivist treatments of first-person privilege. So, the largest part of the attention given to Wittgenstein's work in the following pages owes to that simple fact: that the views I find in his work are so important to the expressivist program.

Now, it could seem that some of the detailed defenses I give of my reading of Wittgenstein as, indeed, the correct reading should be superfluous, since what matters first and foremost is to get some idea of the best version of expressivism out on the table, and not to address the historical question whether Wittgenstein himself held the view I attribute to him. But this is wrong. Aside from the pleasantness of getting the history right, the real motivation for the coming detailed defenses of the coming exegetical work begins in that an increasingly large faction of the people (the Wittgenstein experts) best acquainted with the kinds of tools I want to give the expressivist are apt to think that any expressivist treatment like the ones to be examined should be hopelessly confused because, as they're apt to see it, grammatically dogmatic and contentiously theoretic. Furthermore, the disagreement between me and those who would see the following work in that negative light amounts to nothing other than our disagreement about how best to understand Wittgenstein's own anti-dogmatic, anti-theoretic pronouncements. Thus, for example, my defense of the expressivist project against the (what I take to be, rather

serious) accusation that expressivism mistakenly seeks to explain what should only be described simply doubles as a defense of my allowing that Wittgenstein himself 'explained', in a suitably defanged sense of the word, lots and lots of things; that is, I give detailed exposition only where I go on to draw directly from the results for expressivistic purposes. The point throughout is to arrive at a sketch of the most defensible expressivism.

CHAPTER I
FIRST-PERSON PRIVILEGE: A DEFINITION AND DEFENSE

Introduction

This dissertation is an investigation into the viability of *expressivist* accounts of the phenomena of *first-person privilege* (privilege, for short). In this opening chapter I simply want to set out what I understand by "the phenomena of first-person privilege", and I want to defend the claim that there are such phenomena. Discussions of how and how well expressivist theories are suited to account for these phenomena will make up the remaining chapters, but the first thing is to fasten on just what these expressivist theories are meant to be accounts of, just what the very real phenomena of first-person privilege are.

Accordingly, in this chapter I'll first (a) set out the definitions of the phenomena of first-person privilege that I want to investigate expressivist theories of, then I'll (b) make a few clarifying remarks about these definitions, then I'll (c) turn to defending these phenomena as indeed characteristic of the first-person mode of avowing[1], and then I'll (d) offer some remarks in the way of concluding this chapter and transitioning to the next one.

First-Person Privilege: A Definition

The definition of first-person privilege I'll work with is a modification of a characterization of privilege set out by Crispin Wright (Wright, 1998). Wright says that avowals exhibit three features which "the basic philosophical problem of self-

[1] A terminological note: the word "avowal" seems to have acquired two different senses, one of which presupposes, the other of which doesn't presuppose, some account of first-person privilege other than a 'perceptual' account. The term in connection with self-knowledge seems to have originated with Gilbert Ryle (cf. (Ryle, 1949, p. 102)) for whom it was clearly meant to oppose a perceptualistic understanding of the processes issuing in 'avowals'. But somewhere along the line the term has become more or less interchangeable with "privileged utterance", abstracted from any particualr account of that privilege; such will by my use of the term.

knowledge is to explain" (Wright, 1998, p. 14); they are *groundlessness*, *strong* and *weak authority* and *transparency*. I'll set out my own versions of three of these right away; for reasons to be explained in the next section, I'll set out just one notion of authority (corresponding to Wright's "strong" notion) instead of two. So these three, groundlessness, authority and transparency, are what I understand as the *phenomena of privilege*, and their (contextual) definitions for the purposes of this dissertation run as follows:

For any subject S and any mental state *m*:

Groundlessness: The sentence "I am (not) in *m*" is groundless for S. =df Necessarily, a sincere request for S's grounds for S's claim "I am (not) in *m*" betrays a misunderstanding of the sentence "I am (not) in *m*".

Authority: S enjoys authority with respect to the sentence "I am (not) in *m*". =df Necessarily, if S sincerely claims "I am (not) in *m*", then S is (not) in *m*.

Transparency: The sentence "I am (not) in *m*" is transparent to S. =df Necessarily, S's sincere claim "I am unsure whether I am (not) in *m*" betrays a misunderstanding of the sentence "I am (not) in *m*".

These phenomena of privilege combine in the definition of "first-person privilege" that I want to assume throughout:

First-Person Privilege: S enjoys first-person privilege with respect to the fact that S is (not) in *m*. =df S is (not) in *m*; the sentence "I am (not) in *m*" is groundless for S; S enjoys authority with respect to the sentence "I am (not) in *m*"; and the sentence "I am (not) in *m*" is transparent to S.

<u>First-Person Privilege: Some Clarifications</u>

Before turning to defending the actuality of the phenomena of privilege (i.e. that there are S's and *m*'s exhibiting groundlessness, authority and transparency as here defined), I want to make a few clarifying remarks about these definitions. The last of these will take some space to set out, but the first ones will go quickly.

There is, first of all, the question what kind of necessity I intend with the word "necessarily" in the *definientia* of the first three definitions. All I believe I can say about this is that I intend something less inclusive than physical necessity and more inclusive than Tarskian semantic validity. Whatever kind of necessity attaches to "Red is darker than white", it's that kind of necessity I intend.

Next I should point out that the quote marks in these definitions are intended Quine-corner-esquely, but not strictly as Quine-corners for two reasons. For one, Quine's original corners aren't suited to my use of "not" in parentheses, the intention behind which is, I hope, obvious. For two, the relation I intend between, for instance, "*m*" as it shows up in and out of quote marks in the above definition of "authority" wouldn't be allowed by Quine either, even if the quote marks were quasi-quotational. Still, what I intend should be obvious, and anyway the occurrences of "S is (not) in *m*" can be switched out with "The sentence 'S is (not) in *m*' is true".

One may perhaps wonder, though, why I've brought sentences into it at all then, why I haven't just defined notions of authority, transparency and groundlessness with respect to facts or propositions throughout. There are two reasons. First, while I'm happy to talk of facts and will do so as is convenient, in the definitions of "transparency" and "groundlessness", crucial reference is made to *misunderstanding*, and it seems to me that they are *sentences* that are misunderstood, not facts or propositions. I've defined "authority", then, with respect to a sentence, instead of a fact, just to show up its continuity with the other two as a component of privilege.[2] Second, if one takes it that "I am in *m*", where I am S, and "S is in *m*" express the

[2] I'll also speak of privilege with respect to sentences (self-ascriptions, avowals, etc.) as is convenient, and in a manner that shouldn't be confusing given the above definitions.

same proposition, or are made true by the same fact, then if, for instance in the case of authority, I were to have, "Necessarily, if S claims that S is in *m*, then S is in *m*", it would turn out far too trivially that no S and *m* meet the condition, since it's always possible that S should make the relevant claim without having any idea that *she* is the one she's talking about. Use of the first-person pronoun is required.[3]

But bringing sentences into it the way I've done calls for the following point of clarification: I understand the possible worlds in which S claims "I am (not) in *m*" as worlds wherein S is speaking English in making her claim (the same goes for "I am unsure whether…"). Thus, while a non-English speaking German speaker enduring a migraine headache, for instance, certainly isn't disposed to make the sincere English claim "I am in pain", she will still meet the authority condition for enjoying privilege with respect to the fact that she's in pain, by the lights of my definition, just in case all the worlds wherein she makes the sincere English claim "I am in pain" are worlds wherein she's indeed in pain (at the time, that is, that she counterfactually makes her English claim). This will perhaps seem somewhat awkward, since, one would think, the privilege with which a non-English speaker sincerely claims, for instance, "Ich habe Schmerz" depends curiously, by my lights, on possible connections between that person's possible *English* claims and her possible states of mind. But the strangeness should dissipate insofar as it's seen that any language at all (comparable to English in the relevant respects) could be taken to fill the role I've assigned to English in the above definitions. Again, the need to bring the sentences of some language into the

[3] One might see a further problem on the predicate side of things, which I want to disarm simply by saying that, as I understand things, and for the purposes of these definitions, S doesn't say the same thing when she says "I'm in pain" as when she says "I'm in the state that Carl is thinking of", even if Carl is thinking of pain and no other state.

picture was just so as to bring in the first-person pronoun and the notion of misunderstanding.

As a further point of clarification, I want to point out that I don't contend that first-person privilege extends to *all* mental state self-ascriptions. While I don't want to compromise on the distinctiveness of the privilege with which one speaks about a good deal of one's own mental life, I'm not going to pretend there isn't a good deal of one's own mental life about which, in fact, one is in no special position to make privileged pronouncements.

Two easy points suggest themselves. First, one's character traits, while indisputably aspects of one's psychological makeup, are nonetheless clearly out of the running for subject matter with respect to which anyone has first-person privilege. A second easy point is that the farther away in time one wants to ascribe some mental property to oneself, the less privileged that ascription will be. If I want to ascribe to myself a headache at the precise moment of ascription, that ascription will be a likelier candidate for bearing first-person privilege than an ascription of headache to a distant past or a distant future time-slice of myself. Whether, indeed, everything goes out the window, privilege-wise, once one moves even the slightest bit from the moment of ascription is an interesting and difficult question; but the way I've already worked around this limitation on privilege is simply by defining authority, transparency and groundlessness only with respect to sentences in the first-person *present*. These sentences are interesting enough, and I'm simply going to leave out of discussion the possibility of first-person privilege with respect to mental state self-ascriptions in other than the present tense.

The next clarificatory point that I want to make, the longer one referred to above, can be got at by way of bringing out why I've defined just one notion of authority (a 'strong' one) and not two ('strong' and 'weak' ones), as would've been more in step with Wright's original scheme; for that we need to see what Wright himself had in mind with his strong and weak notions of authority. Those notions, for Wright, correspond to what he calls "phenomenal" and "attitudinal" avowals, which kinds of avowals correspond to phenomenal and attitudinal mental states. The rough and ready difference between these states, all we need for our purposes, is the difference between states with and without *propositional content*, or *aboutness* as it's sometimes put. Intuitively, pain isn't about anything (though, of course, that's a contested question[4]), and so being in pain is being in a phenomenal state; beliefs, on the other hand, are commonly construed as being about those things in the world that would enter into, as it were, their 'truth-makers', and so they're classed by Wright (quite in accordance with standard usage) as attitudinal states (where belief is seen as an attitude (opposing disbelief, say) one may take toward a given propositional content). Phenomenal avowals, then, are claims running "I am (not) in *m*" where *m* is a phenomenal mental state; *mutatis mutandis* for attitudinal avowals.

Now, as Wright has it, phenomenal and attitudinal avowals both exhibit groundlessness and transparency, more or less of the sort I've defined above. But when it comes to authority, it's only phenomenal avowals that Wright thinks exhibit something like the kind of authority I've defined, which kind of authority he calls

[4] It's rather interesting that Sydney Shoemaker (Shoemaker, 1994a) suggests assimilating sensations (e.g. pains) to the content-bearing *just so as* to secure the distinctiveness of a special first-person epistemology with respect to them, the idea being that if pains were non-intentional, then *attention* to them would have to be *observational* (*perceptual*) and thus inherently subject to breakdown. Wright doesn't address this point.

"strong"; attitudinal avowals, on the other hand, exhibit for Wright what he calls "weak" authority. This weak authority consists, for Wright, in the fact that while sincerely and comprehendingly issued attitudinal avowals don't *guarantee* their own truth, still they "provide criterial--empirically assumptionless--justification for the corresponding third-person claims." (ibid.) His reason for splitting the attitudinal from the phenomenal in this way is that, while:

> Any avowal may be discounted if accepting it would get in the way of making best sense of the subject's behavior…with attitudinal avowals, it is admissible to look for explanations of a subject's willingness to assert a bogus avowal other than those provided by misunderstanding, insincerity, or misinterpretation. (Wright, 1998, p. 17.)

But two things, I think, are not quite right about Wright's approach here. The first is that he only compromises the strength of the authority attaching to attitudinal avowals, but he leaves their groundlessness and transparency, which they share in full with Wright's phenomenal avowals, completely in tact. This seems curious to me, since, insofar as we allow for false beliefs with respect to one's own attitudinal states, it seems we should just as well allow for *having grounds for* claiming *p* and *being in doubt about* whether *p*, where *p* is an attitudinal self-ascription. For example, someone might come to believe, on the basis of her therapist's diagnosis, that she harbors the belief that her sister is out to harm her; and so she'll quite happily claim, on that basis, "I believe my sister is out to harm me". Certainly, this attitudinal self-ascription won't guarantee its own truth; but just as certainly, it seems to me, it won't enjoy anything like groundlessness or transparency either. This therapist-trusting someone will, it seems, be quite ready indeed to produce a *ground* for her claim, "I believe my sister is out to harm me"; the ground will be that her (trustworthy)

therapist told her so. Further, if she isn't sure whether to trust her therapist, she may well perfectly sensibly profess, "I'm unsure whether I believe my sister is out to harm me." If we're considering the kind of mental state about whose presence or absence in one's own case one may trust one's therapist, then it seems clear to me that groundlessness and transparency, as well as authority, are equally out of the question.

The second thing I see mistaken in Wright's approach is that he even retains *weak* authority for attitudinal self-ascriptions. If, again, we're considering the kind of mental state about whose presence or absence in one's own case one may trust one's therapist, then even Wright's weak authority is as much out of the question as groundlessness or transparency. Again, consider someone returning from her therapist, claiming on the basis of her therapist's diagnosis, "I believe my sister is out to harm me". It doesn't seem to me she gives us with her claim any special "criterial justification" for her really harboring that belief. Certainly, our going by the therapist's third-person diagnosis itself in drawing our conclusion that the patient harbors the belief in question will involve us in fewer "empirical assumptions" than our going by the fact that the self-deceived has herself claimed to harbor it; that will certainly be the case, at least, if we're assuming that the self-deceived has based her claim on her therapist's diagnosis. The moral I see in this is simply that once we're considering the sort of self-attributions of attitudinal states that one makes on the basis of the interpretation of one's own behavior, or on the basis of the testimony of an expert interpreter of behavior, no special phenomenon of *avowal* gets a grip at all. So, if it's this sort of self-attribution Wright is thinking of, then he's far too optimistic

about their retaining anything like special authority, however weak, or anything like groundlessness or transparency.

On the other hand, though, if we're talking about the self-attribution of an attitudinal state that *isn't* made on the basis of any interpretation of behavior, or on any testimony, but rather a self-attribution made in the, let's say, usual way, then it seems to me there isn't any reason to opt for a weakened notion of authority instead of the strong one I've defined in the last section (so I'll argue in the next section), and certainly no reason to give up on groundlessness and transparency either. If someone quite sincerely and comprehendingly claims, "I believe my sister is out to harm me", and if, as *per* the usual situation, no reliance on the interpretation of her own behavior enters into the way she arrived at that claim, then I don't see any reason her claim shouldn't be assessed the full strength of authority Wright reserves for phenomenal avowals. So, where attitudinal self-ascriptions are arrived at thus, I don't see that we need to compromise their authority in the slightest (so I'll argue); and where they do turn out to be anything less than strongly authoritative, then, it seems to me, they don't turn out to be weakly authoritative either.

But I believe I need to go somewhat further in drawing out this point by considering the sort of attitudinal self-attribution that might, as a matter of fact, not have been made on the basis of any interpretation of one's own behavior, or on the basis of any therapist's testimony, etc., and which is made quite apart from any kind of inference at all, but which is nonetheless on a par with, as it were, such self-attributions. Consider the person who has a *hunch* that she's harboring some kind of unfriendly attitude toward her sister, say, but who has yet to get confirmation either

from her therapist or from her own detailed observations of her quirks and otherwise unremarkable seeming habits. And let's suppose she sincerely claims, just on 'intuition', that she's harboring this attitude, a belief that her sister is out to harm her, let's say. So, does this person speak with first-person privilege in so claiming? I shouldn't think so. But she hasn't actually based her claim on any behavioral interpretation or testimony, so then why the lack of privilege?

Well, for one thing, while the claim she makes isn't actually based on any inference, testimony, etc., just because it's only a *hunch* of hers that she believes her sister is out to harm her, it's perfectly possible she's wrong; we could even build that into the scenario. But even if she'd turned out to be right, even if it had turned out that she did harbor that belief, still, I should think nothing like a philosophically important first-person privilege would have made its appearance. And the salient point seems to be this, that the sort of state in question doesn't play enough of a role in 'rationally controlling' the actions of the person in that state. Someone who harbors a belief that her sister is out to harm her, but who only knows about this belief of hers on the testimony of her therapist or takes it on a hunch, is someone who won't, for example, act in the ways a rational person would who was, as it were, perfectly in touch with her beliefs. First of all, she wouldn't respond affirmatively to the (first-level) question, "So, is your sister actually out to harm you or not?". (Please disregard scenarios wherein she expects to be rewarded for answering affirmatively to any question whatsoever, etc.) And in other ways too she won't act the ways a rational person would who held that first-level belief. If, on the other hand, it happened that she not only held that belief, but was disposed to act just the ways a

person would whose behavior was rationally guided by that belief, then (so I'll argue) she wouldn't also sincerely claim not to have it, and her claim to have it would carry the kind of first-person privilege it's important to get a satisfying theory of.

What I'd like to do, then, is to define a notion of *consciousness*, what I'll call "r-consciousness" ("r" meant to indicate the role of *rationality*) such that the claim to be defended next section will be that they are the facts about our *r-conscious* states with respect to which we enjoy first-person privilege, and not facts about our r-*un*conscious states. To stick for a moment with beliefs, the beliefs we might trust our therapist about will be r-unconscious, removed as they are from the rational control of actions; while beliefs not so removed will be the r-conscious ones.

Before setting out the definition explicitly, I should point out that my r-consciousness bears an affinity to Ned Block's a-consciousness (access-consciousness), which he defines thus:

> A state is access-conscious if, in virtue of one's having the state, a representation of its content is (a) inferentially promiscuous, i.e. freely available as a premise in reasoning, and (b) poised for rational control of action and (c) poised for rational control of speech. (Block, 1994, p. 215)

Block seems to have in mind *perceptual* states first and foremost when he discusses access-consciousness[5], but I don't see that his notion won't apply to all the content-bearing mental states; and anyway my r-consciousness is explicitly meant to do just that.

Now, the first difference between Block's a-consciousness and my r-consciousness is that nothing corresponding to clause (c), what he calls "reportability" (Block, 1995), will make its way into my notion, since, as should be

[5] Cf. (Block, 1995)

clear, we won't want to build reportability into our conception of r-consciousness. The second thing will be that clause (a), along with the phrase "a representation of its content", will only apply to doxastic/perceptual r-conscious states (and not to desires, for example). The main affinity between a-consciousness and r-consciousness, then, comes from Block's clause (b), a correspondent to which will do the heavy lifting for my conception of r-conscious states, which conception runs thus:

r-consciousness: Mental state *m* is an r-conscious state of S =df *m* is poised for rational control of S's actions; and if *m* is a doxastic/perceptual state of S, then a representation of its content is freely available as a premise in S's reasoning.

I should note, now, that there can seem to be a problem in combining my contention that they are only the r-conscious states with respect to which we enjoy first-person privilege and my use of the necessity operator in the definitions from the last section. The problem comes if we take it that "I am (not) in *m*" is the same claim regardless of whether *m* is taken to be r-conscious or not. So, somebody could be in the r-conscious state of believing her sister is out to harm her, and sincerely claim "I believe my sister is out to harm me", but because it's *possible* that she should make the *same* claim and be wrong (because in some 'possible situation' she bases her claim, let's say, on faulty testimony about her r-*un*conscious belief about her sister), *whenever* (in *whichever* 'possible situation') she makes that claim, she lacks authority, according to the definition above. There are two ways to handle this: first, we can, by introducing complications to the above definitions, have it so that one sometimes does, sometimes doesn't, enjoy authority with respect to one and the same sentence, making similar adjustments for the other definitions; or second, we can simply have it that "I am (not) in *m*" is *ambiguous* between an r-conscious *m* and an

r-unconscious *m*. For simplicity, and because it seems nearer to the truth anyway, I'll adopt the second way of handling this technicality.

With that, I'll turn to defending the actuality of the phenomena of first-person privilege.

First-Person Privilege: A Defense

So, what about the suggestion that even one's sensory experiences, one's most mundane beliefs, one's strongest desires could be, for all one knows, subject matter for error? This is *prima facie* rather startling; nonetheless, there are more or less reasonable arguments for just such separations between what we're inclined to say about our own mental lives and what may very well be going on with them. I'm going to address these in this section.

One thing to point out is that an important by-product of the discussion of privilege I intend to conduct here is that it will be seen how one can fend off attacks against privilege without appealing at all to traditional intuitions about an inner Cartesian theater. Indeed, it's fending off that will be the main concern here. That is, the humble strategy I'm going to adopt in this section is simply to help myself to the presumption that the burden of proof lies with the deniers of privilege, and see whether any of their arguments casts anything like conclusive doubt on the ostensible phenomena of first-person privilege. I don't think anything like a knockdown argument for privilege is required from my end so as to motivate the search for a proper philosophical account of first-person privilege. I simply want to have a look at the more central anti-privilege arguments and see whether there's very much to dissuade someone from taking privilege seriously. It doesn't seem to me there is, in

fact, as will come out. But it must be borne in mind, in any case, that the way I want to conduct this particular discussion, theirs is the side with the explaining to do.

There certainly isn't any shortage of explanations on their side of things, and the basic themes of these I want to collect together as follows: (1) arguments from the substantiality of self-knowledge, (2) arguments from coherentism/inferentialism, (3) arguments from possible, future brain science, (4) arguments from illusion, and (5) unargued claims about the obviousness of the possibility of inattention to one's mental states. I'll treat these in order.[6]

Arguments from the substantiality of self-knowledge rely on the premise that a necessary condition for the substantiality of self-knowledge is a real possibility of failure in introspection, otherwise it would be pointless to speak of self-knowledge as representing any kind of achievement. David Armstrong writes:

> There is one important line of argument that derives from Wittgenstein. If introspective mistake is ruled out by logical necessity, then what sense can we attach to the notion of gaining knowledge by introspection? We can speak of gaining knowledge only in cases where it makes sense to speak of thinking wrongly that we have gained knowledge. In the words of the slogan: "If you can't be wrong, you can't be right either." If failure is logically impossible, then talk of success is meaningless. (Armstrong, 1963, p. 422).[7]

---

[6] It will be noticed that two oft discussed arguments against first-person privilege are conspicuously absent from the list, viz. Armstrong's "distinct existences" argument (cf. (Armstrong, 1963)) and Timothy Williamson's sorites argument (cf. (Williamson, 2000)). I don't include the first because, as Frank Jackson has pointed out, "A husband is numerically distinct from his wife, but 'I am a husband' entails 'I have a wife'" (Jackson, 1973, p. 58). I don't include the second because, as Selim Berker has shown, it's too crucially question begging (cf. (Berker, 2008)).

[7] Boghossian makes the reverse argument: from the premise that introspection is fallible he concludes that its results are epistemically substantial, that they represent a genuine cognitive achievement. See (Boghossian, Ayer, A. J. (1965). Basic Propositions. In *Philosophical Essays* (pp. 105-124). New York: St. Martin's Press. 1989, p. 19). Putting together the conditionals, then, from both Armstrong's and Boghossian's arguments, we get a plausible biconditional, viz. that self-knowledge is substantial if and only if introspection is fallible.

In the words of another slogan: "One man's *modus tollens* is another man's *modus ponens*." In any case, Armstrong makes a rather astute point here, as I see it. (In fact, I sometimes wonder whether this isn't really at bottom of so many people's resistance to the most serious kind of first-person privilege.) In the way of responding to this particular line of argument, then, against the 'incorrigibility' of introspection, I'll simply point out that, for the purposes of this dissertation, the substantiality of self-knowledge isn't to be taken for granted.[8] How the expressivist would teach us to live without substantial self-knowledge, we'll have to wait until later to see, but if this were the *only* argument against first-person privilege, and it most certainly isn't, the expressivist would have no trouble whatsoever; she could respond simply, "That's exactly what I'm trying to say."

Coherentism/inferentialism based arguments proceed on the premise that groundlessness in introspection is, in fact, impossible. There are three ways this kind of thinking gets started that I want to consider here, the last more general and the first two less so: (a) from a commitment to theories of introspection as essentially-memory-involving, (b) from a concern that, in an important sense, mental states are individuated relationally, and (c) from a general commitment to coherence theories of justification.

Beginning with (a), one characteristic example of this line is the following:

> To judge that a phenomenal occurrence has the property F is to

[8] Armstrong is explicit about using "incorrigibility" to mean something very like what I've called "authority". It's interesting that "incorrigibility" has enjoyed a peculiar dual use, on the one hand being simply substitutable for "authority" more or less as I've characterized it, but on the other hand indicating something (supposedly) slightly less forceful, namely the impossibility of a claim's being rightly overturned on the introduction of further evidence. For a useful exploration of such terminological possibilities, cf. (Alston, 1971). Also, for now I'm going to take on the term "introspection" to signify however it is one comes to make the kind of mental state self-ascriptions I'm concerned to assess expressivist theories of.

> assert that it belongs to the class of F's, that the property it has is just the property that is possessed by other F's. But how could one know this infallibly, if one's memory is intrinsically fallible?--if one may well misremember the peculiarities, the distinguishing features, of F's generally? (Aune, 1967, p. 35.)

This line, let's call it the memory-inferentialist line, would seem all right to me, that is on the presumption that such a reliance on memory is indeed essential to the introspective process. But what's the argument for that? Typically, the argument that introspection does, or had better, rely thus on memory (and so must admit fallibility) is made on the premise that, just like before, otherwise introspectively based claims would fail of substantiality: "[E]ach so-called identification would turn out to be nothing more than a kind of ceremonial announcement or verbal baptism, something very different from an out-and-out claim to knowledge." (ibid.)[9] So, again, how devastating this is we'll have to wait to see. One thing I can say now, though, is that it seems to me a level-headed expressivistic theory of avowing is indeed particularly well placed (along with constitutivist theories, I suppose) to reconcile the epistemic insubstantiality with the *contentfullness* of introspective mental state self-ascriptions. While I would agree with exponents of the memory argument that certain traditional accounts of authoritative introspection indeed cannot sustain the epistemic substantiality of introspectively based claims, I don't agree that in order to cast introspectively based claims as more than "verbal baptisms" we need to cast them as involving reliance on memory. I want to say a little more in response to this momentarily, but first I want to shift to concerns to do with the relational

[9] A number of authors have criticized this line of argument, trying to save the plausibility of introspective infallibility by way of reconciling the exclusion of reliance on memory from introspection with the epistemic substantiality of introspection. Cf. (Ayer, 1965, pp. 116-122), (Chisolm, 1982, pp. 141-5) and (Jackson, 1973, pp. 53-6).

individuation of mental states, since the response I want to give to memory-inferentialism applies equally to those concerns.

Gilbert Ryle's behaviorism famously led him to what, I think, can rightly be called an inferentialist account of self-knowledge. Ryle's principal concern in *The Concept of Mind* was to clear away traditional conceptions of the mind and of introspection (Cartesian Theater conceptions). The behavioristic alternative he worked out, regarding introspection, was to cast the processes issuing in self-directed mental state ascription as no different, in essentials, from those issuing in other-directed mental state ascription. "Our knowledge of other people and of ourselves depends on our noticing how they and we behave." (Ryle, 1949, p 181) Bertrand Russell also theorized, some years earlier, that a proper account of certain mental states will result in the breaking down of self/other asymmetries when it comes to their means of discovery. His more explicitly inferentialist formulation about desires (motives) in particular is the following:

> I believe that the discovery of our own motives can only be made by the same process by which we discover other people's, namely, the process of observing our actions and inferring the desire which could prompt them. (Russell, 1921, p. 31)

His account too arose from a behavioristic analysis of mental states. And actually, it seems inferentialism about introspection will follow any story of mental states that, like a behavioristic story, individuates those states according to certain of their *relational* properties; and so any argument against inferentialism about introspection will be, it seems, an argument against such modes of mental state individuation. On the other hand, the appeal of relational identity conditions for certain, if not all, mental states can be quite strong, and so any account of

introspecting mental states that makes no mention of inferential processes will either have to sever the *prima facie* connection between relational individuation and inferential introspection, or else explain away the attractiveness of relational individuation. I prefer the tack of severing that connection, because I think the pressures to eschew inferential models of introspection and to embrace the relevance of relational factors in mental state individuation are, on their own, quite a bit stronger than the pressure to connect up relational accounts of individuating mental states with inferential accounts of introspecting them.

That said, I want to set out what I see as an important line of response to both these kinds of inferentialism about introspection. The basic idea of my response comes from Wittgenstein. I want to get it on the table by drawing an analogy between, on the one hand, the processes of inference that inferentialists would have mediating introspected states and introspective claims, and, on the other hand, the processes of *interpretation* that (for lack of a better word) interpretationalists would have mediating commands heard and commands obeyed. As Wittgenstein said in the *Blue Book*, our problem is analogous to the following:

> If I give someone the order "fetch me a red flower from that meadow", how is he to know what sort of flower to bring, as I have only given him a *word*?
>
> Now the answer one might suggest first is that he went to look for a red flower carrying a red image in his mind, and comparing it with the flowers to see which of them had the color of the image. Now there is such a way of searching…But this is not the only way of searching and it isn't the usual way. We go, look about us, walk up to a flower and pick it, without comparing it to anything. (Wittgenstein, 1969, p. 3)

As I see it, precisely parallel remarks apply to our struggle with the inferentialists: there is such a way as they propose of ascribing mental states to oneself; but this is

not the only way, and it isn't the usual way. In the vast majority of cases we simply ascribe mental states to ourselves without making any inferences at all. Just ask yourself, without thinking about what *must be* going on according to some philosophical theory or other, whether any such processes ever actually accompany your introspectively based claims. I believe you'll find you plainly don't call to mind (by way of memory or otherwise) samples of red or of the-desire-for-food before you claim, in that effortless manner with which we're all so familiar, to see red or to be hungry. I believe you'll find you don't draw conclusions about your aches and fears by first observing your behavior either.

Of course, it's open to the inferentialist simply to insist that these processes go on without our ever knowing about it. But this strikes me as an unnecessary and, indeed, far-fetched hypothesis. What's needed is a way of doing justice to introspection (as unencumbered by acts of inference) while at the same time doing justice to the thoughts that, first, introspectively based claims aren't mere verbal baptisms and that, second, mental states are, in a legitimate sense, individuated relationally. Indeed, expressivism seems to me perfectly well placed to pull off both tricks, but because this is of some interest it will have to wait until the last chapter before I return to remark on these advantages. In any case, once it's seen that there *needn't be* acts of inference involved in the acts of introspection we've been considering, then I think the response to the above argument that such processes would have to proceed unconsciously will simply fall away.[10] For now, though, as I

[10] Incidentally, the 'thought experiment' Wittgenstein poses in order to help with the idea that there needn't always be intervening acts of interpretation is that we consider the order to *imagine* a red patch (instead of to fetch a red flower, e.g.): "You are not tempted in this case to think that before obeying you must have imagined a red patch to serve you as a pattern for the red patch which you were ordered

said, a discussion to the effect that, indeed, a non-inferential account of (privileged) mental state self-ascription can be reconciled with genuine contentfullness, and such contentfullness as is individuated relationally, is deferred until later.[11]

So then what about a more general commitment to inferentialism? Quite apart from speculations about particular inferential needs in the particular case of mental state self-ascription, there's a general approach in epistemology by the lights of which, strictly speaking, every statement one could possibly hope to make with epistemic green light, so to speak, indeed derives inferential support from other statements one is prepared to make. This is the approach of the coherentist, as opposed to the foundationalist, in the theory of knowledge and justification.[12] Without proposing any particular inferential process, so that we can at least try to look and see, as it were, whether it indeed goes on, the general commitment to inferentialism, that is coherentism, is more purely abstract. This general coherentism won't say anything so particular as that, for instance, they are memory images we call up, or that they are facts about our behavior we appeal to in introspection; and so we can't use the fact that these certainly don't ever *seem* to be appealed to, and the absurdity of a scene wherein such convoluted processes were indeed quite self-

---

to imagine." (Wittgenstein, 1969, p. 3) I would draw the reader's attention to the word "tempted" here as opposed to the word "allowed".

[11] Contemporary (content externalist) concerns about the irreconcilability of the relational individuation of mental states with privileged introspection, unlike Ryle's and Russell's behaviorist concerns, don't rely on arguments that introspection involves inferences we wouldn't have expected. Rather, the contemporary concern is precisely that introspection doesn't involve inference where it seems it should (though, even if it did, privilege is still certain to be drastically compromised). As I say, a discussion of these concerns is deferred until the last chapter.

[12] Actually, the very notion of 'inference' as it applies to coherence theories of justification is arguably different from the notion of the same name that applies to foundationalist theories, which latter contrast inferential from non-inferential justification. Perhaps the primitive "coherence" would be better than "inference", though there's no doubt an important resemblance between the two.

consciously engaged in, as a way to point at its unsatisfactoriness in the case of arriving at certain mental state self-ascriptions.[13] Just the same, I propose we rely still on the considerations I've set up against the particular inferentialisms already mentioned. Zeno of Elea knew quite well that motion was possible; and while this wouldn't itself count as a solution to his paradoxes, still, anyone seriously reasoning from the basis that motion is impossible, is, I think, making a grave error. Similarly, even in the absence of particular inferences to look out for, the general principle telling us there has to be something or other inferential going on in introspection has got to be looked at with deep suspicion given its, apparently, patent falsity as applied to that particular case. Because of this, I don't think there's any conclusive argument that introspection is inherently inferential and so out of the running for being privileged in the way I've said it is. Furthermore, there being no conclusive argument to that effect is sufficient for what I'm after right now. It seems to me that the presumption in favor of privilege is strong enough, the need for a satisfying account of it *prima facie* pressing enough, that nothing so controversial and paradoxical as a thoroughgoing coherence theory of justification should give more than a moment's pause.

The arguments from possible, future brain science proceed on the premise that it could just so happen that correlations between people's brain states and their claims to be in such-and-such mental states should become so well established that they both should count equally well as evidence for a person's mental state, and so that if they should then diverge here and there it would simply be contrary to the spirit of science

[13] Cf. (Bonjour, 1978) for the basic arguments for a general coherence theory of justification.

should we rule *a priori* that the first-person claims-as-evidence automatically override the brain-o-graph claims-as-evidence. Richard Rorty asks us to imagine that:

> [O]ne day (long after all empirical generalizations about sensations *qua* sensations have been subsumed under physiological laws, and long after direct manipulation of the brain has become the exclusive method of relieving pain) somebody (call him Jones) thinks he has no pain, but [an] encephalograph says that the brain-process correlated with pain did occur…it is fairly clear what Jones' first move would be--he will begin to suspect that he does not know what pain is--i.e., that he is not using the word "pain" in the way in which his fellows use it. (Rorty, 1965, p. 43)

Jones is supposed to say, "I don't know what pain is" instead of, "Strange; brain process α is taking place, but no pain process is". It's supposed to be clear that this will be Jones' first move. I have to say, I find the Rortian (Jonesian) first move here extremely perplexing. In his story, do we teach people to talk about their pains by pointing out states of their brains? Or do we teach them the normal way, and then later they learn of correlations? If we teach them the normal way (that is, the normal way *as is stands*, not some *possible future* normal way) then I simply don't understand this line one bit. For one thing, why wouldn't Jones think he doesn't know what a *brain* is; that is, that he isn't using the word "brain" in the way in which his fellows use it? And, does Rorty think that if he lit a candle, for example, but found it emitting no heat, that he would then be unsure what a candle is? Or what heat is? I very much doubt it. But that's because I have a pretty good guess about how Rorty grew up using the words "candle" and "heat", and I'm willing to bet he hasn't been led away from that path of use by any philosophical interest in reducing candles to heat or *vice versa*.

Consider this, if Jones had been brought up in one of the Judeo-Christian faiths, had grown up going to Sunday school and all that, his first move on being

brought to give up that lifestyle (for whatever non-coercive reason, let us suppose) might very well be to give up any talk of morality. He might say, "There being no God, I just don't know what morality is." But that will be understandable given the likely upbringing with moral vocabulary Jones was subject(ed) to. If, on the other hand, he'd gotten a very different upbringing with moral vocabulary, a more, let's say, secular upbringing, the suggestion that absent a creator God one couldn't know moral truths would seem absurd. "Who talks like that about morality?" he might say. Furthermore, given this second kind of upbringing with moral vocabulary, no amount of correlations between what an actual creator God (we'll suppose there is one) approved of (commanded, what have you) and what Jones himself ordinarily judged morally virtuous would have anything to do with how he'd respond to that absurd (from his perspective) suggestion. That is, not insofar as he hadn't been led away from the path of usage he'd been set on by those who had brought him up.

Now, in the case of moral vocabulary, it seems to me people are indeed brought up very differently; and in the case of candles and heat, it seems to me people are pretty much brought up the same. Moreover, in the case of candles and heat, it doesn't seem to me very much tends to lead one away from the way we were all brought up. Now, what I think about the case of pains and brains is this, we were indeed all brought up the same way (so that the possibility that there weren't any brains in our skulls wouldn't have us thinking we didn't know what pain was), but philosophical pressures have led some of us away from the path we were set on. That is, I have a pretty good guess about how Rorty grew up using the words "pain" and "brain", but I also believe I can see plainly why he should be led away from that path

of usage, namely because of a dissatisfaction with dualism, which view he seems to think is the inevitable consequence of admitting first-person privilege.[14] In any case, what it seems to me we have with Rorty, as with Armstrong, is an antecedent philosophical project of materialism and a recognition that incorrigibility is incompatible with that project. And so my response to the brain science argument is simply that if we hold constant the way we *actually* teach people to make first-person claims, then should first-person claims and brain-o-grams begin to diverge after however long a period of convergence, nobody's scientific scruples should get in the way of preferring the first-person claims to the brain-o-graph claims. Philosophical pressures (like those linking first-person authority to dualism, e.g.), may, of course suggest we ought not, in fact, keep that usage constant (the usage, that is, that almost everybody actually grew up in, and which respects first-person authority). But one of the main advantages of expressivism, it will come out later, is that it would dissolve such pressures, allowing us to continue in the path we were actually set on as we grew up talking about our own and other people's pains, thoughts, desires, etc., without demanding from us an acquiescence in dualism.[15]

Arguments from illusion are, in my estimation, the weakest of the arguments against privilege. Such arguments begin with the mundane fact that things can appear otherwise than they really are, and conclude that things can appear to appear otherwise than they really appear. Daniel Dennett gives a materialist version of this kind of argument. After telling us how, from a moderate distance, he "marveled at

---

[14] Cf. (Rorty, 1970, p. 414).

[15] For other replies to the brain science argument, see (Jackson, 1967, pp. 358f.) and (Jackson, 1973, pp. 61f.).

the gorgeously rendered details" of a certain painting on a certain occasion, Dennett says:

> I remember having had a sense that the artist must have executed these delicate miniature figures with the aid of a magnifying glass. When I leaned close to the painting to examine the brushwork, I was astonished to find that all the little people were merely artfully positioned single blobs and daubs of paint--not a hand or foot or head or hat or shoulder to be discerned." (Dennett, 2002, p. 489)

The conclusion Dennett draws is that, while his "brain 'filled in' all the details", still there weren't any "particular neural representations" of those details "created by [his] brain". (Dennett, 2002, p. 490) Translated into talk of beliefs and experiences, something I don't see Dennett would have any problem with, this says that while he (part of his brain) *believed* his visual experience had certain details, his (brain's) belief was *mistaken*. Now I want to ask, Is this really the moral to be drawn? Wouldn't it be better to say that he had a visual experience *as of* detail, which was *itself* illusory, that is, non-veridical? Or even (probably closer to the truth) that he didn't *really* believe his visual experience was all that detailed, but he believed that the visual experiences he would *soon* have would be, and that *that* belief was mistaken? Why should he think he believed he had a particularly detailed visual experience when he had no such experience? Just because he thought the painting was detailed when it wasn't?

To get at my point it helps to note that Dennett's surprising conclusion is something one could arrive at just as easily in the case of any visual illusion. Consider the Müller-Lyer, under the grips of which people will take two lines of *actually* equal length to be other than actually equal. Isn't the typical diagnosis of the mistake here simply that they took one *line* to be longer when it really wasn't? Do

we need to back up further, supposing their mistake actually began when they took *themselves* to be taking one line for longer than the other? So that if they'd just known how they took the lines, they wouldn't have said the one was longer? One wonders how far back one could go with this. What one shouldn't wonder, though, is how far back one should bother going with this. There simply isn't any need to describe this situation with the Müller-Lyer in anything other than the standard way, and I don't think there's any need to describe the illusion of a detailed painting (however gorgeous) in anything but the same way. Of course, if Dennett, like one version of Rorty's Jones, had been brought up to talk of his visual experiences by being shown diagrams of "neural representations" and being told, for example, "*That* is what your visual experience of red looks like in your brain; when you say you're experiencing red you're saying that *that* is going on in your brain", then I suppose I would understand why he takes the surprising turn he takes. He believed that that was going on in his brain when it wasn't; no problem. But I don't think Dennett was brought up that way, nor do I think he should have been (not that he shouldn't have been, of course). Outside philosophical pressures, pressures to abandon the practices we grew up with, are surely at work here. In any case, this is not the way we've been brought up to describe visual illusions, and I don't see that any amount of parading visual illusions *themselves* should dissuade us from that practice.

Finally, I'd like to consider the view that it's just obvious that we can pay greater or lesser attention to our mental states (from which the joint denial of authority, groundlessness and transparency as attaching to sincere mental state self-ascriptions is an immediate corollary). To some writers, for instance Paul

Boghossian, the demand for a "cognitively substantial" first-person epistemology (that is, for Boghossian, one grounded either in observation or in inference (Boghossian, 1989, p. 17)) is ensured by the, apparently obvious, fact that "one can decide how much attention to direct to one's thoughts or images, just as one can decide how much attention to pay to objects in one's visual field." (Boghossian, 1989, p. 19) Surprisingly, however, Boghossian is content to grant that perhaps, in the case of sensations, the situation is quite different. About pain in particular he writes that: "[I]t seems not conceivable, in respect of facts about pain, that we should be either ignorant of their existence or mistaken about their character…" (Boghossian, 1989, p. 7)

Brie Gertler seems to conceive just what Boghossian can't, then, when she writes: "The doctor asks you whether the pain in your arm is an ache or, rather, a burning sensation. You reflect for a moment before answering. What process do you undergo during this interval which informs your answer?" (Gertler, 2001, p. 305) The process, Gertler eventually discloses, is precisely that of *attending to* the character of the pain in one's arm, which kind of attention is, she tells us, unanalyzable. (Gertler, 2001, pp. 317, 8)

And Richard Fumerton doesn't even bother about finer-grained differences between aching and burning pains; he conceives a scenario wherein one's very pain (not just its aching or burning quality) goes unattended to by one; and conceives it precisely in order to bring to mind the kind of attention one sometimes *does* pay to one's pains:

> [T]hink about a situation in which you were aware of an intense pain, got lost in an engaging conversation, and for a while no longer *noticed* the

> pain…If one *can* make sense of [this], then one can isolate direct acquaintance. Direct acquaintance is the relation you were in to the pain before you got lost in conversation, which ceased during the conversation, and which came into existence again as the conversation ended.[16] (Fumerton, 2006, p. 64)

Here it's strictly a matter of what one *notices*, instead of what one *attends to*; but I can't think that particular word choice marks anything of significance. In any case, I'd like to work backwards from Fumerton's stimulating conversation scenario, to Gertler's doctor scenario, to Boghossian's unargued assertion that, excepting sensations, greater or lesser attention may be paid to one's own mental happenings.

Fumerton himself points out that "there are two main reactions philosophers have to this thought experiment" (ibid.). The one is to grant with him that, in the course of stimulating conversation, one may fail to notice a pain that hasn't itself diminished one iota. The second is to insist that "in such situations the pain itself temporarily ceased". (ibid.) My reaction is precisely the second, and there only three reasons I can imagine one might be inclined to the first: (a) because of inclinations toward materialism (so that, for instance, brain states settle the question independently of what one is disposed sincerely to claim), (b) because one imagines the person exhibiting *conflicting* signs of pain and its absence (for instance, rubbing one's arthritic knees and moving slowly while smiling, disavowing pain and declining pain-killers), and (c) because one hypostatizes an inner theater, as it were, the players on whose stage are not all always in view. As for the first reason, I believe the proper

[16] It's important to note that what's being proposed here is *not* the kind of (deeply troubling) situation wherein a person quite reflectively avows being in pain, but not being *bothered* by it. Such scenarios are often brought up in the way of arguing against privilege, but they clearly have no such application. I don't want to get into what conceptual upheaval perhaps *is* in order, given such scenarios, but it's clear to me that a person suffering this kind of 'dissociation' hasn't been deprived of *privilege* with respect to *whether* she's in pain. Indeed, it's the presumption of just such privilege that makes the situation so troubling in its own way. For an important discussion of this phenomenon, see (Nikola Grahek, 2007).

response is as before: one of the main points of expressivism about first-person authority is that justice can be done to a serious kind of first-person privilege without succumbing to any kind of immaterialism. As for the second reason, I think what we can say is that we have something analogous to an r-unconscious state, and I'm willing to grant that someone may be in pain in this sense while sincerely claiming not to be. Note, however, if the person were indeed to grimace, seek relief, etc., on top of exhibiting the more subtle self-tending behaviors but *still* seem sincerely to avow the absence of pain, then it seems to me we should have to question her sincerity or understanding in so 'avowing'. Finally, concerning the thought that surely not all of what happens on the stage of the inner theater happens directly in front of one's inner eye(s), as if *that* were Descartes' mistake, my response is simply that this is to retain precisely what was *wrong* with Descartes' fantasy. That is, Descartes' problem, I would suggest, wasn't his naïvely taking it that, for instance, pains on the inner stage couldn't recede into the shadows on the stage, but it was in his taking our ability to speak with authority about our pains to have anything to do with our, as it were, spying pains on some inner stage, however well or poorly lit, however conspicuous or inconspicuous that stage's inhabitants may be. To transcend Descartes is not, I would suggest, to get *realistic* about the various possible arrangements and saliences of players on an inner stage, rather it is to respect a serious first-person privilege *without* dreaming of any inner production whatsoever. So, I think we do just fine to conceive of someone enduring pain, becoming engaged in stimulating conversation, sincerely claiming while so engaged to be free from pain, and enduring pain again when the conversation ends, as indeed being relieved of pain

during that conversation. Getting engrossed in a conversation, it shouldn't be any surprise, is one way of causing pains to subside for a while.

Turning now to Gertler's thought that one sometimes takes a moment to reflect whether certain pains are achings or burnings, and that this is a matter of paying attention to the persisting character of one's pain, my suggestion is that something quite different goes on during those reflective moments that has nothing to do with paying attention to what one previously could've been mistaken about for lack of attention. What it seems to me does go on is better (or at least as well) conceived as taking a moment to try on for size, as it were, the two descriptions ("aching" and "burning"), seeing which fits the situation better. One perhaps calls to mind images of someone rubbing her shoulders slowly, complaining of a deep ache; or one imagines burning one's skin on a hot pan; but these have nothing to do with getting a better view, as it were, of the pain one is asked to characterize. Much the same could be said about rating the intensity of one's pain: one may pause to consider whether one's pain deserves a 6 (out of 10), but these considerations principally involve *imagining other painful situations* and not *attending to the present one* (where "attending to" means putting oneself in a better position to ascertain what was already afoot). Again, these pauses one takes at the doctor's office or wherever are better thought of, on my view, as opportunities to let one's imagination come to the aid of settling on an apt characterization, *not* as opportunities to rub the blur from one's inner eye(s), as it were, and studiously observe the inner object.

Lastly, then, Boghossian's claim that one's thoughts and images (and let's include beliefs and other content-bearing states here too) are obviously subjects for

greater or lesser attention, just the way we can pay greater or lesser attention to objects in our visual field, seems to me quite avoidable. Sydney Shoemaker makes, I think, the right point about what it could mean to turn one's attention to one's beliefs:

> If it means anything to speak of my attending to my belief that Boris Yeltsin is President of Russia, it means either to think about the proposition and presumed fact that is the content of my belief, i.e., to think about Boris Yeltsin's being President of Russia, or to think about the fact that I have that belief. So all that answers to talk of shifting one's attention from one belief to another is *shifting from thinking of one thing to thinking of another thing*. (Shoemaker, 1994a, p. 265, my italics.)

And this point, indeed, goes for sensations, thoughts and images too, as I see it. If *turning one's attention to x* simply comes to *contemplating* (*the fact that*) *x*, or *taking note of* (*the fact that*) *x*, then I have no objection to talk of attending to one's mental happenings; but when attention to one's mental happenings is modeled on attention to objects in front of one's eyes, attention given so as better to position oneself to describe those objects, then it should be resisted. At least, I should say, it can be resisted by those who would.

In sum, then, my recommendations are: first, that we conceive of unnoticed sensations (where curious and conflicting behaviors aren't in question) as non-existent sensations; second, that we conceive of the pauses one takes before settling on "aching", "burning", "throbbing", "stabbing", "fluttering", "deep", "shallow", "dark", "light", "sharp", "dull", etc., as the best characterization of one's sensation to be a matter of trying these descriptions on for size, as it were, and not a matter of getting a better view of the sensation itself; third, that we conceive of whatever leftover sense we have of 'attending to' our mental states as a matter of contemplating or taking note of them, rather than of perceiving them more clearly and distinctly.

And the spirit behind these recommendations is the hope thoroughly to transcend Cartesianism, but not by getting realistic about what on the inner stage a less idealized inner eye can take in at any given moment, rather by rejecting *in toto* that model of what goes into mental state self-ascription.

Conclusion

We've seen that the main enemies to the kind of first-person privilege I want to take philosophically seriously are (1) adherence to the epistemic 'substantiality' of self-knowledge, (2) worries about the relational nature of mental states, (3) adherence to thoroughgoing coherentism, (4) adherence to (some forms of) materialism, and (5) adherence to a watered down kind of Cartesianism. My responses have been that (1′) perhaps epistemic 'substantiality' is not to be had, though the contentfullness of first-person privileged claims is indeed to be retained, (2′) worries about the implications of the relational nature of mental states for first-person privilege can be allayed, (3′) thoroughgoing coherentism is too controversial to cast serious doubt, (4′) materialism (in certain forms) simply diverts us, and unnecessarily so, from our ordinary and tenable practices of privileged mental state self-ascription, and (5′) inner-theater-Cartesianism can be rejected *in toto*.

(1′) and (2′) will, as promised, be expounded upon in further discussions, and in the final chapter in particular they'll receive more focused attention. (3′) will not be expounded upon, but a brief discussion of the consequences of expressivism for foundationalist projects in epistemology will be included in the final chapter. (4′) will as well be revisited in the final chapter, wherein the compatibility of expressivism about first-person privilege with various mind-body solutions will come

under discussion. Finally, how expressivism can yield (5′) will become clear as early as the next chapter, wherein the basic motivations for expressivism are evinced.

CHAPTER II
INTRODUCING EXPRESSIVIST THEORIES OF FIRST-PERSON PRIVILEGE

Introduction

The point of these discussions, remember, is eventually to work ourselves to a place from which we may best assess the prospects for expressivist theories of first-person privilege. But in order to do that it must first be made clear what expressivism about first-person privilege is precisely, and what drives it, so to speak; and just as much as that it needs to be made clear exactly what drives its opposition. Accordingly, in this chapter I want to (a) set out the most basic contours of what I'll call the *expressivist insight* concerning the problem of first-person privilege, in particular as I see them unfolding in Wittgenstein's later work, (b) set out a quartet of questions that will immediately crowd in on anyone trying to work from that expressivist insight, (c) set out one *simple* way of handling those questions while operating within the spirit of the expressivist insight[17], and (d) set out the most pressing difficulties for that kind of simple expressivism.

Before getting directly into setting out the expressivist insight as it unfolds in Wittgenstein's later work, though, I would like to mention, very briefly, just three components of the expressivist insight that we should be on the lookout for in the next section. This will give us at least something minimal to keep hold of from the outset, just in case the very idea of anything like an expressivist treatment of the phenomena of first-person privilege seems something initially rather obscure. The three components of the expressivist insight we should look out for are: (α) that it emphasizes a likeness between privileged utterances and natural *expressions* of

[17] I follow Dorit Bar-On in calling the sort of expressivist treatment to be set out later "simple expressivism". (Bar-On, 2004, pp. 228-40)

mentality (likening "I'm thirsty" to reaching for a drink, e.g.), (β) that it casts those utterances' privilege as a "natural concomitant" (Finkelstein, 2003, p. 102) of their likeness with natural, expressive behavior, and (γ) that in virtue of (α) and (β) it stands opposed to any perceptual/observational conception of the processes issuing in privileged mental state self-ascription (and, closely related to this, it should stand opposed to the inner-ostensive-definition conception of one's coming to understand the terms involved in privileged mental state self-ascription).[18] With these things in mind, then, we can move on to seeing how it was Wittgenstein managed to articulate, for the first time, an expressivistic strategy for approaching the philosophical problem of first-person privilege.[19]

Wittgenstein's Expressivist Insight

It will be largely agreed that the elements for expressivist theories of first-person privilege were first mined by Wittgenstein; furthermore, it seems to me that Wittgenstein's remarks on the uses of first-person psychological ascriptions remain the most compelling source of inspiration for the expressivist about first-person privilege. As will come out next chapter, it's not clear to me that Wittgenstein

[18] For a useful discussion of the "stereotypes" informing perceptual models of introspection, see (Shoemaker, 1994a, pp. 252-4) and (Shoemaker, 1994b, pp. 271f.).

[19] It should also be borne in mind that nothing in these components of the expressivist insight is so strong as to entail what Hacker has called the "truth-valueless" doctrine of avowing. (Hacker, 1975, p. 265) Perhaps due to a few ill-chosen remarks by Norman Malcolm (*e.g*. that, according to Wittgenstein, avowals have the same "logical status" as "outcries and facial expressions" (Malcolm, 1966, p. 83)) and due to the (modest) popularity of non-cognitive (truth-valueless) treatments of ethical discourse (cf. (Stevenson, 1959), (Ayer, 1952) and (Blackburn, 1984) for classical non-cognitivist treatments of ethical discourse), expressivism about first-person privilege seems to have been stuck until quite recently with the presumption that privileged because expressive mental state self-ascriptions ought to be devoid of truth-value. How, in fact, expressivism about first-person privilege could be divorced from the truth-valueless doctrine will, of course, be integral to our understanding of its viability in today's philosophical climate. In any case, that the door is open to this possibility is perhaps too easily neglected and should be borne in mind by the reader who might otherwise be confused by some of the remarks to follow.

propounded anything that would rightly be called a fully worked out expressivist theory of first-person privilege, but just for that reason it's easier to see in his work the expressivist insight at its purest, so to speak.

The first thing to note about the expressivist insight, as it unfolds in Wittgenstein's later work, is that it's very much bound up with his revolutionary approach to questions having to do with meaning and reference. So in order to point out the expressivist insight in Wittgenstein, it will be necessary to discuss his later approach to those questions. To that end, I'll first set out in this section the relevant Wittgensteinian themes with respect to questions about meaning and reference, and then I'll show how these set the stage for his expressivist insight concerning first-person privilege.

As good a place as any from which to begin is one of Wittgenstein's most enigmatic remarks. It comes in response to the suggestion that, by his lights, there's no difference between pain-behavior *with* pain and pain-behavior *without* pain, because, by his lights, "the sensation itself is a nothing":

> Not at all. It's not a Something, but not a Nothing either!...We've only rejected the grammar which tends to force itself on us here.
>
> The paradox disappears only if we make a radical break with the idea that language always functions in one way, always serves the same purpose: to convey thoughts--which may be about houses, pains, good and evil, or whatever. (Wittgenstein, 2009a, §304)

Clearly, what's under fire in *PI* §304 is an overly simplistic idea of the way language functions. And, somehow, making a radical break with that idea is supposed to remove the paradox from "not a Something, but not a Nothing either". Peter Hacker makes a suggestion about how this is supposed to work in his section-by-section exegesis of the *Philosophical Investigations*. In dealing with §304 he first

points out, rightly, that the "grammar which tries to force itself on us" is the "grammar of name and object" for talk of sensations (from the §293 story of the beetles in the boxes). (Hacker, 1990, p. 254) He also rightly observes that rejecting that "grammar" goes hand-in-hand with rejecting the idea that language always serves to "convey thoughts". But his explanation of why these should go hand-in-hand is simply that both the name-object grammar and the conveying-thought idea belong to an "Augustinian *Urbild*":

> The appearance of paradox will disappear only when one frees oneself from the trammels of the Augustinian picture of language according to which language always functions in the same way (the function of words is to name, sentences are combinations of names that describe a state of affairs). According to that pervasive *Urbild*, the purpose of communication by means of language is equally uniform, viz. to convey thoughts. (Hacker, 1990, p. 255)

While I don't want to dispute what all the Augustinian *Urbild* brings with it, I do want to draw out a little more than Hacker does why these two, the conveying-thoughts idea and the name-object grammar at play in §304, do go hand-in-hand.

There's a hint in Wittgenstein's sardonic "or whatever"[20]. The point of that jab, it seems to me, is to call into question the rather tempting idea that, as Searle puts it, "we can separate our analysis of the proposition from our analysis of kinds of illocutionary acts". (Searle, 1969, p. 31) Taking this Searlean perspective on the independence, for the purposes of philosophical analysis, of propositional content from illocutionary act, one perhaps tends to get the idea that there are simply these 'thoughts' ('propositions', 'Gedanken') floating around, as it were, with such various subject matters as houses, pains, good and evil; and then we use language either to

[20] In German: "oder was immer".

command, or to assert, or to promise various things regarding these subject matters. But the Wittgensteinian point, as I read it, is that the *availability* of these subject matters of 'thought' cannot be divorced from the "illocutionary acts" in which they're embedded.[21]

Very much related to this is another big-picture Wittgensteinian moment nearer the beginning of the *Investigations*. There, after making an analogy between the various uses of tools in a toolbox and the various uses of words in a language, he continues the analogy in connection with our tendency to think that "every word in language signifies something":

> Imagine someone's saying: "All tools serve to modify something. Thus the hammer modifies the position of the nail, the saw the shape of the board, and so on."--And what is modified by the rule, the glue-pot, the nails?--"Our knowledge of a thing's length, the temperature of the glue, and the solidity of the box."--Would anything be gained by this assimilation of expressions? (Wittgenstein, 2009a, §14)

Again, what's under fire is a misleading oversimplification, here an "assimilation of expressions". The point here about tools could be recast somewhat in the manner of §304 as follows: Isn't it easy to make a not-so-radical break with the idea that tools always serve the same purpose: to modify things--which may be the position of the nail, the shape of the board, our knowledge of a thing's length, the temperature of the glue, the solidity of the box, or anything else you please? The point here is not so much that the oversimplified assimilation of expressions is *wrong*,

[21] I hesitate about the term "illocutionary act" in rendering Wittgenstein's thought here, just because it's bound up with the locutionary/illocutionary/perlocutionary scheme. (Cf. (Austin, 1962, pp. 94f.).) And it's not clear that Wittgenstein's notion of the "uses" of words maps neatly onto any part of that scheme. Also, this point about not divorcing the subject matter from the speech act is very much on Austin's mind as well in his later lectures in *How to do things with Words*. Surprisingly, however, Austin never considers the possibility that *expressing* one's state of mind, as opposed to giving a *report*, say, about something observed with one's sense organs, should make for a distinct kind of 'illocutionary act'. Though, his discussion of "behabitive" speech acts shows him coming somewhat close to this idea.

more that it's *empty*, that putting things this way (saying tools "always serve to modify something") doesn't begin to tell us how the tools in any given toolbox are used. If after explaining to someone how to use a hammer (incorporating the not-quite-natural "modifying" locution in our explanation), that someone were to pick up the ruler in our toolbox and ask, "So, what does this thing modify?", one answer would be, "In the sense in which the *hammer* modifies the position of the *nail*, the *ruler* doesn't modify *anything*." If he took this to mean that the ruler had stopped working (whatever work it does), he wouldn't be so much mistaken as confused. He would need to make a break with his overly simplistic idea of the ways tools are used, a break that for most of us, when it comes to tools, is not at all difficult to make. One way of, as it were, shaking him out of this confusion would be to say, "The ruler doesn't modify something, but it doesn't modify nothing either!".

Wittgenstein's intended lesson, as it applies to language-use, is this: we shouldn't construe talk of houses, pains, good and evil, and anything else you please as a matter simply of conveying thoughts (with whatever "illocutionary force"); but not because we should in that case be factually mistaken: our construal will not be so much *mistaken* in that case as *empty*. And this emptiness rests on the emptiness of the idea that a thought's having the subject matter it has is always a matter of its (constituents') having this or that 'signification' (the way, for example, "houses" signifies houses). Of course, pointing this out doesn't, on it's own, tell us anything about how we're actually supposed to understand talk of houses, pains, good and evil, etc.. Simply saying that we need to make a radical break, or, for that matter, that meaning is use, or anything so programmatic, isn't going to get anyone anywhere

until it's said *just how* we're to make the radical break, *just what* about the different uses of different words gives them their different "lives"[22], or "meanings", and so on. Indeed, an important part of this chapter is aimed at tracing some of the paths Wittgenstein took in the way of getting clearer about the just-how's and the just-what's, in particular as they pertain to the expressive character of psychological self-ascriptions.

So, what Wittgenstein's radical break demands is that, to put it most generally, we stop seeing words as always 'working the same way'. But then we'll need to know what it means to see words at work, and what it means to see them working this way or that. It seems to me there are roughly three sources of illumination that Wittgenstein repeatedly drew from when he was trying to see words at work; that is, when he was trying to see their different uses rightly. And it's by drawing from these sources that we're supposed to break the spell that has us assimilating the functions of too many words one to another. These are (i) how the words are taught, (ii) how sentences containing them are verified, and (iii) in what sensible manners they may be combined.[23] And whether he's considering (i), (ii), or (iii), the running theme of Wittgenstein's treatment of our practices of self-applying psychological terms seems to be that there are important connections between those self-applications and unlearned, natural behavioral expressions of our psychological lives. Now, as I see it, the expressivist insight (which addresses a problem in epistemology), is most clearly at work when Wittgenstein draws from (ii) and (iii);

---

[22] Cf. (Wittgenstein, 1969, p. 5) and (Wittgenstein, 2009a, §432).

[23] These are certainly closely related, and (ii) and (iii) are so closely related as to be sometimes seen, perhaps, as versions of each other, but for expository purposes it's useful to break them apart.

but when he draws from (i) too, a kind of corroboration of the expressivist insight quite definitely emerges.[24]

Beginning with (i), we should consider *PI* §244, wherein Wittgenstein (or one of his voices) asks "How do words refer to sensations?":

> How do words *refer* to sensations?—There doesn't seem to be any problem here; don't we talk about sensations every day, and name them? *But how is the connection between the name and the thing named set up? This question is the same as: how does a human being learn the meaning of names of sensations?* [my italics] For example, of the word "pain". Here is one possibility: words are connected with the primitive, natural, expressions of sensation and used in their place. A child has hurt himself and he cries; then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior. (Wittgenstein, 2009a, §244a)

Here, Wittgenstein quite unapologetically turns the question of how the connection between "the name and the thing named" is set up into the question of how a human being learns the meaning of the name; of the word "pain", for example. Then he suggests one way it might go: the word "pain" is taught as a replacement of primitive pain-behavior, and is itself a bit of pain-behavior.[25] If it really is like this with sensation words, then the name-object picture as applied to them indeed seems to begin to give way. For Wittgenstein anyway, looking at our training in his preferred way, our training in the mastery of mental state self-ascription, that is, is an important

---

[24] As it should be, since the inner-access picture is cut from the same cloth, so to speak, as the inner-ostensive-definition picture. Also, it should be noted that there a quite a few more remarks of Wittgenstein's that move in the expressivist direction than I'm going to canvass simply by proceeding with (i), (ii), and (iii) as entering wedges. Often enough, Wittgenstein makes movements in the expressivist direction without any explicit consideration of (i), (ii), and (iii), and some of these will be invoked later when I discuss the underdevelopment of an expressivist theory in Wittgenstein's later work. All the same, tracing the expressivist insight like this is a way of keeping an eye on the all-important connections between the expressivist insight in Wittgenstein and his radical views on the multifarious "uses" of words.

[25] It's clear that Wittgenstein's setting this out precisely as "one possibility" is an important and easily neglected feature of his remark here (no sooner does Hacker acknowledge this feature than he seems to forget it--see (Hacker, 1990, pp. 38,9); but what's more to the present point is that even this (the expression/replacement story) should count as a possibility.

phase in the general erosion of the picture of mental state self-ascription as an affair made possible by inner ostensive definitions and consisting of inner observations and reports.

A second, indispensable phase in the erosion process comes from Wittgenstein's source (ii) of illumination: attending to the ways first-person psychological ascriptions are verified (or, *aren't*). And this again, as Wittgenstein sees things, brings out likenesses between such self-ascriptions and unlearned expressions of mentality. One of the most important (because one of the most explicit) remarks of Wittgenstein's in this area is the following (rather broad sketch of a) "plan for the treatment of psychological concepts", wherein the connection between the method of verifying first-person psychological utterances and those utterances' expressive character is undeniable:

> Plan for the treatment of psychological concepts.
> Psychological verbs characterized by the fact that the third person of the present is to be verified by observation, the first person not.
> Sentences in the third person of the present: information. In the first person present: expression. ((Not quite right.))
> The first person of the present akin to an expression. (Wittgenstein, 1980b, §63)[26]

Elsewhere Wittgenstein thrusts in much the same direction:

> To call the expression of a sensation a *statement* is misleading because 'testing', 'justification', 'confirmation', 'reinforcement' of the statement are connected with the word "statement" in the language-game. (Wittgenstein, 1970, §549)

In the *Investigations* a similar contrast is gestured at, however enigmatically, between describing one's state of mind and describing, for example, one's room:

[26] Cf. (Wittgenstein, 1970, §472)

> It is not, of course, that I identify my sensation by means of criteria; it is, rather, that I use the same expression. But it is not as if the language-game *ends* with this; it begins with it.
>
> But doesn't it begin with the sensation -- which I describe? -- Perhaps this word "describe" tricks us here. I say "I describe my state of mind" and "I describe my room". One needs to call to mind the differences between the language-games. (Wittgenstein, 2009a, §290)

It would seem the relevant differences between the language-games is supposed to have something to do with the fact that describing one's state of mind is importantly related to expressing (non-linguistically) one's state of mind, and in such a way as to exclude questions about testing, justification, confirmation, etc., questions that can arise when one, for example, describes one's room.

Of course, this kind of exclusion of questions about testing, justification, etc., is of a certain very special sort, as Wittgenstein sees it. It's of a sort such that it would be *nonsensical* even to raise such a question, and so we're brought to Wittgenstein's source (iii). As might be expected, when elsewhere in Wittgenstein's corpus it comes time to weigh what one can and can't sensibly say about one's own mental states, Wittgenstein's ruminations steer toward expressivism. An interesting place this happens is in the *Blue Book* where Wittgenstein is explicitly and specifically concerned with a certain, as it were, unassailable use of the word "I":

> To ask "are you sure that it's *you* who have pains?" would be nonsensical. Now, when in this case no error is possible, it is because the move which we might be inclined to think of as an error, a 'bad move', is no move of the game at all. (We distinguish in chess between good and bad moves, and we call it a mistake if we expose the queen to a bishop. But it is no mistake to promote a pawn to a king.) And now this way of stating our idea suggests itself: that it is as impossible that in making the statement "I have toothache" I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me. To say, "I have pain" is no more a statement about a particular person than moaning is…The difference between the propositions "I have pain" and "he has pain" is not that of "L. W. has pain" and "Smith has pain". Rather, it corresponds to the

> difference between moaning and saying that someone moans. (Wittgenstein, 1969, pp. 67, 8)

As noted already, the main thing in focus is a particular use of the word "I", not the "have pain" part of the first-person ascription. Nonetheless, the impossibility of error attached to the use of "I" in such first-person ascriptions is supposed to be on equal footing, in some sense ("it is *as* impossible"), with the impossibility of *moaning* with pain by mistake, having mistaken someone else for oneself. It's significant, of course, that we don't get from Wittgenstein an outright identification of the role of "I have pain" with that of a moan, but it's equally as significant that the impossibility of *moaning* by mistake is the one Wittgenstein chooses to shine light on the impossibility of error attached to the use of "I" "as subject".[27]

A notable point in the *Investigations* where the tie-up is made between the impossibility of mistake (and doubt, which impossibilities seem to go hand-in-hand for Wittgenstein) and natural expressions of mentality is §288:

> I turn to stone, and my pain goes on. -- What if I were mistaken, and it was no longer *pain*?...That expression of doubt has no place in the language-game; but if expressions of sensation -- human behavior -- are excluded, it looks as if I might then *legitimately* begin to doubt. My temptation to say that one might take a sensation for something other than what it is arises from this: if I assume the abrogation of the normal language-game with the expression of a sensation, I need a criterion of identity for the sensation; and then the possibility of error also exists. (Wittgenstein, 2009a, §288)

Here, the key phrase is "the normal language-game *with* the expression of a sensation". Exactly what Wittgenstein has in mind here is somewhat unclear perhaps. As I see it, there are two plausible ways of construing that "with" relation

---

[27] Wittgenstein calls the error-proof uses of "I" uses of "I" as subject, and other uses he calls uses of "I" as object in this part of the *Blue Book*. For discussions of Wittgenstein's use of "I" as subject in connection with the (broader) phenomenon of 'immunity to error through misidentification', see (Shoemaker, 1994c, pp. 80f.) and (Evans, 1982, pp. 215f.).

Wittgenstein sees obtaining between the normal language-game, or games, of mental state self-ascription and the various expressions of sensations in human behavior. It could be that the normal language-games of mental state self-ascription are games "*with*" expressions of sensations in the looser sense that (a) the games get a grip somehow only against a background of normal human expressive behavior, or in the tighter sense that (b) self-ascriptions of sensations are themselves to be thought of as (somehow enlighteningly) *akin to* pre-linguistic expressions of sensation in human behavior. When Hacker treats §288 in his four-volume exegesis (Hacker, 1990, pp. 180-3), he seems to discern only the first sense of that crucial "with"; but as I read §288, it doesn't seem at all far-fetched to have it that this word "with" is meant in *both* senses at once. In any case, it's easy to see why a game "with" the expression of sensation in sense (b) should be a game that excludes the possibility of error (even in predication), while it's more difficult to see why the possibility of error should be excluded if the sense of "with" didn't include (b).

In that case we get a radical suggestion about how to think of first-person privilege. We get the suggestion that the exclusion of the possibility of doubt and error rests not in an intimate, *direct* epistemic relation, by comparison to which *inference* from behavior, for example, is *indirect*; rather, that exclusion rests in the kinship between (certain) mental state self-ascriptions and natural, pre-linguistic expressions of mentality.

And with that, I think, the expressivist insight is sufficiently on the table. The main points are: (A) talk of sensations is different from talk of, say, houses, in a way that renders idle, as Wittgenstein might put it, the name-object conception of sensation

talk; (B) this conception begins to be eroded by likening sensation talk to pre-linguistic behavioral expressions of sensation according to insights gleaned from considering (i), (ii), and (iii); (C) this at once begins to erode the inner-observation conception of first-person privilege, supplanting it with a radically different, expressivist, conception. We can see, then, how components (α) and (β) of the expressivist insight mentioned at the outset (i.e., likening privileged utterances to natural expressions and letting this tell the story, as it were, of privilege) emerge from point (B) (i.e., from what Wittgenstein gets from considering (i), (ii) and (iii) with respect to (at least certain) privileged mental state self-ascriptions); and we can see how component (γ) (i.e., using components (α) and (β) to oppose the inner-access conception of privileged mental state self-ascription) is embodied in points (A) and (C).

It's important, furthermore to see how points (A) and (C) fit together for Wittgenstein. That is, it's important to see how, for Wittgenstein, the inner-access (or, inner-observation/perception) picture of the epistemological (metaphysical) situation and the name-object picture of the semantical (metaphysical) situation are eroded both at once by the expressivist considerations gleaned from sources (i)-(iii).[28] It doesn't seem one could, for instance, attain his kind of expressivistic reconceptualization of what was traditionally conceived as a matter of *perceptual* contact with inner items while retaining any traditional (name-object) conception of

[28] It's important to note that no amount of attention to the deliverances of (i)-(iii) will itself bring down the thought that the name-object grammar is appropriate to every 'domain of discourse', not for someone under the grips of that conception (the 'Augustinian' conception, See (Baker and Hacker, 1980, pp. 1-28)). Only *after* (or, perhaps, *as*) one is ready to release the idea that such a 'grammar' is appropriate to *every* discourse will attention to the deliverances of (i)-(iii) begin to seem at all relevant to bringing down the unthinking acquiescence in the appropriateness of that grammar *specifically* to talk of the mental.

'*reference*' to one's sensations, etc. And certainly, trying for the reverse, that is, looking to expressivist considerations for an alternative to the name-object conception of first-person sensation talk while looking to hold onto an inner-access conception of the epistemology of such talk, makes scant sense. For Wittgenstein, it seems, the inner-access picture and the name-object picture inform and sustain each other; and his expressivistic tendencies, it seems, are geared toward undermining these, as I've been saying, in one stroke. This will be important later when it comes time to discuss what kinds of approaches to the problem of mind and body are friendly to expressivistic approaches to the problem of first-person privilege.

Four Questions

Now, it's all well and good to stress that there are important likenesses, kinships, etc. between privileged utterances and pre-linguistic behavioral expressions, but when we try to move on from there, working out a serious theory, it's very soon that important and perhaps inevitably regrettable decisions have to be made; that is, if we want to continue drawing on the expressivist insight. I want to discuss here four pressing questions for anyone serious about fully working out an expressivist theory of first-person privilege. Much of the remaining discussions will be in the way of seeing how various versions of expressivism arise as various ways of responding to these four questions.

(Q1) What is to be made of the truth-aptness of privileged utterances in light of the expressivist insight? Could it be that "I'm thirsty!" is, just like a natural expression of thirst, without truth-value? This can seem a particularly embarrassing

question for the expressivist, and no sensible expressivist can hope to avoid it while also hoping for any chance in philosophical combat (as it might be seen).

(Q2) What is to be made of the epistemic status of privileged utterances in light of the expressivist insight? Could it be that "I'm thirsty!", just like a reach for a drink, doesn't represent anything known or even justifiably believed by the person exclaiming it? It will be of particular interest to see how the expressivist theories to be surveyed handle this question in their different (or perhaps not so different) ways. In any case, it's clear that this question must be addressed by anyone working from the expressivist insight.

Next there is the question, (Q3) Exactly which privileged mental state self-ascriptions in the first-person present are expressive as opposed to reportive? Could it be that the unsolicited *exclamation* "I'm thirsty!" made in the Sahara desert, say, is in all relevant (expressive-as-opposed-to-reportive) respects just like the response "I'm thirsty," given on a questionnaire in a psychology lab, say? And if not, how are the lines drawn? Wittgenstein himself remarked that "I am afraid" may be sometimes very far removed from a cry of fear and sometimes very close to it. (Wittgenstein, 2009b, §83) He didn't say much about what that farness and nearness consisted in, but he certainly saw that things weren't so straightforward as to count *all* uses of "I am afraid" equally as akin to cries of fear.

And finally, (Q4) What is the relationship between the far removed cases (if any) and the 'basic' cases vis-à-vis their privilege? Does the privilege attaching to the far removed cases in any sense *derive from* some essential connection they bear to the basic cases? A central concern here, for the expressivist, is that it would be

upsetting if “I’m thirsty,” said in a psychology lab were every bit as distinctively privileged as “I’m thirsty!” exclaimed in the desert, but the privilege of the former sort of utterance didn’t in any sense derive from an essential connection to the privilege of the latter sort of utterance. The scope of the expressivist program would be, in that case, so limited as to render it pointless; expressivists are in effect forced out of the discussion of privilege if they only get to deal with the privilege attaching to mental state self-ascriptions that end in exclamation points.

So, these are the questions facing the philosopher who would work from the expressivist insight in sorting through the problem of first-person privilege. And the evaluative task of this dissertation will be to judge expressivist theories on two scores: first, whether they can satisfactorily, or even plausibly, answer these four questions; and second, whether they can do so in such a way as to remain the distinctive, radically non-Cartesian sort they were originally prized for being.

Now, the ways these questions have actually been dealt with by careful theorists working from the expressivist insight will be the subject of detailed discussion in later chapters. My concern in the remainder of this chapter, though, is to set out how these four questions can quickly lead to serious difficulties; that is, given a certain simple way of following through with the expressivist insight. While serious advocates of the expressivist insight won’t be caught proceeding in the simple way I’ll be setting out presently, it can be useful to cast their sophistication as a matter of making improvements on the simple way of proceeding. Of course, there isn’t the slightest bit wrong with being an improvement on a simple theory; so casting things this way shouldn’t, in itself, be upsetting to sophisticated advocates of the

expressivist insight. Moreover, casting things this way sets the stage for a later question as to whether the point, whether the original spirit, of the expressivist insight is plausibly preserved in any given sophisticated version of expressivism about first-person privilege. That's because the sophisticated theories will have to face a comparison with the simple theory on this score, and there will be no doubt that the point of the expressivist insight is in full bloom, so to speak, with the simple version to be set out forthwith.

Simple Expressivism

The simple version that I want to take as a kind of basic and instructive version of expressivism casts the kinship between privileged utterances and pre-linguistic behavioral expressions of mentality as a very close one indeed. It thus fares quite well with respect to tidiness. Regarding (Q1), the simple response is simply that privileged utterances lack truth-values; they aren't literally true or false, just like pre-linguistic expressions aren't. And regarding (Q2), the simple response is that privileged utterances, just like pre-linguistic expressions, aren't 'privileged' in any *epistemic* sense: the privileged utterance doesn't betoken any very highly, or infinitely highly, justified belief; nor does it betoken any degree of certainty, however great; nor does it betoken any knowledge, however secure. Regarding (Q3), the simple response is simply that *all* privileged mental state self-ascriptions are *equally* as expressive (as opposed to reportive), equally as near to cries, reaches for water, etc.. Of course, given this response to (Q3), (Q4) becomes irrelevant; there simply won't be any privileged mental state self-ascriptions that are supposed to be far

removed from the basic kinds, and so there won't be any question of relating their privilege to the privilege of the basic kinds.

Now, clearly, this way of keeping with the expressivist insight in response to (Q1)-(Q4) is crying out for opposition, and I'll set out the opposition presently, but right now I want to say a little about how someone could be so impressed by the expressivist insight that she would even consider responding thus to (Q1)-(Q4).

Starting with the first two responses, it seems to me that surely something like the following kind of thinking would be at work in a simple expressivist's mind. The features that place non-linguistic, behavioral expressions of mentality beyond epistemic reproach, as it were, are very much bound up with their being truth-valueless and their bearing, indeed, no epistemic status whatsoever; so how could we, as it were, borrow one set of features without borrowing the other? To put it another way, how are we supposed to liken the *senselessness* of asking "How do you know?" when somebody offers "I'm in pain" to the *senselessness* of asking that question when somebody cries out "Arghhh!" without to that extent likening the *sense* of "I'm in pain" to the *sense* of "Arghhh!". But *there is no* sense of "Arghhh!". In any case, "Arghhh!" isn't true or false, though clearly it may be sincere or insincere; that is, it may be a genuine expression of pain, or a pretended one. Furthermore, as regards specifically the simple expressivist's response to (Q2), it would seem the whole point of the expressivist tack, meant as it is to provide a radical alternative to the Cartesian scheme of inner-observation, demands nothing short of an unequivocal rejection of what Hacker calls the "cognitive assumption" (i.e. the assumption that avowals represent genuine self-knowledge):

> So what the objection to the cognitive assumption rules out is [the use of] the form of words 'I know I am in pain'…that lies at the heart of the philosophical tradition which informs our thought, i.e. its use to ascribe to oneself or to claim a form of knowledge of the subjective which is derived from introspection and is both indubitable and infallible. That conception, which was part of the picture of the mental as better known than the physical and of self-knowledge as based on privileged access to, and private ownership of, experience, is philosophers' nonsense. (Hacker, 2005, p. 269)

Turning now to the simple way of handling (Q3) (which itself handles (Q4)), it would likely stem from a recognition of a point made already made, that expressivists won't have too much to say in the discussion of first-person privilege if it's admitted that the distinctively first-person privilege attaching to sentences that, putting it roughly, don't end in exclamation points doesn't in any important way derive from the first-person privilege attaching to sentences that do so end (in a way, that is, that respects the goal of replacing the inner-access picture by means of the expressivist picture). The simple response to (Q3) quite clearly obviates any such concern.

Difficulties for Simple Expressivism

I think it's no understatement to say that these responses to (Q1)-(Q4) invite quite severe criticism, especially regarding (Q1) and (Q2), but certainly regarding (Q3) (and (Q4)) as well.

For the standard line against a simple expressivist response to (Q1) (the truth-valueless response) we can turn to Peter Hacker's first edition of *Insight and Illusion*. There he writes that, "The fundamental features around which the counter-argument must revolve concern the fact that 'I am in pain' has, after all, a structure. As such it is complex or articulated." (Hacker, 1975, p. 266) He proceeds to give nine arguments in all that revolve around these fundamental features; I present just three:

> Fifthly, 'I am in pain' seems to contradict 'No one is in pain' and to imply 'Someone is in pain'. Sixthly, 'I am in pain' can appear in molecular sentences, e.g. 'I am in pain and the doctor has not come', without the molecular sentence lacking a truth-value or being non-truth-functional. Seventhly, 'I have a pain' can appear as a premise in a valid argument, e.g., 'All persons with a pain of such and such a kind suffer from disease *D*, I have a pain of such and such a kind, therefore I suffer from disease *D*.' (Hacker, 1975, p. 267)

For the standard line against simple expressivism's response to (Q2) (the 'non-cognitive' response), we turn again to Hacker's first edition of *Insight and Illusion*:

> Firstly…either the non-cognitive thesis is false, or one can know the truth of the conclusion of some sorts of valid arguments, without knowing the premises. Secondly, 'He is in pain' is knowable, and 'I know he is in pain' is undoubtedly a legitimate sentence. But what of 'He knows that I am in pain'? Can it be that I know that he knows that I am in pain without it being possible for me to know that I am in pain?...Finally, in the standard analysis, 'I remember that *p*' implies that I previously knew that *p*. So 'I remember that I was in pain' implies that it was the case that I knew that I was in pain, or else the standard analysis of 'remember' must be modified. (Hacker, 1975, p. 276)

Turning to (Q3), even those sympathetic to the notion that certain, as it were, sensation-heralding sentences are importantly akin to primitive expressions of sensation (and who will in the first place admit a distinctive first-person privilege) are likely to find it highly implausible that mental state self-ascriptions are one and all equally near to cries, reaches, etc., or else lacking any distinctive first-person privilege. I suppose the most incisive example to be brought against the simple response to (Q3) would be that of saying about oneself that one is *not* in pain, or *not* thirsty, etc. That's because, on the one hand, it's hard to deny that "I'm *not* in pain" can be every bit as privileged as "I *am* in pain", and on the other hand, it's hard to see that the former can be, like the latter, learnable as some kind of replacement of any pre-linguistic expression of mentality in behavior. Of course there's more working

against a simple response to (Q3) than just this negation trick. Looking one last time at Hacker's first edition of *Insight and Illusion*, we see him writing about such self-ascriptions as "I see red", "I remember my third birthday", "I am thinking of Gödel's theorem" and "I believe that it is Tuesday today" that:

> [T]he assertion of such sentences (and of their negations) does not, in any sense, replace a primitive natural form of behavior, and cannot, by any stretch of the imagination, be thought of as having the logical status of outcries and facial expressions. Yet such sentences do share the central perplexing features of 'I am in pain'. (Hacker, 1975, p. 268)

The "central perplexing features" Hacker has in mind are more or less those of groundlessness, authority and transparency.

Such, then, are the difficulties facing a simple expressivist. In the following pages we'll examine expressivist options for dealing with these difficulties. These will come in two varieties: either the expressivist can defend (parts of) simple expressivism against these objections, or else she can abandon (parts of) simple expressivism in the face of these objections. To the extent that an expressivist defends simple expressivism, to that extent I'll be calling her approach a *conservative* one; and to the extent she abandons simple expressivism, to that extent I'll be calling her a *progressive* expressivist. The problems facing the two approaches are really two sides of the same coin, a ubiquitous phenomenon in philosophy: the conservative arguments in defense of simple expressivism are, to the extent that they are indeed valid, in danger of being recast as *reductios*; and the progressive movements away from simple expressivism leave the expressivist in danger of losing all contact with the original spirit of the expressivist insight.

Needless to say, given the many ways one can make adjustments here or there, the ways one can take a more or less conservative line on any given question, a more or less progressive line on any given question, the number of possible responses to the difficulties facing simple expressivism is too high to estimate. It would be absurd to try to map them all out and evaluate all their strengths and weaknesses, saying how far each one succumbs to this or that danger. Instead, in the following chapters I'll examine a small handful of expressivisms that have actually been worked out, starting with a fuller examination specifically of Wittgenstein's own expressivist theory (or lack thereof) in the next chapter. After seeing how these expressivist theories handle the difficulties set out above, it will remain to assess the prospects for any satisfactory expressivist theory of first-person privilege.

CHAPTER III
WITTGENSTEIN'S (LACK OF AN) EXPRESSIVIST THEORY

Introduction

In this chapter I want to consider to what extent Wittgenstein himself could be thought of as a serious expressivist theorist about first-person privilege. How far, and in what ways, does he diverge from simple expressivism, if he does at all? That is, how does he handle the four questions set out last chapter? He constantly points out likenesses, kinships, and correspondings, but what's to be made of it all? Does Wittgenstein or doesn't he chart a path from the expressivist insight to a fully worked out expressivist theory of privileged utterance? It should be clear from what I've been saying hitherto that I think the answer to this last question is, No. So I'm going to take this chapter to discuss why and in what sense Wittgenstein managed to avoid advancing any fully worked out expressivist theory of first-person privilege.

The first thing to point out is that this question cannot be treated properly out of the context of a bigger question, a bigger question about Wittgenstein as a systematic, if you please, avoider of philosophical theorizing. That is, it can seem Wittgenstein's 'way' of doing philosophy is rather self-consciously constructed just so as to leave one wanting more; and this bit about the absence of an *expressivist* theory is only a special case, as it were, of the way he does philosophy in general. Then again, only those will be left wanting more, presumably, who somehow or other *miss the point* of whatever it was Wittgenstein had to say concerning this or that philosophical dispute. And, presumably, if one really gets Wittgenstein's point, this feeling of a theoretical job left undone simply won't afflict him, as Wittgenstein might say. This is roughly the way it's supposed to go; if Wittgenstein was right, that

is. In any case, he repeatedly and explicitly denounced *at least some kind of* philosophical theorizing in general, and so it's best to locate the specific concern about Wittgenstein as a serious expressivist within the larger concern about Wittgenstein as an anti-theorist.

Norman Malcolm approached this larger question, in connection with a particularly vexed, smaller question about how to read Wittgenstein on private language, when he reviewed the *Philosophical Investigations* in 1953:

> [Wittgenstein] says that he changes *one's way of looking at things* (§144). What is it that he wishes to substitute for that way of looking at things that is represented by the idea of private language? One would *like* to find a continuous exposition of his own thesis, instead of mere hints here and there. But this desire reflects a misunderstanding of Wittgenstein's philosophy.[29] (Malcolm, 1966, p. 77)

Malcolm's follow up to this leaves a lot to be desired. He more or less proceeds to produce a barrage of the most problematic of Wittgenstein's remarks on method, as though they were some great help:

> [Wittgenstein] rejects the assumption that he should put forward a *thesis* (§128). "We may not advance any kind of theory" (§109). A philosophical problem is a certain sort of confusion. It is like being lost; one can't see one's way (§123). Familiar surroundings suddenly seem strange. We need to command a view of the country, to get our bearings. The country is well known to us, so we need only to be *reminded* of our whereabouts. "The work of the philosopher consists in assembling reminders for a particular purpose" (§127). "The problems are solved, not by giving new information, but by arranging what we have always known" (§109). When we describe (remind ourselves of) certain functions of our language, what we do must have a definite bearing on some particular confusion, some "deep disquietude" (§111), that ensnares us. (ibid.)

And then on the heels of this, after everything's supposedly crystal clear, we get: "Thus we may not complain at the absence from the *Investigations* of elaborate theories and classifications." (ibid.)

---

[29] Section references to (Wittgenstein, 2009a).

In this chapter I'm going to try to shed some light on these various anti-theoretic slogans themselves. Whereas Malcolm seems to take them for an explanation of the absence of "elaborate theories", I'm taking them as my problem in this chapter. Just what is going on with all this talk of *theories* and *confusion* and *reminders*? I could add: *dissolution*, *therapy*, and *nonsense*. There are others too. These are familiar buzzwords for readers of Wittgenstein, but if anything is clear about Wittgenstein scholarship, it's that there isn't any general agreement about what work these words are doing in his later thought. Indeed, this is, and will be for a long time, I'm sure, *the* exegetical question in Wittgenstein scholarship.

The structure of this chapter, in large part a contribution to that exegetical question, will be as follows: I'll first (a) set out my own view of Wittgenstein as a methodically non-theorizing philosopher, but this in the largest part only with an eye specifically toward setting out my view of Wittgenstein's (lack of an) expressivist theory of first-person privilege later in the chapter; I'll then (b) discuss *non-assertory* and *non-argumentative* readings of Wittgenstein as therapist, trying to show the relative benefits of my reading by contrast; and then finally I'll (c) set out my view of the Wittgensteinian responses to (Q1)-(Q4) set out last chapter; that is, my view of Wittgenstein as an expressivist about the problem of first-person privilege.

Wittgenstein's Anti-Dogmatic Later Philosophy

I'll begin with three preliminary points. The first point has to do with something that might be called "the problem of the criterion" for understanding Wittgenstein's later philosophy. About arriving at the correct understanding of certain problematic philosophical concepts (e.g. 'knowledge' and 'justification') and

how to apply them, Roderick Chisholm pointed out that a problem seems to arise in that one would seem forced to choose between what have been called a *methodist* approach and a *particularist* approach. (Chisholm, 1982, pp. 61ff.) On a methodist approach, one seeks first to arrive, in abstraction from any particular application of a concept, at the proper 'analysis' of the problematic concept. And then whether and how it applies in any given instance is left to sort itself out in accordance with that analysis, with surprising results to be expected and accepted. On the reverse approach, the particularist approach, one begins with particular instances about which one already has an idea as to whether and how the problematic concept applies, and one only works backwards from there to an analysis of the problematic concept. The so-called problem of the criterion is, roughly, that it seems neither of these approaches makes sense precisely because it presupposes and yet excludes the other.

Now, it seems we get something analogous to this situation as we set out to understand not only Wittgenstein's overall conception of (good) philosophy, but also the ways it gets applied to any particular philosophical puzzle. So, on the one hand, as 'particularists', we might say that to deal with the difficulty of understanding Wittgenstein's later work we need to begin by working through the particular ways Wittgenstein deals with particular problems; and only after that, says the particularist, will we be able to arrive at any kind of understanding of Wittgenstein's methodological pronouncements.[30] On the other hand, from the 'methodist' approach, we'll be more inclined to say, with Robert Fogelin, that:

> The task of interpreting Wittgenstein's later writings is not -- as with the *Tractatus* -- one of deciphering opaque passages. Sentences in his later

[30] This seems to be the sort of approach at work in David Stern's introduction to Wittgenstein's *Philosophical Investigations*, for instance. Cf. (Stern, 2004).

> writings are usually transparent as they stand. The challenge, where there is one, is to appreciate the philosophical significance that Wittgenstein assigns to them. (Fogelin, 2009, p. xi)

Appreciating the "philosophical significance" of Wittgenstein's "usually transparent" sentences just is, it becomes clear in Fogelin's work, understanding Wittgenstein's *method*, namely as something *therapeutic*.

Just like with the classical problem of the criterion, now, it seems to me each of these approaches, insofar as it's meant to exclude the other, doesn't make sense precisely because it presupposes the other. It seems to me you couldn't possibly begin to understand what to make of Wittgenstein's particular treatments of particular problems if you simply *left it open* at the outset whether it was characteristic of Wittgenstein's method, for instance, never to assert anything.[31] And on the other hand, it seems to me you couldn't possibly begin to understand a methodological dictum like "don't ask for the meaning but for the use" precisely as a *therapeutic* tool if that understanding were, *per impossibile*, arrived at in abstraction from an appreciation of Wittgenstein's way with particular philosophical puzzles. But the answer lies in the statement of the problem, of course, and it's simply this: appreciation of method and appreciation of method in action *arise together at once*.[32] As with many things human, it hardly needs to be said, the arising together at once of these dual appreciations will only seem paradoxical if we're concerned to program a computer to model it.

---

[31] We'll eventually come to inspect one particular non-assertory construal of Wittgenstein's method.

[32] This is, indeed, analogous to the solution to the classical problem of the criterion. Cf. (Fumerton, 2008).

For the next preliminary point, I'd like to look again at Malcolm's line on Wittgenstein as an anti-dogmatist. I've already gestured at something peculiar in it, something awkward, and I'd like to put my first point as a reaction to that awkwardness. The problem I see for Malcolm lies in his talk of "hints" and "elaborate theories". He says that the wish for something *more than* hints betrays a misunderstanding, and he says that what we're left without are *elaborate* theories. But the slogans he cites don't seem to have anything to do with unelaborated upon hints. Rather, these slogans are about refraining from advancing *any thesis* and about arranging what we *already know* (to pick two of the least picturesque). Now, these wouldn't seem to have anything to do with failing to elaborate on hints. Indeed, these would seem to leave hinting decidedly out of the question.

I think this is an important awkwardness, because I think that the underdevelopment of Wittgenstein's views on some topics can very easily be mistaken for having something to do with his therapeutic approach in philosophy. While I believe Malcolm is dead right about hints and the lack of elaboration all throughout Wittgenstein, I think he's simply wrong in the way he connects that up with Wittgenstein as a systematically anti-dogmatic, therapeutic philosopher. Actually, Malcolm doesn't do much at all in the way of plausibly connecting these two up; he simply places them side by side, trying to let the therapeutic aspect explain the hinting aspect; hence the awkwardness. What I want to do is to take the occasion of Malcolm's awkward juxtaposition to say right away that I don't think there is, actually, any significant way in which these two do connect up. Wittgenstein *did* leave some philosophical work undone; and *good* philosophical work, not just *bad*. I

think it helps to remember what Wittgenstein said in his preface to the *Investigations*: "I should not like my writing to spare other people the trouble of thinking. But if possible, to stimulate someone to thoughts of his own." (Wittgenstein, 2009a, p. 4) An important part of my reading of Wittgenstein as therapist is that room is made for his having left hints that *should* be elaborated. Why this can seem problematic will come out in due course, but it's an extremely important point to fasten on in order to get straight about the real workings, as I see them, of Wittgenstein's anti-dogmatic approach.

The third preliminary point I want to make is in the way of encouraging a certain freedom of mind with respect to certain problematic "titles". I draw inspiration from the *Blue Book*:

> If, e.g., we call our investigations "philosophy", this title, on the one hand, seems appropriate, on the other hand it certainly has misled people. (One might say that the subject we are dealing with is one of the heirs of the subject which used to be called "philosophy".) (Wittgenstein, 1969, p. 28)

What I want to suggest, and what will perhaps seem sacrilege to many Wittgenstein sympathizers, is that we bring a whole family of terms in with "philosophy" for similar treatment, viz. "solution", "thesis", "argument", "explanation" and even "theory". It seems to me that every one of these, no less than "philosophy", has an on-the-one-hand-appropriate and an on-the-other-hand-misleading quality to it when it comes to discussing Wittgenstein's later thought. Just like, for Wittgenstein, there is, apparently, a *good* kind of philosophy and a *bad* kind of philosophy, I think it helps greatly to allow at the outset for good and bad kinds of solution, argument, etc.[33]

[33] The details about how this could be will only emerge in the discussions to follow.

And besides good and bad kinds of solution, argument, etc., it seems the terms "solution", "argument", etc., all admit of such perfectly thin, innocuous construals that forbidding them any favorable entrance into discussions of Wittgenstein begins to seem itself something of a superstition. [34] For instance, if we allow that "S's theory of x" can amount to nothing more than simply what S would have to say in response to, "Tell me about x", or "What do you think of R's theory of x?", then I don't see any reason we shouldn't allow for W's theory of x. And if we think of "S's making an argument to R that *p*" as sometimes amounting to nothing more than "S's trying to arrange *q*'s and *t*'s so that R's appreciation of them will have R acquiescing in *p*", then, again, I think we can have Wittgenstein making arguments. Similarly for "solution", "thesis" and "explanation", these admit of such thin senses that one should avoid straining too much at them.

So much for preliminaries. I'd now like to try and cut very closely to what I see as the heart of Wittgenstein's anti-dogmatism. For help, I'll turn to an indispensable selection from Wittgenstein's *Lectures on the Foundations of Mathematics*:

> You might say, "How is it possible that there should be a misunderstanding so very hard to remove?"
>
> It can be explained partly...by a quotation from Hilbert: "No one is going to turn us out of the paradise which Cantor has created."
>
> I would say, "I wouldn't dream of trying to drive anyone out of this paradise." I would try to do something quite different: I would try to show you that it is not a paradise--so that you'll leave of your own accord. I would

---

[34] Stewart Candlish and George Wrisley, in their work on the private language arguments in Wittgenstein's *Philosophical Investigations*, point out that anxiety about anything called "argument" in Wittgenstein depends on a "tendentiously narrow notion of argument" and is "a reaction against some drastic and artificial reconstructions of the text by earlier writers". (Candlish, Wrisley, 2008) I think their point is made all the more plausible in light of the discussions to follow.

> say, “You’re welcome to this; just look about you.” (Wittgenstein, 1976, pp. 102, 3)

Wittgenstein is here responding to the question how there could possibly be the kind of misunderstanding he sees himself battling against. His response is unforgettable. He says we misunderstand him because we think he’s trying to *drive us out* of paradise, when instead he’s trying to show that there *isn’t any* paradise, so that we’ll leave of our own accord. This, to me, encapsulates Wittgenstein’s anti-dogmatism as much as anything could. My reading of it is simple and relies on a contrast between falsehood and emptiness (alternatively: senselessness). As I read him, Wittgenstein isn’t out to show us that any particular philosophical view is *false*, he rather wants to show us the *emptiness* of the pictures and words with which we would express the philosophical view in question, so that we’ll give these up of our own accord. We won’t give them up because it’s been *proved* that we *have* to, but simply because we aren’t captivated by them anymore. We give them up, again, not as false, but empty.

In my estimation, this is perhaps the single most significant aspect of Wittgenstein’s anti-theoretic approach; and to the extent that rival readers of Wittgenstein would agree in acknowledging this aspect of his approach, it seems to me they would agree on the largest part of what really matters about Wittgenstein as a non-theorist; they can all say, quite rightly and confidently, that Wittgenstein is no (traditional philosophical) theorist in the straightforward sense that he doesn’t *deny* any (traditional philosophical) theory, but tries to show its *senselessness*.

Well, I say that that’s straightforward, but I suppose it can only be as straightforward as the term “senseless” itself is, and what it would be to “show” the senselessness of some traditional philosophical theory. Importantly, however, the

sense in which traditional philosophical theories are 'senseless' and can be 'shown' to be so is worth being careful about. Indeed, the most novel part of my reading of Wittgenstein's anti-dogmatism consists in a crucial, as I see it, distinction between kinds of senselessness, or kinds of nonsense, and what this means for the ways traditional philosophical nonsense can be shown to be so.[35] In order to get my distinction out on the table I want to set out what I think are four unargued, highly contentious (and two of them even insulting to some, I imagine) claims of Wittgenstein's, the embracing of which largely *constitutes* his turning away from what he would call philosophical dogmatism. The first three, while, as I say, unargued, are still supported in a different sense by Wittgenstein, a sense of support that, as I see it, is crucial to the *therapeutic* aspect of his later philosophy. The last, though, goes completely without support as far as I can tell in Wittgenstein's work. Now, if it's complained that, by my lights, these four will land Wittgenstein right back in the land of the dogmatists, so to speak, and that my reading is therefore uncharitable and unsatisfying, then all I can ask is that it be considered whether by my lights it must be that these four are the *same sort* of dogmas Wittgenstein wouldn't even call "false", but "nonsensical", the same sort that, on my reading, he used these very four 'dogmas' as weapons against. If not, and I do believe not, then I think we can say *at least* that, by my lights, Wittgenstein can't be brought in on the charge of breaking his promise to abstain from *what he would call* "philosophical dogmatism". Any such charge will, I hope, be seen as akin to a creationist's charge against an evolutionist that, because the latter appeals crucially to vast stretches of

[35] I use "senseless", "nonsensical" and sometimes "absurd" (and their forms) interchangeably throughout.

time (billions of years) in explaining the origins of species, therefore Time is the evolutionist's God.

The first two of these Wittgensteinian 'dogmas', or better "insight[s] into the workings of our language" (Wittgenstein, 2009a, §109), go together; they have to do with ways Wittgenstein saw that we can be misled by our 'forms of expression'. To begin with, probably the most fundamental Wittgensteinian idea (along side the next one coming up) is that "one of the great sources of philosophical bewilderment [is that] a substantive makes us look for a thing that corresponds to it." (Wittgenstein, 1969, p. 1) Just to get an inkling how important this idea is for Wittgenstein, consider:

> But let me remind you here of the queer role which the gaseous and the aethereal play in philosophy,--when we perceive that a substantive is not used as what in general we should call the name of an object, and when therefore we can't help saying to ourselves that it is the name of an aethereal object…This is a hint as to how the problem of the two materials, mind and matter, is going to dissolve. (Wittgenstein, 1969, p. 47)

Here Wittgenstein is talking specifically about the "problem of the two materials, mind and matter", but anyone who knows Wittgenstein knows that this sort of thing, this understanding and curbing of the impulse to find an object corresponding to a substantive, is supposed to help with nearly every traditional philosophical ontological puzzle.

The next way we can be misled by our forms of expression (the second Wittgensteinian insight) is that certain associations called up by our forms of expression can have us mistaking matters of *convention* (that is, matters of "grammatical rule") with matters of fact. On the impossibility of one's being able to

know whether someone else is in pain, and so having to be content with conjecturing about the other's pain, Wittgenstein wrote:

> [Y]ou did not state that knowing was a goal which you could not reach, and that you have to be contented with conjecturing; rather, there is no goal in this game. Just as when one says "You can't count through the whole series of cardinal numbers", one doesn't state a fact about human frailty but about a convention which we have made. Our statement is not comparable, though always falsely compared, with such a one as "it is impossible for a human being to swim across the Atlantic"; but it is analogous to a statement like "there is no goal in an endurance race". (Wittgenstein, 1969, p. 54)

These two points are the backbone, you might say, of Wittgenstein's later approach. They encapsulate so much of what he's trying to do with traditional philosophical puzzles: he's trying to point out the ways they result from our being misled (through false comparisons, false analogies) into "philosophical bewilderment". What is the *argument* that these are the analogies, the comparisons, that take us in, or that they are indeed false analogies? On my reading it's important that Wittgenstein never offers any. That's because on my reading it's part of the *therapeutic* aspect of his philosophy that *no argument* could possibly be in order, that Wittgenstein's only hope for a 'breakthrough' is to put his finger right on the analogy that misleads us in such a way that we would be compelled to admit that, indeed, this was the confused path we had ourselves taken. He writes:

> One of the most important tasks is to express all false thought processes so true to character that the reader says, "Yes, that's exactly the way I meant it".
>
> Indeed, we can only prove that someone made a mistake if he (really) acknowledges this expression as the correct expression of his feeling.
>
> For only if he acknowledges it as such, *is* it the correct expression…What the other person acknowledges is the analogy I'm presenting to him as the source of his thought. (Wittgenstein, 2005, p. 303)

Here it comes out clearly that the only sense in which Wittgenstein would take himself to 'prove' to anyone that the person had been misled by an analogy is by "*presenting*" to him, that is *suggesting* to him, that such-and-such was the "source" of his thought in a way "so true to character" that the person should acknowledge it for the "false thought process" that it is.[36]

The third unargued insight has, again, to do with what pertains to grammar, with what pertains to the rules (however open-ended) we play by when we use language. Not only are many 'metaphysical' statements, for Wittgenstein, disguised grammatical statements, but descriptions about how our sentences are *explained* and certain of the ways they're *verified* are disguised grammatical statements as well:

> It is part of the grammar of the word "chair" that this is what we call "to sit on a chair", and it is part of the grammar of the word "meaning" that this is what we call "explanation of a meaning"; in the same way to explain my criterion for another person's having toothache is to give a grammatical explanation about the word "toothache" and, in this sense, an explanation concerning the meaning of the word "toothache". (Wittgenstein, 1969, p. 24)

Wittgenstein was fond of flatly asserting such things, though in *PI* §354 he did, in vintage therapeutic form, have a try at diagnosing why we might lose sight of his all-important distinction between *criteria* and mere *symptoms*. There he wrote, foreshadowing one of the most important post-Wittgensteinian contributions to philosophy, viz. Quine's holism, that it's the "fluctuation in grammar between criteria and symptoms [that] makes it look as if there were nothing at all but symptoms".

[36] Cf. (Wittgenstein, 2009a, §254).

Still, of course, we don't have anything at all here that could count as an *argument* that his distinction holds.[37]

One last bold assertion from Wittgenstein, the last 'dogma' he used in the fight against *what he called* philosophical dogmatism, concerns the "point" of our "language-games":

> It is only in normal cases that the use of a word is clearly laid out in advance for us; we know, are in no doubt, what to say in this or that case. The more abnormal the case, the more doubtful it becomes what we are to say. And if things were quite different from what they actually are--if there were, for instance, no characteristic expression of pain, of fear, of joy; if rule became exception, and exception rule; or if both became phenomena of roughly equal frequency--our normal language-games would thereby lose their point. (Wittgenstein, 2009a, §142)

This one goes completely without support in any way, as far as I can tell. No argument for it, certainly, but no diagnosis of why we might be led to think otherwise either. Just a bold (and incomparably inspired, if you ask me) Wittgensteinian insight into the workings of our language.

These four insights are important in three respects: first, in my judgment, taken together they make up the very spirit of the slogan "meaning is use" so often used to sum up Wittgenstein's later philosophy; second, they aren't, as I've been saying, in any proper sense argued for in any place I know of in his work; and third, an immensely important part of my reading of Wittgenstein, they aren't themselves *rules of grammar*, the very kind of grammatical rules that, according to the second insight, get mistaken for describing necessary features of reality. I want to spend some time on this last point, because it's central to so much of what I'm going to say about Wittgenstein as a (non)theorist.

---

[37] For a (Wittgenstein-friendly) discussion of the relationship between Wittgenstein and Quine on this point, see (Glock, 1996, pp. 209f.).

Wittgenstein writes that philosophical problems are "solved through an insight into the workings of our language, and that in such a way that these workings are recognized -- *despite* an urge to misunderstand them." (Wittgenstein, 2009a, §109) What are these workings of our language? And in what does our misunderstanding them consist? I think that in order to understand Wittgenstein's later philosophy, we need to see that, for him, our "urge to misunderstand" the "workings of our language" is *not* an urge to take such-and-such for grammatical rules when really they weren't any such thing. That is, not when 'grammatical rules' are the sort seen by Wittgenstein as the better expression of certain metaphysical claims of necessity, à la the second of his insights just laid out.

Essentially, what I'm trying to combat with this is the view that would take such a Wittgensteinian claim as "private ostensive definition is an illusion" to be doing anything like what's done by the claim that "every rod has a length", which latter is best understood as a rule of grammar, something one may well violate. One can, because of 'grammatical illusions'[38], be seduced into thinking that "every rod has a length" is, for instance, a statement made true by the essential nature of rods, a statement made true by every possible rod's having a certain property, length. But I don't see that someone has been seduced into anything, misled by any form of expression, is under any grammatical illusion, who says, "The rod in the freezer has no length"; I can only see that he's breaking a rule of grammar, that is, a rule of *our* grammar, a rule about the grammar of *our* word "rod". If someone were to be under the illusion that, for instance, one could promote a pawn to a king in chess, I don't

[38] Cf. (Wittgenstein, 2009a, §110) where Wittgenstein said that one's temptation to say that "language (or thinking) is something unique" is the result of grammatical illusions.

think this would be the kind of grammatical illusion that, on my reading, Wittgenstein saw as the "cloud-castles" his investigations were meant to "destroy"[39] (Wittgenstein, 2009a, §118). The idea is that *certain* metaphysical claims, like for instance that one must know whether she's in pain, *can* be usefully compared to a claim like "every rod has a length"; but certain *other* metaphysical claims, claims that involve the philosopher in such fantasies as that of private ostensive definition, or that of necessities in the natures of things, these other metaphysical claims *cannot* be usefully so compared, and *neither* can the claims that the metaphysician's fantasies are just that, fantasaical.

Consider the following remark from the *Blue Book*:

> The cases in which particularly we wish to say that someone is misled by a form of expression are those in which we would say: "he wouldn't talk as he does if he were aware of this difference in the grammar of such-and-such words, or if he were aware of this other possibility of expression" and so on. (Wittgenstein, 1969, p. 28)

I'd like to point out, first of all, that the possibility of someone's being unaware of any "difference in the grammar of such-and-such words" seems just to be, for Wittgenstein, the possibility of someone's neglecting his four unargued insights set out above. At least, that's the reading I'm proposing. Most notably, it's neglect of the third, that certain of the ways our statements are verified pertain to their grammars, that will have one overlooking certain grammatical differences, or, better put, that will have one overlooking that certain differences *are grammatical*. While most people will agree that two sentences (e.g. one in mathematics and one in astronomy) get verified differently, many won't agree with Wittgenstein that these are the grammatically significant differences he believes they are.

---

[39] The German for "cloud-castles" being "Luftgebäude".

The second thing, the main thing, about the above remark that I'd like to point out, I believe I can point out by changing slightly what Wittgenstein says "we would say", and then by asking a question. I'd like to change it to: "The confused philosopher wouldn't talk as he does if he were aware of this *rule of grammar*", and I'd like to think of rules of grammar as straightforward, conventional, violable rules like "every rod has a length". So, I ask, Is it supposed to be that Wittgenstein's benighted philosopher will stop talking as he does because the rule of grammar Wittgenstein would point to, the one the philosopher is unaware of, runs something like "Don't talk that way" (e.g. like the rule running "Don't talk of rods without lengths")? Or will the metaphysician stop talking that way, not because he found a rule saying "Don't", but because he sees *other* rules, other grammatical features, which, when seen *in the right light*, relieve him of any temptation to talk the way he wanted? We can also take Wittgenstein's remark from a page earlier that "philosophy, as we use the word, is a fight against the fascination which forms of expression exert upon us" (Wittgenstein, 1969, p. 27), and we can ask, Is "the fascination" Wittgenstein battles simply an erroneous belief that such-and-such was the grammatical rule, when it wasn't? Like the mistaken belief that you could promote a pawn to a king in chess? I simply don't think that's right. I don't think that just because we lose our philosophical "fascination" upon surveying rules of grammar, that therefore it's among the rules of grammar that our "fascination" is to be lost.

Consider again a remark from the *Blue Book*, the closing remark:

> The kernel of our proposition that that which has pains or sees or thinks is of a mental nature is only, that the word "I" in "I have pains" does

> not denote a particular body, for we can't substitute for "I" a description of a body. (Wittgenstein, 1969, p. 74)

Applied to this remark, the idea I'm after would be that, while the remark "we can't substitute for 'I' a description of a body" is indeed a grammatical rule, a convention whose violation and corollary bit of nonsense anyone could recognize as such[40], the "proposition that that which has pains or sees or thinks is of a mental nature" is a *confusion*, not a violation of any rule, though it has a (misunderstanding of a) grammatical rule as its "kernel".

Another way I want to put my point is as one about the importance of a threefold distinction, a distinction between types, or notions, of nonsense. Really, the importance of the distinction lies between the second two types I'll set out presently, but I'm including the first type here so as to anticipate something that will come up later. (There's probably room for further distinctions besides just these three, but I'm putting up the minimum it will take to make the points I want to make in this chapter.) While mine is a distinction between three notions of nonsense, none of these is a *mistaken* notion of nonsense (again, anticipating something to come up later). Though, as I see things, a particularly important mistake does arise from conflating the second two. The three kinds of nonsense, or three kinds of failure, are these: type (0) nonsense like "ab sur ah" whose failure to make sense nobody would be able seriously to *diagnose*; type (1) nonsense like "I remember going to bed tomorrow" whose failure anybody (with the relevant English competences) could

[40] Anyone sharing Wittgenstein's view of our actual grammar, that is. I'm not trying to argue that what Wittgenstein counted among the rules of grammar, the violable ones, would be universally agreed to be so. I don't suppose it would matter much to him, though, whether these or different rules were adopted. Cf. (Wittgenstein, 1969, p. 55): "The matter is different, of course, if we give the phrase 'unconscious pain' sense by fixing experiential criteria for the case in which a man has pain and doesn't know it."

rather easily diagnose as a matter of breaking certain grammatical rules, rules like "every rod has a length"; and type (2) nonsense like "The number two is a mysterious immaterial object" whose failure Wittgenstein found it terribly elusive and important properly to diagnose (which diagnosis seemed one and the same with the diagnosis of one's *temptation* to say it), and whose failure is not a matter of violating our grammar, but of being confused by it.

The real interest in this threefold distinction, as I said, lies in the distinction between types (1) and (2). (1) is the sort of nonsense that, as I read Wittgenstein, is indeed a matter of flouting well-known grammatical rules, the sort of rules that can be mistaken for descriptions of necessary features of reality and that can be innocuously translated back into the 'formal mode'; (2), on the other hand, is a far more interesting sort of nonsense. (2) is the sort of nonsense in action when we, in various ways, neglect the four Wittgensteinian unargued insights set out above; that is, when we're misled by our forms of expression, or when we fail to recognize that differences in criteria make for grammatical differences, or when we've lost sight of the point of some language-game. Producing this kind of nonsense isn't, as I said, *violating* any rule of grammar, though our impulse to produce it indeed arises, according to my reading of Wittgenstein, in part from a misunderstanding about what rules of grammar *are* (that is, from a neglect of the middle two of the four insights laid out above). So, it's not by *breaking* rules that, on my view, we produce type (2) nonsense, but it's by *misunderstanding* them. It's not that we *misuse* our language, it's that we *misunderstand* it.

Corresponding to my types (1) and (2), then, are two distinct notions of 'describing our grammar'. We could say, for instance, (1*) that it's a rule of grammar that one can sometimes doubt whether others are in pain, but that one can't doubt it of oneself. But then we can also say, for instance, (2*) that the point of this rule of grammar is tragically parodied in the inner/outer picture of mind and behavior. Keeping (1) and (2) distinct keeps (1*) and (2*) distinct as well, and this changes greatly the spirit in which much of Wittgenstein's later philosophy will be taken; that is, if (1) and (2) had been hitherto run together.

And so we can see that when Wittgenstein says we've been misled by our grammar when we're tempted to say such-and-such, he doesn't mean that such-and-such is a combination of words that gets excluded by conventional, violable rules of grammar. We can see that Wittgenstein's philosophy isn't a matter of pointing out where rules of grammar have been violated (like they would be violated, e.g., if someone were to say "The number three is heavy"); it's a matter of getting us to see our rules in such a way that certain pictures, questions, and theses will come to seem empty. That they are empty will not itself be a grammatical rule, something that can be broken, but saying that they're empty will be, in sense (2*), to describe our grammar; it will be something Wittgenstein thinks we will acquiesce in once we see the grammar of our language (however open-ended) in the right light. The steps to seeing our grammar in the right light, again, are embodied in the four insights set out above: quickly, (i) that noun phrases shouldn't be taken uniformly to pick out (however mysterious) entities, states and processes, (ii) that *certain* metaphysical statements are conventional rules of grammar in disguise, (iii) that our grammar is

shaped not only by these rules but also by explanations and verifications we associate with our sentences, and (iv) that our language-games have points that can be lost in various abnormal situations. And, on my view, these four insights are, just like Wittgenstein's claims that their neglect results in confusion, not themselves rules of grammar. Anyway, not the kind of grammatical rules that are the better expression of certain necessity claims. It'd be best not to call them "rules" of grammar at all, really; though, just the same, they are descriptive, in sense (2*), of our grammar.

(I want to note in passing that my distinction between types (1) and (2) isn't to be thought of as the Wittgensteinian distinction between obvious and unobvious nonsense.[41] Transitioning from unobvious type (2) nonsense to obvious type (2) nonsense importantly *won't* be a matter of making type (2) nonsense out to be type (1) nonsense. We don't find it, or make it, a rule of grammar that gaseous, immaterial entities, for example, are a confusion, though we do dissolve that confusion by attending to grammatical rules, attending to them *as* grammatical rules.)

I don't know of anywhere in Wittgenstein's writing that he explicitly makes this distinction between types (1) and (2), but I have to say it seems to me a highly important one, one that can be plausibly read into his work, and one that, indeed, should be read into his work.

With that, I'd like to leave off setting out my own reading of Wittgenstein as a non-theorist, turning to a pair of readings of Wittgenstein with which I want to contrast my own.

<u>A Non-Assertory and a Non-Argumentative Reading</u>

[41] Cf. (Wittgenstein, 2009a, §464).

I want to set out two readings of Wittgenstein that, as I discern things, both labor under a conflation of the types of nonsense (1) and (2); that is, they don't see things the way I'm suggesting, that Wittgenstein's urgings against such fantasies as private ostensive definition are *not* intended as articulations of, or reminders of, violable rules of grammar.[42] They both, in their ways, accept uncritically, if I may say, that Wittgenstein is in the business of, first and foremost, articulating, or reminding us of, rules of grammar, whose violation is the spring of all nonsense, and then they both try to let Wittgenstein off the grammatical-dogma hook, so to speak, in some way or other. The hook being, of course, that "The picture of private ostensive definition is empty" doesn't seem anything at all like the grammatical triviality that runs "'3 is heavy' is nonsense", and so putting the former as though it were of the latter sort would seem, indeed, to commit one to a kind of, what should be clearly un-Wittgensteinian, 'grammatical dogmatism'.[43] I'll try to show that these other ways are unpromising and unnecessary.

The first way I want to look at of letting Wittgenstein of the grammatical-dogma hook is due to Oskari Kuusela; I think of his as a non-assertory reading. Now,

---

[42] Both these readings are very much in line with the so-called New Wittgensteinian movement (principally, but not restricted to, a movement of interpretion of the *early* Wittgenstein; cf. (Crary, 2000)), and they're very much inspired by the later interpretation of the later Wittgenstein set out by Gordon Baker; cf. (Baker, 2004). The definitive feature of their shared approach is that Wittgenstein's later work is to be seen as a matter of therapy *as opposed to* a matter of the "tabulation of grammar and the interrelations of concepts". (Kahane, Kanterian, Kuusela, 2007, p. 6) For important 'orthodox' reactions to both the New Wittgensteinian interpreation of Wittgenstein's earlier work and to Baker's later interpretation of Wittgenstein's later work, see (Hacker, 2000) and (Hacker, 2007).

[43] Cf. the following remark of Wittgenstein's, one of many that commits him to dealing in trivialities: "It is…nonsense to ask where the number 1 is. This comment may be trivial, like all the comments we shall make; but what is not trivial is seeing them all together." (Wittgenstein, 1979, p. 44) On my reading, what are also not trivial in the "comments we shall make" (no trivial matters of violable grammatical rule, that is), and what Wittgenstein forgot to mention here, are the comments he'll make about what one will be led to give up as fantasaical upon "seeing them all together" ("them" referring to his grammatical trivialities).

for me, Wittgenstein does make assertions, striking ones, and he even left some important assertions unasserted; but these, importantly, aren't simply articulations of violable grammatical rules. That the picture of private ostensive definition is empty is something it will take an appreciation of our grammar to see, no doubt, but, I can't stress it enough, that doesn't make "No private ostensive definition" itself a rule of grammar. As I understand Kuusela, though, he would take this as, indeed, a rule of grammar; but, given its admitted non-triviality, he would try to account for its place in Wittgenstein's work by insisting on its being nothing Wittgenstein would've *asserted* as a rule of grammar. How could this be? Well, it turns on his view of the role of grammatical rules in Wittgenstein's later thought.

Kuusela writes:

> It seems that the philosopher's statement has two elements: i) a rule for the use of language and ii) a statement that the actual use accords with or is governed by this rule. Importantly, the second element can be understood in more than one way. (Kuusela, 2005, p. 111)

When Kuusela says "the philosopher" he means to include Wittgenstein, and his understanding of Wittgenstein's use of grammatical rules turns on how he construes the second element of the "philosopher's statement". For Kuusela, there's one good way and there are two bad ways of construing that second element; that is, one good and two bad ways of understanding the role of Wittgenstein's appeals to rules of grammar. The two bad ways are, first, thinking that the rules of grammar articulated by Wittgenstein are meant as descriptions of the actual grammar of our actual language, and second, thinking that his rules of grammar are meant as descriptions of what rules we would *have* to follow if we were to stay faithful to our actual rules. The first is supposed to be bad (or at least un-Wittgensteinian) because it's supposed

to involve the philosopher in *hypotheses* about empirical affairs (ibid.), which Wittgenstein is supposed to have repudiated: "There must not be anything hypothetical in our considerations." (Wittgenstein, 2009a, §109) The second way is supposed to be bad because:

> To say that a word must be used in such and such a way in order to express a particular concept is still to make a statement about how that word *must* be used…Accordingly, the danger that in stating his rule the philosopher overlooks some other ways of using the word as an expression of the concept in question seems as great as ever. (Kuusela, 2005, p. 111)

Aside from the dubious, and completely unexplained, notion of "other ways of using the word as an expression of the concept in question", it seems obvious to me that the second way couldn't really be anything more than the first way. Compare: "Johnny is at the door now" and "It will have to be Johnny at the door later, if the person at the door later is to be the same person as the person there now."

So, about what's supposed to be wrong with the first way of understanding Kuusela's element ii), I just don't see that Wittgenstein's claim that nothing "hypothetical" can enter "our" considerations counts out our giving descriptions of our actual language. Consider the following remark from Wittgenstein's *Ambrose Lectures*:

> I only describe the actual use of a word if this is necessary to remove some trouble we want to get rid of. Sometimes I describe its use if you have forgotten it. Sometimes I have to lay down new rules because new rules are less liable to produce confusion or because we have perhaps not thought of looking at the language we have in this light. Thus we may make use of the facts of natural history and describe the actual use of a word; or I may make up a new game for the word which departs from its actual use, in order to remind you of its use in our own language.[44] (Wittgenstein, 1979, p. 97)

[44] This passage should mean something to Kuusela, at least, who himself seems to put much stock in the Ambrose Lectures.

Whatever the claim that nothing hypothetical will enter in is supposed to be doing, I don't see that it can be used to take Wittgenstein out of the business of describing actual usage.

So what about the alleged *good* way, Kuusela's third way, of construing his element ii)? For Kuusela, the crucial point about the third way is that, according to it, the rules of grammar Wittgenstein gives us are only meant as *objects of comparison*. And for Kuusela, the dogmatic mistake we are in danger of making is that of taking Wittgenstein's objects of comparison for something more than they are. His key Wittgensteinian remark is the following:

> For we can avoid unfairness or vacuity in our assertions only by presenting the model as what it is, as an object of comparison -- as a sort of yardstick; not as a preconception to which reality *must* correspond. (The dogmatism into which we fall so easily in doing philosophy.) (Wittgenstein, 2009a, §131)

Now, it's undeniable that what Wittgenstein had in mind here, the "objects of comparison", were his "clear and simple language-games" (Wittgenstein, 2009a, §130), those he used to "disperse the fog" of the "general concept of the meaning of a word". (Wittgenstein, 2009a, §5) Kuusela's radical suggestion is that a whole lot else of what would pass for Wittgenstein's descriptions of our grammar are really only meant to be held up as such objects of comparison. The principal 'rule' Kuusela seems to have in mind is the slogan "meaning is use".[45] For Kuusela, as I read him, "meaning is use" is indeed put forward as a rule of grammar (like, e.g. "There's no castling in checkers"); and that's element i) of a Wittgensteinian statement that meaning is use. But the rule in that statement is only put forward as an object of

[45] Cf. (Kuusela, 2006).

comparison, something to which reality may or may not correspond; and that's element ii) of the Wittgensteinian statement. It is crucially, for Kuusela, not asserted by Wittgenstein to be the actual rule we play by.

Beyond Wittgenstein's §131 warning not to take his objects of comparison for anything more, Kuusela's main textual evidence for this reading of this particular Wittgensteinian dictum ("meaning is use") comes from the *Ambrose Lectures*:

> I shall not proceed by enumerating different meanings of the words "understanding", "meaning", etc., but instead shall draw ten or twelve pictures that are similar in some ways to the actual **use** of these words…To begin with, I have suggested substituting for "meaning of a word", "use of a word", because *use of a word* comprises a large part of what is meant by "the meaning of a word".[46] (Wittgenstein, 1979, p. 48)

As Kuusela reads this, Wittgenstein's suggestion to substitute "use of a word" for "meaning of a word" is itself meant as one of the "ten or twelve pictures" that he'll draw around the words "meaning" and "understanding". I think Kuusela is mistaken here. My boldfaced "use" in the quoted passage is supposed to bring out what I think renders his construal implausible. It's clear, if we're careful, that the point of the "ten or twelve pictures" is that they're supposed to capture the *uses* of the words "meaning" and "understanding", so that it's *presupposed* in this that it's *use* we're interested in capturing. Wittgenstein's interest in use is here patently *anterior* to his drawing any given picture (object of comparison) to shed light on our words "meaning" and "understanding". Reading the passage this way, now, we can see "to begin with" not as signaling Wittgenstein's beginning to draw an object of comparison, but as signaling his taking a step back, so to speak, and explaining why it was the *use* of the word "meaning" (instead of the *meaning* of the word "meaning")

[46] My boldface.

that he was after with his promised objects of comparison. This seems to me to make perfect sense of the boldfaced occurrence of "use", while I don't see that Kuusela has any way of doing so. The substitution of "use" for "meaning", then, is, I think, highly implausibly thought of as setting out an object of comparison.[47]

Furthermore, on my view, it's simply out of the question that "meaning is use" is a violable rule of grammar like "no castling in checkers" or "every rod has a length". Rather it amounts to a recommendation precisely to take up the four unargued Wittgensteinian insights laid out above as (therapeutic) tools used to "destroy" "cloud-castles", whose to-be-destroyedness is *itself* no violable grammatical rule.

But besides "meaning is use", what if we try out Kuusela's object of comparison strategy on another would-be rule of grammar Wittgenstein is supposed to have set forth (that is, given a conflation of my types (1) and (2))? Let's pretend that "Private ostension is a fantasy" is supposed to be a rule of grammar that can be violated, and that Wittgenstein indeed set it forth as such. And let's ask whether it's at all plausible that this was set forth as indeed a rule, though only as a rule *qua* object of comparison. Well, I have to say I find this highly implausible. They are the *clear and simple language-games* that are the objects of comparison, but I don't see any way that Wittgenstein's contention that private ostension is a fantasy can possibly be made out to be definitive of some such clear and simple language-game, language-games that are instructive only in their similarities and dissimilarities with our actual linguistic practices. Then again, I don't see any way it's supposed to contain a *rule* in

[47] I have to admit that I don't see anything in the pages immediately following this remark of Wittgenstein's that I would readily count as "drawing ten or twelve" pictures, but I suppose that can be explained by the fact that it came up in the course of delivering lectures.

the first place, certainly no rule in the sense that "'3 is heavy' is nonsense" is a rule of our grammar.

While I share Kuusela's aversion to any dogmatic pronouncement that to dream of private ostensive definition is to break our grammatical rules, that's only because I think that the rules of our language can be, on the one hand, broken, and on the other hand, misunderstood, and I think that the dream of private ostensive definition belongs squarely on the other hand. So, if I'm right about the implausibility of Kuusela's object of comparison strategy, and if the goal is to explain away any appearance in Wittgenstein that he was simply laying down the grammatical law, so to speak, with his rejection of such fantasies as that of private ostensive definition (and I do believe such is Kuusela's and my shared goal), then I think my breaking-rules/misunderstanding-rules strategy has more going for it than Kuusela's strategy has going for it.

The next reading of Wittgenstein I want to contrast with my own, and one I want to spend a little more time on than I have with Kuusela's, is Stephen Mulhall's; I think of his as a non-argumentative reading. On my reading of Mulhall's reading, it, like Kuusela's, is framed as a response to the problem of Wittgenstein's not-so-trivial seeming claims about our rules of grammar, such as that a private language is a fantasy, is nonsensical. The specific question I see Mulhall dealing with is the question by what authority Wittgenstein thinks he can make such radical claims. The traditional philosopher who would insist that, for instance, one is directly acquainted with her own mental goings on would *also* insist that it makes sense to say that. So, what about the philosopher who *denies* that it makes sense to say that? As I

understand Mulhall, his principal concern, as an expositor of Wittgenstein, is with whether this second kind of philosopher can *prove* to the traditional philosopher that his claim doesn't, in point of fact, make sense. After all, if anyone would seem to, Wittgenstein would seem to deny that such claims make sense. So, does Wittgenstein have a proof? And what would it be to 'prove' of some philosophical question, or some try at answering it, that it was nonsense?

Mulhall defines his understanding of the Wittgensteinian strategy with these difficulties in opposition to a certain, in ways obviously un-Wittgensteinian, understanding of the Wittgensteinian strategy. He sees his opponents, rival readers of Wittgenstein, as working with a *substantial* conception of nonsense, while he sees himself, or his Wittgenstein, as working with an *insubstantial* conception.[48] Our question, the question Mulhall addresses, comes to this: How do we uphold, on the one hand, Wittgenstein's talk of the nonsensicality of various traditional philosophical theories, while upholding, on the other hand, his promise to abstain entirely from philosophical theorizing himself? And, as Mulhall sees it, whether we can uphold both turns on which of these two views we take of nonsense. If we take a substantial view of nonsense, then we can't; if we take an insubstantial view, then we can.

As Mulhall sets it out, we adopt a substantial view of nonsense when we adopt the following: (a) some nonsense is distinct from mere gibberish, because its individually intelligible signs are combined in illegitimate ways, thereby specifying a

[48] The terms derive from James Conant's distinction between "substantial" nonsense and "austere" nonsense; cf. (Conant, 2002, p. 400). For an important 'orthodox' reaction to Conant's (and others') reliance on this distinction (which reliance involves a particular story about Wittgenstein's relationship to Carnap), see (Hacker, 2003).

thought we can't think; (b) a philosophical theory of the conditions of sense is required to demonstrate nonsensicality, "as if our everyday abilities to distinguish sense from nonsense require at the very least a philosophical grounding or foundation". (Mulhall, 2007, p. 9) The insubstantial view amounts, more or less, to the denial of these. It should be clear that Mulhall, with his substantial/insubstantial distinction, is aiming at precisely that difference between trying to *drive people out* of paradise, and trying to get them to *give up* a would-be paradise as *illusory*. It's not enough, as he sees it, just to call traditional philosophical theories nonsensical, as opposed to outright false, in order to land on the 'giving up' side, as opposed to the 'driving out' side; one must also have the correct, the insubstantial, view of nonsense itself.

One thing worth pointing out is how closely (a) and (b) fit together for Mulhall. For him, indeed, they're two sides of the same dubious coin. That's because, for Mulhall, it's only someone working with a theory of the conditions of sense who would respond to a traditional philosophical claim with something like, "That's nonsense, and I can tell you why". And, clearly, we shouldn't need to appeal to any theory of the conditions of sense in the way of pointing out the nonsensicality of mere gibberish; we don't have to tell anyone why gibberish is gibberish; there's no *diagnosing* such nonsense. So, the very idea of an appeal to theory already sets the problematic utterance apart from mere gibberish. And, for Mulhall, the difference between, on the one hand, mere gibberish, whose nonsensicality couldn't require explaining, and, on the other hand, nonsense whose nonsensicality can be, and needs to be, explained by a theory, is the difference between nonsense whose parts are

themselves nonsensical, and nonsense whose parts are themselves, indeed, quite possessed of sense. On this kind of view, being able to *say why* this or that philosophical claim is nonsensical amounts to just the same thing as being able to say why *these* meanings can't fit together *like that*.

> After all, one might think, how can we know that the philosopher cannot say or think what he wants to say or think, without knowing what exactly it is that he wants to say or think? (Mulhall, 2007, p. 9)

For Mulhall, if we ever wanted to specify what exactly a (traditional) philosopher was trying to say or think (with his nonsensical utterance) we should have to specify a thought one couldn't think, "an identifiable place in the region that lies beyond the limits of sense". (Mulhall, 2007, p. 8) But this would, obviously, run contrary to the whole Wittgensteinian outlook in the later work.[49] Mulhall recalls two important remarks in the *Investigations* that rather unequivocally (to those familiar with Wittgenstein, anyway) close the door on any such conception of the unthinkable at work in Wittgenstein:

> The great difficulty here is not to represent the matter as if there were something one *couldn't* do. (Wittgenstein, 2009a, §374)

> When a sentence is called senseless, it is not as it were its sense that is senseless. But a combination of words is being excluded from the language, withdrawn from circulation. (Wittgenstein, 2009a, §500)

So much for thoughts we can't think in Wittgenstein. And, for Mulhall's Wittgenstein, so much for theories, proofs, or explanations of the nonsensicality of philosophers' nonsense. Mulhall's denial that there can be, in any sense, an *argument* in back of some charge of nonsensicality comes across particularly strongly when he

[49] I'm not going to try to tackle the question of continuity in Wittgenstein on this point.

addresses the question in what our ability to distinguish sense from nonsense in fact consist:

> It is simply the capacity to recognize when a sign has not been given a determinate meaning--even when it appears that is has been…[T]he practical know-how we are thereby drawing upon is such that any attempt to state it in words will produce an utterance that anyone who possesses that know-how must recognize as itself nonsensical. (Mulhall, 2007, p. 13)

It's interesting that Mulhall doesn't go so far as to say that even the recognition of nonsense is something that can't be put into words, that any attempt even to *say* that this or that is nonsensical is bound to misfire. That would've been the very last stop, I suppose, before denying any real, as opposed to apparent, charge of nonsense at work in Wittgenstein's later thought. For Mulhall's insubstantialist, there *is* nonsense, we *do* recognize it, we *can* say so, we simply *can't* say how or why. There *is* a how, that is, a know-how; but this, for Mulhall's insubstantialist, cannot be put into words.

But why doesn't Mulhall take the further step and embrace an inexpressible recognition of nonsense, or even do away with that recognition altogether? After all, handling the problem about understanding Wittgenstein as a non-theorist means handling the problem about *all* his apparently non-trivial trivialities. But it isn't just the apparent *arguments* for nonsensicality in Wittgenstein that would threaten his commitment to triviality; the very *accusations* of senselessness themselves (should they be more than apparent in Wittgenstein) would just as well threaten that commitment.

Mulhall, however, doesn't want to go this far and doesn't see that he needs to. For Mulhall, any real charge of nonsense in Wittgenstein comes only from a humble

competence with English, German, French, or whatever; and so it really is trivial, that is, not to be contended with. He admits that sentences can *seem* to bear sense when they actually don't; but, on his view, nobody with "everyday capacities for linguistic expression and understanding" would actually be taken in by such seemings; that is, so long as her everyday understanding were "deployed" (Mulhall, 2007, p. 6) in the unusual philosophical context that gave rise to the illusion of sense. Clearly, then, Mulhall isn't adopting any rigid notion of triviality that comes simply to "won't be called 'false'". For, some people will, indeed, fail to deploy their everyday understanding in the problematic philosophical context, and that's why they'll call the trivially nonsensical thing "sensical", and that's why they'll call the claim that it's nonsensical "false". I think this is a step in the right direction for Mulhall, or at least a stop in the wrong direction.

Mulhall takes it as "obvious" that certain "mainstream readings" of the *PI* take the substantial view of nonsense. (Mulhall, 2007, p. 10) My misgivings with Mulhall's view begin with my objection to his (implied) claim that *any* mainstream reading of the *PI* is committed to unthinkable thoughts. In fact, I don't know of one that is. Of course, I don't deny that many mainstream readers of Wittgenstein are committed to the possibility of specifying conditions of sense in the way of explaining the nonsensicality of some particular nonsensical sentence; I only deny that they are thereby committed to unthinkable thoughts.

For Mulhall, the substantialist's commitment to unthinkable thoughts is supposed to come by way of his commitment to *determinate* nonsense, which is supposed to come by way of his commitment to explanations of the senselessness of

senseless strings by appeals to *conditions of sense*. The conditions of sense appealed to by Mulhall's substantialist are otherwise known as *rules of grammar*, and his substantialist casts the recognition of some bit of nonsense as a matter of the recognition of the violation of some grammatical rule or other. Of course, there's going to be more than just one rule of grammar for Mulhall's substantialist, and there will be at least as many ways to produce nonsense as there are rules of grammar to violate (and ways to violate them). So, the substantialist has at his means (or, for Mulhall, on his hands) a rather natural procedure for sorting strings of nonsense apart: according to which grammatical rules they violate (and in which ways).

Now, I suppose I agree with Mulhall up to this point: a conception of nonsense as a matter of the violation of specifiable grammatical rules does, indeed, bring along a commitment to determinate nonsense, at least *in this sense of* "*determinate*". That is, I can agree with Mulhall that "[t]he notion of substantial nonsense is that of pseudo-propositions that are unintelligible, but determinately so". (Mulhall, 2007, p. 8) I simply don't see, however, that we need to follow him in his very next step:

> [These pseudo-propositions] therefore seem to specify a thought that we cannot think--an identifiable place in the region that lies beyond the limits of sense, something specific that exceeds our mental grasp. (ibid.)

To see that we needn't follow Mulhall here, it helps to see that his insubstantialist too has at her means a procedure for sorting nonsense. Mulhall's insubstantialist, remember, still recognized certain strings as nonsensical; only the know-how deployed in her recognition was supposed to be such as couldn't be put

into words. I'll let one obvious criticism slide[50], and I'll simply point out that, on any reasonable construal, it won't be some kind of vague, general linguistic competence whose deployment will have the insubstantialist recognizing any particular string of nonsense as such. I can only presume it will be competence with *the very signs* figuring in a particular nonsensical string whose particular deployment will be singularly relevant to her recognizing the string's nonsensicality. So, while the substantialist says that, given the meanings of (grammatical rules for) such-and-such particular words, such-and-such combinations of them are nonsensical, the insubstantialist will say that, given only her competence with (*not* propositional knowledge of grammatical rules for) such-and-such particular words, she can tell that such-and-such combinations of them are nonsensical. This, it seems to me, is certainly a straightforward sense in which Mulhall's insubstantialist operates with a notion of determinate nonsense. Bits of nonsense won't be individuated for her according to (perfectly statable) grammatical rules violated; but they can be individuated according to linguistic competences deployed in the course of (or instant of) recognizing nonsensical combinations as such. And while she will insist that all nonsense is gibberish, she'll have to admit that her recognition of the nonsensicality of "ab sur ah", for instance, isn't a matter of her deploying any competence with any English word in particular, but that in more interesting cases such particular competences of hers are indeed pressed into service.

[50] Incidentally, Mulhall's defense against the obvious point here is that his kind of know-how isn't itself mysteriously ineffable, because there isn't any ineffable *truth* about *how* things are that is its "object or content", and because our know-how is indeed made "manifest" in our abilities, among others, to tell sense from nonsense. (Mulhall, 2007, p. 7)

So I ask, Does this kind of determinacy of nonsense commit one to any "place in the region that lies beyond the limits of sense, something specific that exceeds our mental grasp"? I should think not. I should think one could quite happily individuate nonsense according to competences without coming anywhere near taking each determinate bit of nonsense to be getting at some unthinkable thought. But then why should a commitment to the statability of grammatical rules and their violations generate such an untoward commitment as that to unthinkable thoughts? Because such a commitment brings a commitment to the determinacy of nonsense? But it can't be the determinacy of nonsense *per se* that does one in; after all, the insubstantialist, like it or not, can tell strings of nonsense apart just as well as the substantialist. Is there something in the notion of stating a grammatical rule, or pointing out that one has been violated, that renders a correlated notion of determinate nonsense particularly outrageous? Mulhall certainly doesn't give any argument to that effect. In fact, I don't see that there's one to be given. That is to say, I don't see that anything in a conception of nonsense whereby particular nonsensical utterances are explained as violations of particular grammatical rules brings along, all by itself, *that* kind of determinate nonsense.

Nonsense-strings that get their determinacy (distinctness from other nonsense-strings) from their associations with determinate places in an unthinkable region beyond the limits of sense are indeed to be avoided at all costs. I just don't see that the way to avoid them is to opt for insubstantialism over substantialism. Whatever other merits or demerits it may merit, I don't see that Mulhall's unstatable know-how, a know-how apparently without "object or content" though duly sensitive to the

presence of these, as opposed to those, word-sized sign-strings in the context of any particular nonsensical sentence-sized sign-string, is the desperately needed way out of such metaphysical clap-trap as "regions beyond the limits of sense". Mulhall is barking up the wrong tree.

I say that, because I do think Mulhall is on to something. I share his aversion to that conception of our being captivated by particularly captivating bits of nonsense (our insistence on the reality (and ubiquity) of super-private ostension, for instance) as a conception of our being in violation of some grammatical rule or rules. And I share Mulhall's aversion to that conception of Wittgenstein's way with such captivation as a conception of his setting out an assortment of grammatical rules of which we remain in violation so long as we remain captivated. But if it's not that such a conception of nonsense, and of Wittgenstein's way with it, leaves room for our being able to tell apart various captivating nonsensicalities from each other and from less interesting nonsensicalities, then what is it about such a conception that unsettles so many readers of, and sympathizers with, Wittgenstein? The answer I want to give, of course, is that such a conception wrong-headedly conflates types of nonsense (1) and (2). The crucial point will be that, if we are mindful to keep (1) and (2) separate, claims about the second kind of nonsense (e.g. that the picture of platonic objects is empty) aren't in any way the *grammatical consequences of* claims about the first kind of nonsense (e.g. that "3 is heavy" is nonsense). On my view, Wittgenstein's 'showing why' a type (2) nonsense is indeed nonsensical only amounts to using the four insights from the last section to show someone, among other things, which analogies they had been led by in coming to express their puzzlement or its 'solution',

and in such a way that they'd then *give up* at once the puzzle and its solution as empty. Given this, I'd like to take a moment to see how Mulhall's distinction looks in light of my own.

Unless I'm quite mistaken, the spirit of Mulhall's distinction between substantial and insubstantial conceptions of nonsense can captured thus: if a traditional philosopher persists in his absurd claim, a substantialist will respond "You haven't appreciated the force of my argument", whereas an insubstantialist will respond, "You haven't deployed your everyday understanding of your own words." But let's suppose first that substantialist and insubstantialist both are conflating my (1) and (2). That is, let's suppose they both think that forsaking, for example, platonic objects as nonsensical is of the same sort as forsaking "3 is heavy" as nonsensical. In that case, I don't think "You haven't deployed your everyday understanding" is any more reasonable a thing to say than "You haven't appreciated my argument". I don't think any amount of forceful argument, nor any amount of everyday understanding, can have one recognizing that platonic objects are nonsensical just the way heavy numbers are.

But now let's suppose that substantialist and insubstantialist both are working with my distinction; that is, let's suppose they're keeping (1) and (2) distinct. In that case, I don't see that the insubstantialist has any substantial complaint against the substantialist's talk of *deriving* a claim about the (2) kind of nonsense from a claim about the (1) kind. If you recognize that claims about the empty picture kind of nonsense aren't intended to be the grammatical consequences of any amount of grammatical reminders, then you'll have to admit that assembling grammatical

reminders isn't anything in the way of constructing, *per impossibile*, an *argument* whose logical conclusion is that the picture in question is empty. If you call what you've set up "an argument", but you keep (1) and (2) distinct, then yours isn't any argument in any objectionable sense. If it's said that the words "argument", "demonstration", "proof", etc., *must* betray a misunderstanding of the way (2) nonsense relates to (1) nonsense, then I believe the rejoinder is simply that the word "philosophy" itself is as at least as out of place in Wittgenstein's thought as any of these.

Such are my misgivings about the productiveness of Mulhall's substantial/insubstantial distinction. And such are the considerations in favor, again, of my own distinction as a way of letting Wittgenstein off the grammatical-dogma hook; or, rather, of explaining away any appearance of his being so hung up. I turn now to consider Wittgenstein's place as an expressivist theorist in light of everything I've been doing up until now in this chapter.

Wittgenstein' (Lack of an) Expressivist Theory

The time has come to ask how Wittgenstein handles my (Q1) through (Q4). I believe I've laid the groundwork so that each of the responses I'll be attributing to him will be seen to harmonize with his overall, non-dogmatic approach to philosophizing.

We'll start with the first question, whether the utterances most nearly approximating non-linguistic, behavioral expressions of mentality are rightly thought of as truth-apt. Surprisingly perhaps, Wittgenstein never addresses this question exactly, though he addresses a very closely related one. The question he does address

is whether such utterances are rightly thought of as *descriptions*. His response is characteristically therapeutic:

> If someone wants to call the words the "description" of the thought instead of the "expression" of the thought, let him ask himself how anyone learns to describe a table and how he learns to describe his own thoughts. And that only means: let him look and see how one judges the description of a table as right or wrong, and how the description of thoughts: so let him keep in view these language-games in all their situations. (Wittgenstein, 1980a, §572)

Admittedly, what's in question here is thought in particular, and not the whole of the avowably mental. But the lesson, as I see it, generalizes. Now, there are two ways of reading this passage corresponding to whether we do or we don't make use of my distinction between the types (1) and (2) of nonsense in Wittgenstein's work. It could be that Wittgenstein is asking us to remember the relevant language-games, because there we will find the rule that "description" doesn't apply to avowals about what one thinks. But I don't believe that's right. I believe we should see Wittgenstein as directing us to grammatical facts about our language-games (here, facts about how one judges the correctness of the various 'descriptions') so that we will be struck by the emptiness and the highly misleading character of using the word "description" in both connections. But not only is Wittgenstein trying to get us *not to want* to use the empty word "description" in both connections, he also seems to think that we *will want* to use the word "expression" in connection with mind-descriptions once the relevant grammatical features of the language-games are appreciated, or at least that we should prefer it to "description".

Now, in order to see how Wittgenstein's treatment of the question about "description" connects up with a Wittgensteinian treatment of the neighboring one

about “true or false” it helps to set up two different notions of description. You might call them the propositional (maybe, locutionary) notion and the pragmatic (maybe, illocutionary) notion. The propositional notion of a description is simply that of setting truth-conditions; the pragmatic notion I have in mind is rather like that of a *report*. Suppose you were to ask someone you’d never met to describe his room, and then suppose you asked him to describe your room. He’d probably say he couldn’t, because he’d never seen it. But this could be understood two ways, corresponding to the propositional and the pragmatic notions of description. On the one hand, he might be telling you he’s very likely to *mis*describe your room, never having seen it (or heard anything about it, etc.). On the other hand, he might be complaining that nothing he could say would even *count as* a description of your room, “describing” in that case amounting to something like “reporting on”.

Now, if we take Wittgenstein’s strategy with the question about “description” in the above quote as engaging with the pragmatic notion of description only, then his way with the question doesn’t quite reach to my own original question (1) about truth-aptness. That is, if Wittgenstein’s only target in that quote is an insensitivity to the diverse natures of the grammars of room-description and mind-description as embodied in the uncritical collecting together of both under the *pragmatic* notion of description, then the question remains whether these room and mind ‘descriptions’ aren’t both quite happily, that is non-misleadingly, collected under the *propositional* notion of description. Someone could admit that these mind-descriptions are more like expressions than *reports* while seeing nothing problematic about their setting truth-conditions, just like room-descriptions do, and having truth-makers, and all the

rest. Things are, obviously, quite different if we take Wittgenstein to be engaging with the propositional (truth-conditional) notion of description. If we read him this way, then it seems he's trying to tell us that reflection on the grammars of mind-description and room-description leads us to give up as empty any insistence that both kinds of description are alike in setting truth-conditions.

So, it could be that, when it comes to mind-descriptions, Wittgenstein is telling us that "expression" is better than "report", or it could be that he's telling us that "expression" is better than "truth-evaluable-content-bearer". ('Better', now, in the sense of being, what he thinks, will be the preferred choice of anyone duly sensitive to the *points* of the grammars of the two language-games; not better in the sense simply of correctly laying out the relevant, violable grammatical rules.) As I read him, he's telling us both. As I read him, the very fact that "expression" is better than "report" is just what goes to show why "expression" is better than "truth-evaluable-content-bearer". In fact, these two seem two aspects of the same point. For Wittgenstein, as I read him, insofar as collecting mind and room-descriptions alike under "bearers of truth-evaluable content" brings along a picture of mental truth-makers (states of affairs, objects and relations) right along side physical truth-makers, it is precisely the picture of inner-access brought along by collecting them both under "reports" that is to blame. That is, I think that Wittgenstein's expressivist treatment of (at least certain) privileged utterances is meant as much as a treatment of the *ontology* of the mental as a treatment of the *epistemology* of the mental. While this will seem an undergraduate mistake to the post-Kripkean philosophical masses, it seems clear that Wittgenstein's fight against the inner/outer picture of mind and

behavior is at once an ontological and an epistemological affair. So an appreciation of Wittgenstein's way with mind-descriptions *qua* reportive illocutionary acts does indeed lead to an appreciation of his way with mind-descriptions *qua* setters of truth-conditions.[51] Because of this, I think the quoted passage reaches all the way to the question about whether avowals are of the sort to be true and false. If thinking of them as truth-apt suggests the setting of truth-conditions, then it is to that extent misleading; that is, if we take truth-conditions to be something set just as well by room-descriptions as by mind-descriptions, and then we go on to be puzzled about just what kind of things, properties and processes could figure in the world's satisfying a truth-condition set by a mind-description. But the Wittgensteinian line here, on my reading, would certainly have to have run something like: "Go ahead and call them true and false, only don't be misled by that."

So, is this confronting (Q1) or not? Well, yes and no. I don't doubt Wittgenstein would have acknowledged that avowals belong to *language*, that they bear grammatical relations to each other and to non-avowals, that their competent use involves an acknowledgment of these facts, and furthermore that they are the sort of things that can be uttered sincerely and insincerely. But I don't think these considerations would have him giving an unqualified "yes" on the question whether avowals are truth-apt, not so long as he had any suspicion that empty pictures of mental objects, events, relations, states and processes would be ushered in thereby.

Now, it could seem this were an unnecessary hyper-evasiveness; Why wouldn't Wittgenstein simply unqualifiedly concede truth-aptitude and then fight like

[51] Later, when I discuss David Finkelstein's expressivism, it will be seen that an important difficulty arises for the expressivist who would keep entirely distinct the two questions of the appropriateness of "reportive" and "truth-apt".

mad against unwarranted inferences therefrom? For that matter, Why not read Wittgenstein precisely this way? But I think there is, from a Wittgensteinian perspective, good reason for being chary of that strategy. Namely, it presupposes that two people might agree about the truth (*a fortiori* the sense) of a sentence without agreeing about its logical consequences. More than that, though, it presupposes that two people might agree about the truth (sense) of a proposition, while what the one saw as a logical consequence of it, the other saw as nonsensical, confused and empty. Furthermore, for Wittgenstein, philosophy is a matter of therapy, that is, a matter of talking things out in order to dispel empty pictures, not a matter simply of getting the facts right (grammatical or otherwise) in order to disprove empty pictures. So I don't think any definitive Wittgensteinian statement about the truth-aptitude of avowals can be set out once for all, not so long as there would be so many who would grossly misunderstand that statement.

On the other hand, though, while I said Wittgenstein seemed to think that the preferred choice, after sufficiently taking in the relevant grammars, would be for "expression" over "description" as regards an avowal of one's own mental state, I don't think he'd see the preferred choice would necessarily be for "non-truth-apt" over "truth-apt" as regards a mental state avowal. I think that precisely because of the grammatical interconnectedness of avowals with other sentences, and because of the role of sincerity and insincerity in avowing, Wittgenstein's own preference, after taking stock of the relevant grammatical situations, would, in fact, very likely have been for "truth-apt". I think Malcolm had it more or less right when, in a very interesting exchange with Peter Hacker, he wrote:

> I know of no evidence that Wittgenstein ever held that the utterance "I have a toothache" is neither true nor false. What we should expect him to say is that the words "true" and "false" are used differently, have a different logical grammar, when applied to verbal "expressions," from what they have when applied to a statement such as "I have a decayed tooth." (Malcolm, 1978, p. 65)

One might hesitate to count this a *deflationary* approach to the truth-predicate as applied to verbal "expressions", since, one would think, a true deflationary approach won't have *different* 'logical grammars' for the truth-predicate depending on what it's applied to. On a truly deflationary approach, one might say, the truth-predicate always works the same way. But I think that if we take Malcolm's different logical grammars for "true" when applied to "I have toothache" and when applied to "I have a decayed tooth" to be, as it were, *transparent to* the different logical grammars of those two sentences themselves, then things look different. It does, indeed, begin to look as though we have something like a deflationary approach to truth and falsity in general, and that this helps to avoid certain immediate inferences one might be tempted to make just on the assumption that avowals are truth-apt.[52] So, all things considered, and for the sake of getting on with things, I suppose we can ascribe to Wittgenstein, on the question whether avowals are truth-apt, a qualified "yes". Thus, Wittgenstein diverges on this first point from the simple expressivist's way of

[52] See (Jacobsen, 1996) for an attribution of a deflationary view of truth to Wittgenstein on the basis of (Wittgenstein, 2009a, §136). It should be noted, of course, that the standard deflationist's points, *viz.* that (something like the) Tarskian disquotational schema exhausts what's to be said about truth and that truth is not a substantial property, aren't in precisely those forms insisted upon by Wittgenstein; though they seem to me rather Wittgensteinian indeed. See (Stoljar, 2007) for a review of the motivations for and against deflationary views of truth. See (Quine, 1986, ch. 1), for an early deflationary view of truth completely disconnected from, indeed contrary to, Wittgensteinian concerns about emphasizing differences in the uses of, for instance, various substantives by way of attention to sources of illumination (i), (ii) and (iii) from chapter two. See (Horwich, 1998, ch. 4) for a much more Wittgensteinian use of deflationism about truth in connection with setting out just such differences.

working out the expressivist insight; and thus, he obviates any of the familiar technical worries attending that way of working it out.

I turn now to (Q2), the question of epistemic status. This question is posed principally as one about knowledge: Do privileged (because expressive) utterances represent items of knowledge on behalf of their utterers? But very closely related to this are questions about justification and belief. Are privileged (because expressive) utterances justified? And are they even expressions of self-directed beliefs (when made sincerely), beliefs about one's own mental states? The go-to Wittgensteinian remarks in this connection would seem rather matter of (grammatical) fact:

> It can't be said of me at all (except perhaps as a joke) that I *know* I'm in pain. What is it supposed to mean -- except perhaps that I *am* in pain?...This much is true: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself. (Wittgenstein, 2009a, §246)

> Let's imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation…I speak, or write the sign down, and at the same time I concentrate my attention on the sensation -- and so, as it were, point to it inwardly…[B]y concentrating my attention…in this way I commit to memory the connection between the sign and the sensation…But in the present case, I have no criterion of correctness. One would like to say: whatever is going to seem correct to me is correct. And that only means that here we can't talk about 'correct'. (Wittgenstein, 2009a, §258)

> "When I say 'I am in pain', I am at any rate justified *before myself*." -- What does that mean? Does it mean: "If someone else could know what I am calling 'pain', he would admit that I was using the word correctly"?
>
> To use a word without a justification does not mean to use it wrongfully. (Wittgenstein, 2009a, §289)

> I can know what someone else is thinking, not what I am thinking.
>
> It is correct to say "I know what you are thinking", and wrong to say "I know what I am thinking".
>
> (A whole cloud of philosophy condenses into a drop of grammar.) (Wittgenstein, 2009b, §315)

The first thing to point out is that, while knowledge and justification, but not belief, come up explicitly here, the question about belief can be got at rather easily by pointing out that, if S believes that *p*, then it seems to S that *p* is correct. And if we "can't talk about 'correct'", then we can't talk about 'seeming correct' either; and then we can't talk about 'belief'. The second thing to point out is that, while sensations and thoughts specifically are the concern in these remarks, I should think it's perfectly within bounds to apply the sentiments expressed here to the whole range of privileged mental state self-ascriptions. The third thing to point out is that the Wittgensteinian treatment laid out in these remarks has struck very many philosophers as being quite trivially (that is, obviously) off the mark, and that that treatment would therefore seem an egregious renege on the promise to deal only in trivialities (trivial truths, that is). So we must ask, Is Wittgenstein really saying it's a grammatical fact, is it a type (1*) description he's giving, when he says we don't know of our own pains, that we can't be justified in making claims about them? A type (1*) description that we can't have true beliefs about which sensations we're having? A type (1*) description that it is "incorrect" to say "I know what I'm thinking"? My answer is, yes and no.

The reason I'd hesitate to call these type (1*) descriptions is that, as I read Wittgenstein, one first has to acquiesce in certain type (2*) descriptions of our grammar in order to see these striking pronouncements as indeed grammatical trivialities, as indeed type (1*) descriptions. And so, just in seeing these as type (1*) descriptions one is *eo ipso* taking on board certain Wittgensteinian type (2*) insights into the workings of our language. For that reason, I don't think it's right just to have

it that anyone caught saying something like "I know I'm in pain" can be straightforwardly corrected with the grammatical law, the way we *could* straightforwardly correct someone caught trying to tell us about the lengthless rod in his freezer. Now, these type (2*) insights, the insights that will make "It's incorrect to say 'I know I'm in pain'" out to be a trivial type (1*) matter of grammar, are themselves insights into the workings of certain type (1*) matters of grammar, which lower-level, as it were, matters of grammar are, I should think, indeed quite trivial.[53]

Let's begin with the last of the above quotes, wherein Wittgenstein said that it's correct to say we know what others think, but that it's incorrect to say we know what we ourselves think, the "drop of grammar" into which a "whole cloud of philosophy" is supposed to condense. This drop of grammar is, as I see things, what you might call a black belt version of something considerably more mundane. On my view, *PPF* §315 is nothing but a paradoxical sounding (to those not acquiescing in certain Wittgensteinian type (2*) insights) reworking of the same sort of thought as that at the end of *PI* §246, wherein Wittgenstein said, far less paradoxically, "This much is true: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself." (Wittgenstein, 2009a, §246) (It just seems obvious to me that one could in perfectly good conscience substitute "whether I'm thinking that *x*" for "whether I am in pain" in this grammatical description; and I wouldn't know where to begin arguing for or against that presumption.)

Now, in the 'material mode', Wittgenstein's claim here about the word "pain" could be put thus, "It's possible for others to doubt whether I'm in pain; but

[53] Though, again, only as trivial as the claim that "Every rod has a length" is trivial. Certainly not trivial in the sense that nobody ever in the history of humanity will be inclined to respond "False" to the humble grammatical offering.

impossible for me to doubt it". Similarly, one could, in the 'material mode', very well have Wittgenstein saying, "It's possible for others to doubt whether I'm thinking of *x*; but impossible for me to doubt it". This seems a far cry from *PPF* §315, but I think that's an illusion. As we've seen, the material mode is not Wittgenstein's preferred mode of expression, not given the second of his four type (2*) insights ('dogmas') laid out in the section before last, and this is going to help close the gap between *PPF* §315 and the material mode possibility and impossibility claims just set out.

Translation into the formal mode is all-important for understanding Wittgenstein on this point, because once these possibility and impossibility claims are put in that mode, the stage is set for another striking, vintage, Wittgensteinian type (2*) insight into the workings of our language, viz. that the response, "Of course I know…", is the "wrong method of brushing aside the question" (Wittgenstein, 1969, p. 30), "Are you sure that…?".

> It is similar when we ask, "Has this room a length?", and someone answers: "Of course it has". He might have answered, "Don't ask nonsense". On the other hand "The room has length" can be used as a grammatical statement. It then says that a sentence of the form "The room is --- feet long" makes sense.[54] (ibid.)

For Wittgenstein "Don't ask nonsense" is the better way to brush aside the question about whether a particular room has any length, and the same goes for questions about whether someone is quite sure that she's in pain, or that she's thinking *x*, etc.. For Wittgenstein, the question "Are you sure?" about such matters as one's own thoughts and pains is brushed aside wrongly with the response, "Surely I know such a simple thing as this", and the better thing to say is "The question which you asked me

[54] Cf. (Wittgenstein, 2009a, §§251, 2).

makes no sense". (ibid.) Though, on the other hand, just like with "The room has length", Wittgenstein does see that "Of course I know" can be used as a grammatical statement, viz. the statement that "It makes, in this case, no sense to talk of a doubt". (ibid.)

Now, as I read him, these two points, that "Of course I know" is the wrong way to brush aside the question about one's certainty and that "Of course I know" can be used as a grammatical statement, are really one and the same point. But it takes another type (2*) insight into the workings of our language to see this. What it takes is seeing that when, as a matter of type (1*) grammar, the applicability of a certain term in a certain case is senseless, then *so is its antithesis*. This, like the rest of Wittgenstein's type (2*) insights, isn't anything we should expect to see him arguing for. And I can't think of anything to say in favor of it myself, except that once it is indeed granted and fully appreciated that a given term's application in a given case is indeed *senseless*, then nothing seems more obvious to me than that its antithesis should be likewise without sense. If anyone should balk at this insight, I can only suspect that it's because he's still caught up in the material mode of expressing necessities, and against this I don't believe there's any argument to be given.

In any case, if we take it on board, then we can see better what Wittgenstein was after. If, as a matter of type (1*) grammar, doubt doesn't apply to certain claims we make, then neither does certainty. And, while Wittgenstein always wanted to translate "I must know whether…" into the grammatical exclusion of doubt, I think we can translate it just as well into the joint grammatical exclusions of doubt, mistake and insufficient justification. That's because, if doubting that *p* is simply entertaining

the possibility of one's being mistaken should she claim that *p*, or entertaining the possibility of one's having insufficient justification for claiming that *p*, then insofar as the possibility of doubting is excluded, so are the possibilities of mistake and insufficient justification.[55] But then, insofar as these exclusions are type (1*) *grammatical*, type (1*) grammatical exclusions of certainty, success and sufficient justification are brought in their wake; that is, given Wittgenstein's type (2*) insight about the sensical applications of terms and their antitheses.

So we can see that, for Wittgenstein, "It's wrong to say 'I know that I'm thinking of *x*'" traces back to "It's wrong to say 'I'm confident that I'm thinking of *x*'", which itself traces back to "It's wrong to say 'I'm doubtful that I'm thinking of *x*'". For Wittgenstein, "It's wrong to say 'I know what I am thinking'" just is the flipside of the type (1*) insight that doubt (and mistake and insufficient justification) is (are) excluded here. For these reasons, it's easy to see why, for instance, "I know I'm thinking of a white rabbit" should amount to nothing more than "I'm thinking of a white rabbit." If we can analyze, and I do believe we can, "I know I'm thinking of a white rabbit" into "I'm thinking of a white rabbit, and I know what I'm thinking of", then we can see that the second conjunct in the analysis is, at best, a disguised grammatical rule to the effect that doubt doesn't enter in. On the Wittgensteinian view of things, this would be rather as if someone said "I'm thinking of a white

[55] Strictly speaking, I suppose, it's the sensicality of *entertaining* the possibilities of mistake and insufficient justification that would be excluded grammatically by excluding doubt grammatically. So, someone might insist that it might happen that, while S couldn't *entertain* the possibility of S's being mistaken about such-and-such, still the *possibility* of S's being mistaken would be a real one. But, first, possibilities that someone, anyone, couldn't possibly entertain, possibilities that one couldn't possibly suppose true (not as a matter of psychological fact, now, but of grammatical rule), make scant sense in my estimation. And second, it doesn't matter, since I think Wittgenstein would've been happy to say explicitly that mistake and insufficient justification are themselves just as grammatically excluded from avowing as doubt itself is, regardless of the connections among them.

rabbit, and white is lighter than red". What is this supposed to mean except that I'm thinking of a white rabbit? (Well, I suppose it could be a statement about what I'm thinking of *plus* some instruction in English; I suppose as well that Wittgenstein would've been happy with that rendering.)[56]

This view of Wittgenstein's, that when, as a matter of type (1*) grammatical fact, doubtful, mistaken, and insufficiently justified beliefs are counted out, then, given his type (2*) grammatical insight, so are confident, veridical, and sufficiently justified beliefs counted out type (1*) grammatically, is probably what Wittgenstein is most famous for in the philosophy of mind (without my implicit distinction between types (1*) and (2*) of grammatical fact, clarification, etc.). And in the way of wrapping up my discussion on Wittgenstein's response to the second question, I want to bring out two more things about this view of his. First, while his view doesn't, as we see, diverge from the simple expressivist's response to my second question, on one way of looking at it, his motivations don't seem to have much of anything to do with the simple expressivist's motivations for that very same response. The simple expressivist's response, remember, was motivated by a wholesale assimilation of avowing with, for example, grimacing; and the inapplicability of epistemic (or doxastic) vocabulary thereto derived wholly from that assimilation. The simple expressivist, importantly, didn't work his way through a type (2*) insight about type (1) nonsensical terms and their antitheses. So, and this is something I'm going to return to at the end of this chapter, unless, for Wittgenstein, those basic nonsensicalities (of doubt, mistake and insufficient justification) are somehow connected up with the expressivist insight, it doesn't seem we can properly see

[56] Cf. (Wittgenstein, 2009a, §247).

Wittgenstein as taking up a particularly *expressivist* line, not with regard, anyway, to this second question about the epistemic status of privileged utterances. The second thing I want to bring out is that, while perhaps not motivated by expressivist concerns, Wittgenstein's view does inherit the familiar technical difficulties attending the simple expressivist line on the question of epistemic status. I'm not, however, going to attempt in this chapter to marshal a Wittgensteinian response to those difficulties; that will have to wait until next chapter, where I'll be considering a most impressive effort by Peter Hacker to do just that.

Now for my Wittgenstein on (Q3), of the far removal of cases. I mentioned in chapter two that, according to Wittgenstein, not every use of "I am afraid" approximated equally closely to cries of fear. The remark I had in mind comes from *PPF*:

> A cry is not a description. But there are intermediate cases. And the words "I am afraid" may approximate more, or less, to being a cry. They may come very close to one, and also be *very* far removed from it. (Wittgenstein, 2009b, §83)

So, it might be thought that that settles it. Wittgenstein admits there are mental state self-ascriptions that are 'farther away' than others are from pre-linguistic behavioral expressions. But, actually, that doesn't exactly settle it, because the question wasn't just whether there are (or which exactly are the) mental state self-ascriptions that are 'farther away' than others from cries and smiles and the like, the question was whether there are (which exactly are the) *privileged* mental state self-ascriptions that are 'farther away' than others from cries, smiles, etc.. I don't doubt that any expressivist about first-person *privilege* would be more than happy to grant that there should be *unprivileged* mental state self-ascriptions that are 'far removed' from, even

bearing no interesting relation to, pre-linguistic, behavioral expressions of mentality. Consider the following remark of Wittgenstein's:

> Does it makes sense to ask "How do you know that you believe that?" -- and is the answer: "I find it out by introspection"?
>
> In *some* cases it will be possible to say some such thing, in most not.
>
> It makes sense to ask, "Do I really love her, or am I only fooling myself?", and the process of introspection is the calling up of memories, of imagined possible situations, and of the feelings that one would have if… (Wittgenstein, 2009a, §587)

Four things about this remark: First, as I read him, Wittgenstein is opposing "I've found out what I believe by introspection" to, perhaps among other things, "I haven't found it out anyhow, but I've *expressed* my belief authoritatively"; Second, it should be clear that by "introspection" Wittgenstein doesn't mean anything like 'direct acquaintance', and that what he does say here about what introspection would amount to in the case of "Do I really love her…" (viz. "the calling up of memories…") leaves it open whether one could indeed give an expressivist treatment of one's authority (if any) with respect to what one thus "introspects"; Third, it seems clear to me that the clause "am I only fooling myself" indicates an *absence* of privilege attaching to the claim "I really love her" that's being called into question; and Fourth, this absence of privilege attaching to "I really love her" would seem to attach just as well, given their juxtaposition, to the (in the background) claim above it that "I believe that", whatever "that" may be. So the point is, while Wittgenstein admits here that one's belief-claims[57] can in *some* cases be arrived at introspectively, as opposed to a belief-claim's simply expressing one's belief, one's privilege (if any) with respect to the immediate data, you might say, of this kind of introspection might still be given an expressive (as opposed to introspective) treatment, and anyway the belief-claim based

[57] I mean claims of the form "I believe that…".

on such introspection *isn't privileged*. And so, even though we have a commitment here to an introspective (as opposed to expressive) way of arriving at a belief-claim, we don't yet have a commitment to any *privileged* mental state (here belief) self-ascription that is other than expressively arrived at. Again, what we have to find, if Wittgenstein is to diverge from the simple expressivist's response to my third question, is a commitment on his part to *privileged* mental state self-ascriptions that are somehow or other far removed from cries, smiles, and the like.

This is important because certain other of Wittgenstein's remarks about the "descriptive", "reportive" or "self-observational", as opposed to the expressive, quality of certain first-person mind-descriptions would seem less than clear on the question of their privilege. Take, for example:

> When someone says "I hope he'll come", is this a *report* about his state of mind, or a *manifestation* of his hope? -- I may. For example, say it to myself. And surely I am not giving myself a report…If I tell someone, "I can't keep my mind on my work today; I keep on thinking of his coming" -- *this* will be called a description of my state of mind. (Wittgenstein, 2009a, §585)

Or, for example:

> "I've heard he is coming; I've been expecting him all day." This is a report on how I have spent the day…The exclamation "I'm expecting him -- I'm longing to see him!" may be called an act of expecting. But I can utter the same words as the result of self-observation, and then they might amount to: "So, after all that has happened, I'm still expecting him with longing." It all depends on what led up to these words. (Wittgenstein, 2009a, §586)

Shall we say that in both of these too, there isn't any commitment to *privileged* first-person mental state self-ascriptions that are other than apt for expressivist treatment ("a manifestation of his hope" and "an act of expecting" indicating an aptness for expressivist treatment)? Well, these descriptions, reports and self-observations all

seem to involve crucial reference to fairly long stretches of time ("all day"), and even, in the last case, to "all that has happened", which presumably doesn't just include what one has privilege with respect to. So maybe Wittgenstein doesn't, just like the simple expressivist doesn't, think there are in fact any 'far removed' cases that are *privileged*? I don't think that's right, though.

Let's look at a particularly telling passage:

> I say "I'm afraid"; someone else asks me: "What was that? A cry of fear; or did you want to tell me how you feel; or was it an observation on your present state? -- Could I always give him a clear answer? Could I never give him one?
>
> One can imagine all sorts of things here: for example, "No, no! I'm afraid!"
>
> "I'm afraid. I am sorry to have to admit it."
>
> "I'm still a bit afraid, but no longer as much as before."
>
> "In fact I'm still afraid, though I'm reluctant to admit it to myself."
>
> "I torment myself with all sorts of fearful thoughts."
>
> "Now, just when I should be fearless, I'm afraid!"
>
> To each of these sentences a special tone of voice is appropriate, to each a different context.
>
> It would be possible to imagine people who, as it were, thought much more precisely than we, and used different words where we use only one. (Wittgenstein, 2009b, §§73, 4)

I take it that "No, no! I'm afraid!" is supposed to be a cry of fear, and something privileged, and I take it that the rest are supposed to be other than cries of fear. Now, we might be able to pull the "not necessarily privileged" trick with most of these, viz. with "I'm *still* afraid", "I torment [as in *have been tormenting*] myself…", and even "I reluctantly admit to myself that I'm afraid" and "Now, just when I should be fearless, I'm afraid!"; these, we might argue, aren't unequivocally privileged ascriptions of fear.[58] But I don't think we're going to be able to pull this off at all

[58] If we interpret the last one as something like "What a coward I am for being afraid at this!", then we could argue that what we have in these cases are either matters to do with *ongoing* affairs, matters with respect to which one may well be *self-deceived*, or matters to do with one's *character*, none of which,

successfully for the whole range of examples Wittgenstein has set out in this passage. The obstinate example will be the second one, "I'm afraid. I am sorry to have to admit it". As I read this passage, this is supposed to correspond to *telling* someone else how one feels, what Wittgenstein began the passage by *opposing* to the use of the words "I'm afraid" as a cry of fear. And this much is true, if the expressivist about first-person privilege either denies first-person privilege in the *telling* to others of one's states of mind, or else if she isn't even talking about the privilege attaching to cases wherein one *tells* another what one feels, then the expressivist isn't making any significant contribution to the problem of first-person privilege. And this much is true too: if simply *telling* another about one's state of mind is supposed to be something one may do with privilege, and if simply *telling* another about one's state of mind is also supposed to remove one (however distantly) from expressing one's state of mind (expressing, it may be granted, in some *language* or other), then the vast majority of what one says with privilege about one's own mental life will have to be (however distantly) removed from a basic kind of expressing of one's mental life in words.

In any case, I find it highly implausible that Wittgenstein would've denied first-person privilege (i.e. the senselessness of doubt, mistake, and insufficient justification) to what one admits (tells) to another about one's own state of mind. So, insofar as he sees this telling as taking us away from a certain kind of basic way of using language to express one's state of mind, it seems Wittgenstein's response to my third question will be "Yes, some privileged mental state self-ascriptions are farther

---

we could argue, is unequivocally a matter of privilege. I'm not going to argue against this line, though I think it's obviously far-fetched.

than others from pre-linguistic behavioral expressions of mentality." But as for *exactly which* and *exactly to what extent*, we aren't going to get much at all from Wittgenstein besides a handful of examples here and there. Presumably, the 'exactly which and to what extent' questions are the sort of questions Wittgenstein thought we could think through for ourselves. In any case, we can see Wittgenstein diverging from our simple expressivist on the third question. The simple expressivist, remember, had it that *every* privileged utterance was to be just as near, that is *very* near, to pre-linguistic behavioral expression. But this kind of divergence from the simple line can come, we shall see, with quite a price; that is, if one wants a comparison with natural behavioral expressions of mentality to be the significant source of light on the problem of privilege that the expressivist is longing for.

So, now for (Q4), the last question, which is meant to get at a looming problem for the expressivist who would follow Wittgenstein's lead on the question of the far removal of cases. The fourth question for the expressivist was, What is the relation between the privilege attaching to the far removed cases and the privilege attaching to the more basic cases, the more basic cases being where the comparison with *expression* was to be the light-shedding comparison? Does the privilege fade, for instance, as one moves farther away from the basic cases? But isn't "privileged only insofar as expressive" just another way of saying "privileged only if expressive"? So the privilege will have to be steady throughout. But is that only because of some important relation the far removed cases bear to the basic cases? It seems to me that if one is going to be a serious expressivist about first-person privilege, and is going to grant privilege in other than basic cases, then one had better

find some way of making sense of the privilege attaching to the far removed cases *in light of* some special relation they bear to the more basic cases. Wittgenstein, as far as I know, never considered doing anything like this. And in this we see a real problem with thinking of Wittgenstein as a serious expressivist about first-person privilege, at least if we think of Wittgenstein as the Wittgenstein left to us in his posthumously published philosophical work.

Worse, though, for those who would make Wittgenstein out to be an expressivist about privilege, is the fact that, while he never bothered to relate far removed cases of privilege to basic, expressive cases, his anti-Cartesianism with respect to first-person privilege in general can be seen to rest *entirely* on his answer to my second question, which answer, as already noted, could seem to have nothing whatsoever to do with an expressivist insight. Indeed, as Crispin Wright has noted, the very attempt to make out Wittgenstein's response to my second question to have anything to do with an expressivist insight would seem to fall afoul of Wittgenstein's explicit disavowal of all things philosophically explanatory:

> We are asking: what is the *explanation* of the characteristic marks of avowals? …Cartesianism takes the question head on, giving the obvious, but impossible, answer. And the expressivist proposal, radical though it is…is not so radical as to raise a question about the validity of the *entire explanatory project*. But Wittgenstein, seemingly, means to do just that. Against the craving for explanation, he seemingly wants to set a conception of the 'autonomy of grammar'. [59] (Wright, 1998, p. 39)

This paves the way for his *constitutivist* reading of Wittgenstein on the question of first-person privilege:

---

[59] Wright's footnote to "'autonomy of grammar'" reads "As Baker and Hacker style it" without any citation.

> Why shouldn't psychological discourse's exhibition of [first-person privilege] be regarded as primitively constitutive of its being *psychological*, so that the first-third-person asymmetries that pose our question belong primitively to the 'grammar' of the language-game of ordinary psychology, in Wittgenstein's special sense--'grammar' which 'is not accountable to any reality' and whose rules 'cannot be justified by showing that their application makes a representation agree with reality'? [60] (ibid.)

Readers of Wittgenstein will have to agree that Wright has a point here. Isn't the expressivist about first-person privilege trying to *explain* certain type (1*) matters of grammar? Isn't she trying to show *why* doubt, mistake and insufficient justification shouldn't enter in to certain mental state self-ascriptions when she compares them with pre-linguistic, behavioral expressions of mentality? Doesn't this run afoul of the autonomy of grammar?[61] Is *this* the sense in which Wittgenstein is no expressivist theorist, the sense in which expressivist theorizing *really is* the kind of bad theorizing (the kind that flouts Wittgenstein's second insight, that grammar is *conventional*) that Wittgenstein wanted nothing to do with? Is it not that he wouldn't see the expressivist insight into a fully worked out expressivist theory, but rather that the expressivist insight completely misses the spirit of Wittgenstein's anti-Cartesianism? I have to say, while I can see where Wright is coming from, so to speak, I find this suggestion somewhat hard to take.

Consider again Wittgenstein's so-called plan for the treatment of psychological concepts:

> Psychological verbs characterized by the fact that the third person of the present is to be verified by observation, the first person not.
>
> Sentences in the third person of the present: information. In the first person present: expression. ((Not quite right.))

---

[60] Wright's citation for these last quotes is (Wittgenstein, 1974, sect. X, §§133 and 134).

[61] As Baker and Hacker style it.

> The first person of the present akin to an expression. (Wittgenstein, 1980b, §63)

Here there couldn't be anything more obvious than that the distance, for Wittgenstein, between sentences in the "third person of the present" and those in the "first person present" is supposed to be the distance between something "verified by observation" and something "akin to an expression". And consider again Wittgenstein's remark from the *Blue Book*:

> Now, when in this case no error is possible, it is because the move which we might be inclined to think of as an error, a 'bad move', is no move of the game at all. (We distinguish in chess between good and bad moves, and we call it a mistake if we expose the queen to a bishop. But it is no mistake to promote a pawn to a king.) And now this way of stating our idea suggests itself: that it is as impossible that in making the statement "I have toothache" I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me. To say, "I have pain" is no more a statement about a particular person than moaning is. (Wittgenstein, 1969, p. 67)

Now, why is *that* the way of "stating our idea" that "suggests itself"? Why not *just* the comparison with an illegal, as opposed to a bad, move in chess? And why, now, can't a comparison of privileged utterances with natural expressions of mind (which comparison Wittgenstein made seemingly countless times) be used in the fight against a Cartesian picture of privilege by way of inner observation? Because that would be trying to *explain* privilege instead of just describing it? And Wittgenstein doesn't try to explain anything?

While I can see why Wright wants to say what he does about reading an expressivist theory of privilege into Wittgenstein, about how such a theory tries to explain where Wittgenstein would only describe, I don't think that, really, Wright's reason is the right reason for counting Wittgenstein out of the expressivist camp on

the question of first-person privilege. First, I think that what's really working on Wright's mind are the technical difficulties in trying to work out a proper expressivist theory of privilege. Second, I think that he's helping himself to an unhelpful slogan when he calls expressivism an un-Wittgensteinian try at explanation. As I said before, such problematic titles for readers of Wittgenstein as "explanation", "theory", and the rest all admit of such thin senses that one shouldn't invest too much in combating anything and everything that might come in under them. There seems to me to be a perfectly innocuous sense of "explanation", a sense in which one's explanans simply *sheds light on* one's explanandum, and in which the expressivist about first-person privilege can explain, quite in the Wittgensteinian spirit, first-person privilege by comparing privileged utterances with natural, behavioral expressions of mentality. On my way of looking at it, a Wittgensteinian expressivist *is* a constitutivist, in the sense that she wouldn't try to explain in the *bad* way any type (1*) fact of grammar, as though it were somehow responsible to some necessary feature of reality, but she would try to explain in a *good* way, shedding light on first-third-person asymmetries with her comparison to expressive behavior.

Still, we're faced with the facts that Wittgenstein didn't always bring expression into the picture when he discussed the problems with "I know I'm in pain", and that Wittgenstein never said anything about relating the privilege of utterances far removed from expression to the privilege of utterances not-so-far removed from expression. Well, about this I can only say that he did *sometimes* (I've just given two examples) bring in a comparison with expression when discussing such things as the type (1) nonsensicality of mistake, doubt, and insufficient justification

attaching to certain mental state self-ascriptions. And anyway, he brought in that comparison in many more places besides just those two; and while perhaps not *explicitly* in connection with "I know I'm in pain", my question is, How could anyone think he wasn't *thereby* discussing the problems with "I know I'm in pain"? And as for his not answering my fourth question, all I can say is that, on my reading, this does pull him out of the camp of serious expressivist theorists, but not because such 'theorizing' is bad bad bad, rather because, as far as I can tell, he had better things to think about.

We might ask, Had he thought through my question, would he have turned out a constitutivist about the far removed cases or would he have seen an important relation between the privilege attaching to the far cases and that attaching to the near? My answer is that I don't know. I don't myself see exactly what an expressivist should say on this question, and this is largely why I don't see that expressivists will be able to avoid a non-expressivist constitutivism in the far removed cases. As we'll see in the next chapter, Dorit Bar-On takes a stab at an expressivist answer to my fourth question, something like, I have to say, the simple expressivist line I set out last chapter. But the other two expressivists, Peter Hacker (or Hacker's Wittgenstein) and David Finkelstein never address the question. I'll have more to say about the threat of constitutivism in the last chapter.

So, in sum, the extent to which Wittgenstein is an expressivist theorist about first-person privilege is that, first of all, he does indeed seem to let the expressivist insight shine its light on the problem of first-person privilege. This is a type (2*) insight into the workings of language, and is in no way grammatically dogmatic, as I

hope will be clear. Secondly, he also works out the expressivist insight at least enough to addresses my questions (Q1)-(Q3). And while his answers to the first and the third questions obviate two sets of technical difficulties that would attend such tries at getting serious with the expressivist insight as my own simple expressivism, his answer to the second question indeed brings along a worrying set of technical difficulties to be addressed in the next chapter. Thirdly, his answer to the first and third questions are anything but grammatically dogmatic, while his answer to the second can seem so only if we don't grant the anti-dogmatic spirit of the crucial type (2*) clarifications involved in his arrival at the paradoxical-seeming type (1*) pronouncements. Fourthly, he doesn't address my fourth question, and so cannot be as fully working out an expressivist theory of privilege; but that doesn't, as I see things, indict the expressivist insight as an un-Wittgensteinian try at explaining what must not be explained; it only points to nobody's having asked him my question. And lastly, on the question what he *would've* said about my fourth question, had somebody asked him, I can only hope with this dissertation to stimulate, however minimally, someone to thoughts of her own. I'll have to wait until the last chapter, wherein I'll consider the fate of expressivism as a viable theory of first-person privilege, before I can offer much of anything specifically intended in the way of such stimulation. Next, however, I turn to examining three other expressivists' expressivisms.

CHAPTER IV
FINKELSTEIN'S, HACKER'S (WITTGENSTEIN'S) AND BAR-ON'S EXPRESSIVISMS

Introduction

In this chapter I'm going to consider three expressivisms. The first comes from David Finkelstein. The main point of interest in Finkelstein's work for our discussion is the way he handles my first two questions. Surprisingly, he doesn't in any place I know of address my second two questions, and so I don't think we can call his a fully worked out expressivist theory of first-person privilege.[62] Much of his work on the subject is toward undermining detectivist (more or less perceptual/observational) and constitutivist views[63], and toward letting the expressivist insight help with the problem of conscious and unconscious mental states[64]. These are no doubt valuable contributions to the expressivist project, and I suppose it's a testament to the inherent interest of that project that there are further important questions to be asked, viz. my (Q3) and (Q4). In any case, he doesn't

[62] As a reminder, (Q1): Are privileged because expressive utterances truth-apt? (Q2): Do privileged because expressive utterances represent genuine self-knowledge? (Q3): Are there privileged utterances that are in any sense farther removed than others from primitive cries and gestures? (Q4): If so, how does the privilege attaching to the farther removed cases relate to the privilege attaching to the more nearly expressive cases?

[63] Detectivist accounts of privilege, says Finkelstein, are those appealing to cognitive processes whereby one "finds out" about one's mental states, and which provide one with "better epistemic access to [one's] own mental states than other people have to them." (Finkelstein, 2003, p. 9). Constitutivist accounts, he says, are those according to which "mental state self-ascriptions are unlike observation reports in that they constitute, to some significant extent, the facts to which they refer." Rather as an Army colonel "*declares* an area off limits…his authority consists in the fact that what he says goes." (Finkelstein, 2003, p. 28) Bar-On distinguishes, rightly, between the "ontological" constitutivism Finkelstein has in mind here and the "grammatical" constitutivism discussed in last chapter, attributing both, rightly, to Crispin Wright. (Bar-On, 2009, pp. 60, 1) This is the last we'll see of the ontological variety, but the grammatical sort will come up again in the last chapter.

[64] See (Finkelstein, 2003, pp. 114-127).

address these two, but the way he addresses the first two will be of interest for our discussion.

The second expressivism comes from Peter Hacker's reading (and defense) of Wittgenstein's work on avowing and first-person privilege. His reading is much like my own (minus my distinction between kinds of descriptions of our grammar), so the main interest in Hacker will be in seeing how he defends Wittgenstein on the matter of the contention that where there is first-person privilege, there isn't any self-knowledge. And also, because his reading is much like my own, it isn't clear whether Hacker's Wittgenstein is properly classed an expressivist about privilege, not even in the harmless sense of "-ist". All the same, it will be clear that expression is supposed to have something to do with the story of privilege as Hacker tells it.

The third expressivism is Dorit Bar-On's. She's the only one who directly addresses all four of my questions, and so hers stands out as the most fully worked out of the expressivisms currently on offer (along with the simple expressivism I set out in chapter two, which, though simple, was still meant as a fully worked out expressivist theory). Hers also helps to show just how dire the situation might be for the expressivist. Why this is so will have to wait until I come to discussing her view later in this chapter, though, and when I reflect more generally on the expressivist's predicament in the last chapter.

Finkelstein's Expressivism

The first thing to say is that Finkelstein isn't expressly out to capture first-person privilege in exactly the way I've defined it, that is, in terms of authority, groundlessness and transparency. Rather, Finkelstein sees himself as trying to help

specifically with his own version of privilege, what he calls "first-person authority", the main (problematic) feature of which he characterizes thus: "If you want to know what I think, feel, imagine, or intend, I am a good person--indeed, usually the best person--to ask." (Finkelstein, 2003, p. 100) While this isn't precisely the definition of "privilege" (or "authority") I've been working with, I think it's fair to say that his account, if it does what's it's supposed to do with respect to his 'authority' thus characterized, can count as an expressivist treatment of the problem of first-person privilege as I (following Wright) have characterized it.[65]

In his *Expression and the Inner*, after criticizing detectivist and constitutivist attempts at understanding first-person authority, Finkelstein sets out the main thrust of his expressivist line on first-person authority, more or less the expressivist insight: "If you want to know my psychological condition, I'm usually the best person to ask, for just the same reason that my face is the best one to look at." (Finkelstein, 2003, p. 101)

And so Finkelstein raises the immediate question, Why do we think of mental state self-ascriptions as *authoritative* in a way we do not think of smiles, e.g., as authoritative? His answer to this is really his answer to my (Q1) about truth-aptness: "smiles aren't authoritative because they aren't assertoric," Finkelstein writes, "avowals, on the other hand, are assertoric, and so can be called 'authoritative'". (Finkelstein, 2003, pp. 101, 2) Of course, this makes it look like "authoritative" is just another word for "assertoric" when it comes to the cases Finkelstein is interested in, and this point will be returned to presently, but right now the interest is in

[65] For the purposes of this discussion of Finkelstein, then, I'll use the term "(first-person) authority" as he does.

Finkelstein's free admission that avowals, of whose authority he aims to give an expressivist treatment, are indeed assertions. He sees himself as working in the vein of Wittgenstein's own expressivism:

> While Wittgenstein does say that mental state avowals are expressions, he does not deny that they are truth-evaluable. The influence of the [non-truth-evaluable] expressivist reading is based, to a large extent, on an assumption that has no real foothold in Wittgenstein's writing, an assumption that could be stated as follows: "A given speech act might describe the speaker's state of mind--i.e., assert what state of mind she is in--*or* it might express her state of mind. But it can't do both these things. Expressing and asserting are, in a way, mutually exclusive." (Finkelstein, 2003, p. 95)

One thing to note is that, for Finkelstein, avowals' being descriptive, assertive and truth-evaluable are all one and the same fact. This would seem to mark a departure from the Wittgensteinian view in *PI* (Wittgenstein, 2009a, §585), wherein 'description' seems closer to '*report*' and is clearly *opposed to* 'manifestation', which would seem very close to 'expression'. For Finkelstein's Wittgenstein, who would deny the assumption in the quote above, avowals are one and all descriptive *as well as* expressive, *contra PI*, §585. I think, though, that whatever tension there is between Finkelstein's Wittgenstein and *PI* §585 can be smoothed over on Finkelstein's behalf by noting that he's most certainly working with the propositional, not the pragmatic (reportive), sense of "description"; his point is simply that, even for Wittgenstein, an "avowal of hope that is not a report may yet be a true assertion". (Finkelstein, 2003, p. 97) And this is precisely what Finkelstein sees at work in Wittgenstein's plan for the treatment of psychological concepts, when Wittgenstein backed off of "In the first person present: expression" with "((Not quite right.)) The first person of the present *akin* to an expression." (Wittgenstein, 1980b, §63, my italics) For Finkelstein, Wittgenstein's retreat to "akin" is supposed to signal just this

allowance for truth-evaluability (truth-aptness). Finkelstein's Wittgenstein "thinks there is an assertoric *as well as* an expressive dimension to mental state avowals". (Finkelstein, 2003, p. 99) This is exactly the view of things Finkelstein himself adopts.

It should be noted as well that it isn't clear whether Finkelstein is working with the deflationary approach to my first question, the one that Malcolm and I both read into Wittgenstein, and so it's not clear to me whether Finkelstein can have Wittgenstein's expressivist treatment of avowing set up at once in opposition to the inner/outer picture of mind and behavior *qua* ontological *and qua* epistemological fantasy. Someone who thought that the expressivist treatment of avowals we find in Wittgenstein was meant to combat, not just the detectivist conception of mental state self-ascription, but also any word-here/referent-there (sentence-here/fact-there) ontology (semantics) of mind (talk) would avoid, I'd think, speaking of an "assertoric dimension" of (at least the most nearly expressive) avowals that could be anything but *transparent to* their "expressive dimension". His talk of an avowal's expressive dimension as something in addition to its assertoric dimension, then, could very well mark a departure from the reading of Wittgenstein I offered in the last chapter. In any case, it's clear he departs from the simple expressivist; Finkelstein admits truth-values for privileged (because expressive) utterances.

Now, while it's not clear to me whether Finkelstein indeed departs from the deflationary view I've attributed to Wittgenstein, what is clear to me is that, given Finkelstein's treatment of my question about the epistemic status of avowals (that is, (Q2)), he'd do well to show how his two 'dimensions' *can* fit into a story of the

ontology of the mental whereby avowals' expressive dimensions somehow or other shape, or are anyway intimately connected with, their assertoric dimensions. Let's have a look at his response to my second question, then, to see why.

So, we saw that, for Finkelstein, the word "authoritative" as applied to avowals just amounts to "assertoric", which itself just amounts to indicating the presence of a truth-evaluable dimension. But the word "authoritative" suggests, ostensibly, epistemic praiseworthiness, while "assertoric", ostensibly, doesn't. Now, for Wittgenstein, on my reading, epistemic praise and blame were indeed non-applicable to avowals, the illusion of knowledge and certainty being products of the (grammatical) exclusion of ignorance and doubt; but also, for my Wittgenstein, it was allowed that one said something truth-apt in the issuing of an avowal. So how is it Finkelstein helps himself to such an epistemologically charged word as "authoritative" just on the strength of his case for admitting "assertoric"? Doesn't Finkelstein accept Wittgenstein's view that where there isn't any identification by means of criteria, where there isn't any observation, there isn't any justification, and there isn't any knowledge either?

When Finkelstein (finally) confronts this question head on in *Expression and the Inner* the response he gives is, more or less, that we can say what we want; we can say that epistemic praises apply or that they don't. As Finkelstein sees things, if we want to talk of knowledge of one's own pain, for instance, we can do so, as long we don't make the mistake of thinking that "knowledge always requires epistemic justification." (Finkelstein, 2003, p. 151) He says we can indeed have it that one knows of one's own pains, as long as we also have it that "knowing this isn't like

knowing that one has termites in one's basement." (Finkelstein, 2003, p. 152) If we allow this kind of knowledge of one's own pain, then, Finkelstein doesn't see that "we need be disagreeing about anything of philosophical import" (ibid.) with someone who, because of sensitivity to the expressive character of avowals, refuses to talk of such self-knowledge. I don't doubt that he would see this as a Wittgensteinian maneuver. I don't doubt he'd be right.

That is, it seems to me that, while the 'grammatical' line I attributed to Wittgenstein in the last chapter is more committal than Finkelstein's own, I don't doubt that Wittgenstein would've been perfectly happy talking about self-knowledge of one's own pain, beliefs, thoughts, hopes, etc., so long as the "logical grammar" (to use Malcolm's phrase) of such knowledge-talk were kept perfectly in view. Indeed, it seems to me that, while Finkelstein seems more cagey than Wittgenstein on (Q2), perhaps his response really amounts to nothing more than part of the Wittgensteinian response already noted: that "I must know whether I'm in pain" *can* be used as a grammatical statement, because then it means that doubt, mistake and insufficient justification are grammatically excluded. If, as it seems Finkelstein may very well intend, *all we mean by* "authoritative", when it comes to avowals, is "assertoric", then calling avowals "authoritative" seems, after all, quite in line with the Wittgensteinian treatment, which only called for the elimination of terms of epistemic praise insofar as they were meant to oppose terms of epistemic blame. Again, if the term of epistemic praise is applied just in the way of noting the grammatical exclusion of the term of epistemic blame, then Wittgenstein is on board, as I read him; though I don't

doubt he would've seen the perfectly grammatically well-adjusted use of such 'terms of praise' as intended anywise other than comically.

This much seems true, however, if Finkelstein *isn't* just making the familiar Wittgensteinian point that admitting such kinds of self-knowledge is simply a highly misleading way of gesturing at the grammatical exclusion of such kinds of self-ignorance, then I can't say I have much of a clue what he's doing with his say-what-you-want strategy. In any case, it's clear that, one way or another, he doesn't think there's anything *substantial*, anything of "philosophical import", in the attribution of knowledge to people of their avowable mental states. And it's clear, furthermore, that it's supposed to be his expressivist account of first-person authority that will lend whatever strength there is to be lent to this say-what-you-want maneuver regarding the question of self-knowledge; I presume you can only say what you want about self-knowledge, for Finkelstein, if you have the expressivist view of first-person authority. In this way, Finkelstein's expressivist view of self-knowledge is *preceded by* his expressivist view of authority.

Now, I said that this would pose a problem for Finkelstein if he didn't have anything special to say about the relation between the assertoric and expressive dimensions of avowals, that is, about the relation between the assertoric content (truth-conditions) of privileged, because expressive, psychological self-ascriptions and the fact that such self-ascriptions are, indeed, *expressive*. Wittgenstein certainly seemed to see a tight connection between them when he wrote, "Of course 'toothache' is not *only* a substitute for moaning. But it is *also* a substitute for moaning, and to say this shows how utterly different it is from a word like 'Watson'".

(Wittgenstein, 1993, p. 298) Here, the fact that "toothache" is, even just *sometimes*, a substitute for moaning (that is, presumably, offered expressively as opposed to reportively) goes to show how "utterly different" the word is from words like proper names for people. The point here has got to be just the anti-Augustinian point running throughout Wittgenstein's later work, and which is arguably the touchstone for orienting oneself to that later work[66], that the name-object (sentence-fact) conception of language use must yield to a conception of language- use more attuned to the ways we actually learn to use words in whole sentences, the ways we actually verify those sentences, and also to the facts about just which sentences containing those words we countenance as actually 'making sense'[67]. Furthermore, it seems to me, unless one goes in for Wittgenstein's anti-Augustinianism, at least about the uses of privileged mental state self-ascriptions, nothing like Finkelstein's dismissiveness of the seriousness of the question of first-person knowledge will even begin to get a grip. That is, no matter how well one can account, from a third-person perspective, for the phenomena of first-person privilege (either the version I set out in chapter one or the stripped-down version with which Finkelstein concerns himself), it seems to me the question of *self-knowledge* remains a most pressing one as long as, *per* the Augustinian mindset, the question "Does the proposition that *p* correspond to reality?" is conceived entirely independently from questions like "What right have I

[66] See (Baker and Hacker, 1980, pp. 1-28)

[67] These were just the sources of illumination (i), (ii), and (iii) we saw Wittgenstein drawing from in chapter two.

to claim that *p*?" and "How is one taught to make the claim that *p*?".[68] In the case at hand, the case of privileged mental state self-ascriptions, it seems to me that unless one has it that the very fact that one has a toothache, for instance, is *constituted* (at least in part) by the expressive character of (at least some) claims to have a toothache, then the question how one *knows* that one's own true claim to have toothache is in fact true will remain very much a live one.[69]

As William Alston has pointed out:

> I can express my enthusiasm for your plan just as well by saying "I'm very enthusiastic about your plan", as I can by saying "What a tremendous plan!", "Wonderful", or "Great!". I can express disgust at *X* just as well by saying "I'm disgusted", as by saying "How revolting!" or "Ugh". (Alston, 1965, p. 16)

And so he saw, just like Finkelstein, and just like anybody can I think, that expressing one's state of mind isn't mutually exclusive with asserting that one is in it. (ibid.) But Alston was no expressivist about the phenomena first-person privilege; this is telling.

So let's consider, then, someone who's annoyed and who says gruffly, "The keys are in the basket". (She was asked where the keys were by someone who should've known already, let's say.) Now, she not only asserts something (that the keys are in the basket), but she expresses her annoyance along with what she asserts. In this we can see what might be called an expressive dimension right along side an

---

[68] It's not for no reason that Wittgenstein has been interpreted as an anti-realist by way of verificationism. See (Dummett, 1993) for a standard interpretation in that vein. Those who would distance Wittgenstein from strict, positivistic, reductionistic, anti-realistic verificationism, however, do well to seize on those of his later remarks to the effect that how one verifies a given proposition is (merely) a *contribution* to its 'grammar'. See (Gasking and Jackson, 1967, p. 54), (Hacker, 1990, pp. 379, 80) and (Malcolm, 2001, p. 55).

[69] For very much the same reasons, Matthew Boyle, in his criticism of Bar-On's expressivist treatment of the 'security' of avowals, makes very much the same point, namely that if an expressivist treatment *admitting truth-values* to avowals is going to help with the problem of *self-knowledge*, then solving that problem will demand from us "not merely to account for how mental states *are known* to their subject, but to rethink our conception of what sorts of things mental states *are*." (Boyle, 2010, p. 12) Finkelstein, we see, doesn't want the problem of how they're known to their subject to be anything substantial, but, as I'm trying to bring out, the ontological question is nonetheless ineluctable.

assertoric dimension; no surprise. Furthermore, we don't see any particularly philosophically interesting connection between them. One thing should certainly be clear: in this case, pointing out that one has expressed one's annoyance right along with saying where the keys were wouldn't itself (how could it?) go toward settling any concerns about how one *knew* where the keys were, what one's *justification* was in claiming they were in the basket, what *right* one had in so claiming. But then let's consider someone who's getting rather annoyed and who says (through clinched jaw and furrowed brow) "I'm getting rather annoyed." Again, she asserts and expresses. And, as Finkelstein has it, her first-person authority (that is, for him, our right to take her word for it) resides simply in that what she expresses (her becoming rather annoyed) is precisely what she asserts to be so. As Finkelstein says, first-person authority emerges as an "unsurprising concomitant" (Finkelstein, 2003, p. 102) of the expressive character of mental state self-ascriptions. But, as I've been trying to say, Finkelstein leaves out something important insofar as he leaves it open whether the expressive and the assertoric are just as distinguishable in the case of someone's claiming (annoyedly) "I'm getting rather annoyed" as in the case of someone's claiming (annoyedly) "The keys are in the basket".

If these two dimensions are just as distinguishable in both cases, then in the case where (growing) annoyance is avowed we can ask as before: Why should pointing out that annoyance was expressed (and not merely asserted) itself go toward settling any concerns about how one *knew* one's claim to be annoyed corresponded to reality, what one's *justification* was in making that claim, wherein one's right to claim it consisted? Granted, from a third-person perspective, that it was precisely

annoyance that was expressed will perhaps go some way toward accounting for our being so happy to take the assertion of annoyance for the truth; but their coinciding this way (what was expressed and what was asserted) won't itself account for our refraining from asking the epistemic question about the avower herself; not if their coinciding is modeled after the coinciding of the assertoric dimension of the keys-in-the-basket claim and its expressive dimension.[70] On the other hand, though, as already mentioned, it seems to me the expressivist has a much better chance at handling the question of the epistemic position of the avower with regard to the *assertoric content* of her avowal when it's granted that such content is what it is (at least in part) precisely because the avowal is (at least sometimes) *expressive*, because, that is, the avowal bears precisely the expressive dimension it does bear.

Now, this certainly seems to be so for Wittgenstein; but, as already mentioned, it's not clear where Finkelstein stands on the question of the relationship between the assertoric and the expressive dimensions of avowals. Finkelstein does have it that avowals are assertions "of a special sort" (Finkelstein, 2001, p. 233), but it's simply unclear whether what he sees special about them is anything besides their asserting precisely what they express (where such distinguishable acts of assertion and expression, though not always coinciding in this way, nonetheless do go on whenever we say anything at all with feeling about how things stand with keys, pains, or anything else you please). On my view, Finkelstein, or any expressivist, simply

[70] In (Gertler, 2008) Brie Gertler writes that while "Neo-Expressivism may help us to understand how avowals provide third-person warrant, that is, warrant for others to attribute mental states to the avower," still "the value of Neo-Expressivism may lie outside its contribution to our understanding of the epistemology of self-knowledge." Gertler, however, doesn't suggest, as I do, that where the value of the expressivist's contribution lies should turn on whether the "Neo-Expressivist" (Dorit Bar-On's term) sees the expressive dimensions of avowals as, to a greater or lesser degree, 'fused with' their assertoric dimensions.

can't leave untouched the question of the semantic (ontological) role of expression in avowing, just treating the problem of privilege (from the third-person perspective), and then leaving whatever concerns there are to do with the question of self-knowledge to take care of themselves, as it were, in light of this (third-person) expressivist treatment of privilege.

As for (Q3) and (Q4), I've already said that Finkelstein doesn't address them. It seems never to occur to him that there might be privileged mental state self-ascriptions that should be somehow or other farther removed than others from cries, smiles, reaches, etc. And so it seems never to occur to him that there should be a problem for the expressivist about relating the privilege of the far removed cases to the privilege of the more nearly expressive cases. Because of this, as I've said, we simply don't get a fully worked out expressivist theory from Finkelstein. For a more worked out version, then, I turn next to Peter Hacker's exposition of Wittgenstein on (Q1)-(Q4).

Hacker's (Wittgenstein's) Expressivism

Peter Hacker is undoubtedly one of the most recognizable names in Wittgenstein exposition; and this is undoubtedly due, at least in part, to his exceedingly sympathetic reading of Wittgenstein's later work. Indeed, the sympathies run so deep that his own personal philosophical views are particularly difficult to disentangle from his best reading of Wittgenstein's. It's plausible to think that wherever he's explicating Wittgenstein he's doing so with approval unless otherwise noted. It's also plausible to think that wherever he's defending any expressivist treatment of avowals, he's explicating Wittgenstein. Because of this it

makes things considerably easier for present purposes not to be too concerned with how much of the following belongs, in point of fact, to Hacker, as opposed to Hacker's Wittgenstein. I'd like to begin with Hacker on (Q3) and (Q4), working my way to Hacker on (Q1), and then getting to the main point of interest, his response to (Q2).

So, on (Q3), of the far removal cases, Hacker, following Wittgenstein, admits a "spectrum of cases" (Hacker, 1990, p. 196):

> The affinity between spontaneous avowals and natural expressive behavior must not mask the fact that the uses of first-person psychological sentences are heterogeneous. Some approximate to primitive cries and gestures, and others are far removed from those paradigms. Wittgenstein was not suggesting that there is no such thing as reporting, informing, telling others how things are subjectively with one. But what is called 'telling someone what one feels', 'describing one's state of mind', or even 'observing one's emotional state' are much more unlike reports, descriptions, and observations of the physical world than one thinks. (ibid.)

In this he departs not one jot from the Wittgensteinian response to (Q3). And in never discussing (Q4), of the connection between the privilege of the far removed cases and those most nearly approximating "primitive cries and gestures", he again follows suit. Unlike Wittgenstein, however, Hacker is rather explicit in connecting up the expressive character of at least some mental state self-ascriptions with their privilege:

> Such first-person psychological utterances, Wittgenstein argued, are, *in the primitive language games out of which their use arises,* essentially *expressive*, not descriptive…They are authoritative (to the extent that they are) not because they are assertions of something the agent *knows*, but because they are *manifestations* of the agent's feeling, thinking or intending whatever he feels, thinks or intends. (Hacker, 2005, pp. 245, 6)

One might suspect that because Hacker thinks that in the primitive language games in which psychological self-ascriptions are essentially expressive they are just to that

extent privileged, that therefore the Hackerian line will be that *all* privileged first-person utterances, even those at some remove from his "primitive language games", are privileged only to the extent that they're expressive. But I think the evidence is too thin to make that leap.[71] And anyway, Dorit Bar-On is explicit about taking such a view on board, so the discussion of that line can wait until later. On my reading of Hacker, then, we have an acknowledgement of the far removal of privileged cases from those most nearly approximating cries and gestures, we have the contention that in the cases nearest to cries the utterances are authoritative because expressive (at least, that's so in the "primitive language-games out of which their uses arises"), but we don't have any try at connecting up the privilege attaching to the far removed cases with the privilege attaching to the more basic cases. In this way Hacker's Wittgenstein, just like mine, and just like Finkelstein, is going to fall short of setting out a fully worked out expressivist theory of first-person privilege.

Turning now to (Q1), it's quite interesting that Hacker changed his mind twice on the matter. When he wrote the first edition of *Insight and Illusion*, he was adamant that the "truth-valueless thesis" was to be rejected.[72] But by the time the second edition came around, when he was then emphasizing the "spectrum of cases" idea, he, so to speak, sang a different tune:

> If we examine our use of first-person psychological utterances in the variety of contexts in which they are appropriate we will find a whole

[71] That is, just because in the "primitive language games" an utterance is privileged only insofar as expressive, that doesn't mean that everywhere and always "privileged only insofar as expressive" is the rule. This might seem like splitting hairs, but I think the more general "privileged only insofar as expressive" plunge is so important (*i.e.* potentially damaging) that one should be very careful about reading such a view into a philosopher's work. Also, it's hard to believe, even though it isn't explicitly addressed, that Hacker would have all those self-ascriptions lying at the "descriptive" end of the "spectrum of cases" lacking in any of the problematic features of first-person privilege.

[72] Cf. (Hacker, 1975, p. 265).

> spectrum of cases. At one end, as a cluster of cases, lie exclamations ('I'm so pleased'), cries of pain ('It hurts, it hurts!') sighs of longing ('Oh, I do hope he'll come'), expressions of emotion ('I'm furious with you') or expectation ('I expect you to come'), avowals of thought or belief, expressions of desire ('I want a glass of wine') or preference ('I like claret') and so forth…At this end of the spectrum the concept of *description* gets no grip, nor does that of *truth*. (Hacker, 1986, p. 298)

It's quite interesting that Hacker should make this shift, since, it was in response to Hacker's first edition of *Insight and Illusion*, wherein he accused Malcolm of propounding the truth-valueless thesis, that Malcolm explicitly gave the reading of Wittgenstein I myself gave in the last chapter, viz. that "true" and "false" are quite appropriately applied to first-person psychological utterances, however near to primitive cries and gestures. What's more interesting still is that Hacker changed his mind again. In a 2005 article, Hacker writes:

> Utterances of pain, e.g. the exclamation 'It hurts' or the groan 'I am in pain'…are acculturated extensions of natural pain behaviour…To say that such utterances are acculturated *extensions* of natural expressive behaviour does not imply that they are just like the primitive behaviour on to which they are, as it were, grafted. *On the contrary* -- unlike the natural behaviour such linguistic behaviour can be truthful or dishonest. What is said by such utterances may be true or false, no less than the third-person counterparts. It stands in logical relations of implication, compatibility or incompatibility with other propositions. (Hacker, 2005, p. 246)

It seems that the end of the spectrum nearest cries and gestures has been something of a problem for Hacker.[73] In any case, I suppose that, since in two out of three of these stages of Hacker's thought on the subject he granted truth-aptness to avowals, that will be the Hackerian response to (Q1) we'll work with. So he, like every expressivist we will have considered besides my simple expressivist, avoids the set of

[73] It's interesting that Hacker doesn't mention the exchange with Malcolm, and also that he doesn't give even the little that Malcolm did (about the various "logical grammars" of "true" and "false") or even a more clearly deflationary line like that I attributed to Wittgenstein. He simply says avowals may be true or false, and leaves it at that.

technical difficulties attending the simple response to (Q1).

This brings us to the Hackerian treatment of (Q2), of the epistemic status of privileged utterances on an expressivistic treatment. Now, Hacker says *a lot* in his 2005 article about all the ways "I'm in pain" is a singular sort of sentence, and in such ways as to make "I know" an inappropriate operator with that sentence as its base. But, it seems to me, everything he says toward that end essentially condenses to this drop of grammar: "There is no room for ignorance, and hence nothing for 'I know' to *exclude*." (Hacker, 2005, p. 267) This will be recognized as the familiar Wittgensteinian type (2*) insight about nonsensical terms and their antitheses. I'm not going to go into the minutiae Hacker goes into, and I'm just going to leave the argument at that. The interesting thing is not so much why Hacker thinks "I know" can't sensibly be prefixed to "I'm in pain", as it is the way he handles the technical difficulties that attend this view.[74]

The technical difficulties, remember, all had to do with the fact that if we give up talking of knowledge of one's own pain, then it seems we'll have to give up a lot else besides. The specific examples Hacker treats in the 2005 article are these: it would seem we can't lie about being in pain, we can't remember being in pain, we can't act for the reason that we're in pain, we can't know that someone else knows we're in pain, and we can't know that everybody in the room is in pain when we're in the room. The reasons these will be problematic are clear. How can one lie about *p* if she doesn't know (or at least believe) what the truth is about whether *p*? How can one remember something one never knew? How can one's action be for the reason

[74] Hacker explicitly restricts his discussion to the case of pain, and I'll follow him in this for the purposes of setting out his view; but it should be clear that his treatment of the difficulties attending the simple response to (Q2) will be more widely applicable.

that *p* when one doesn't know (or at least believe) that *p*? How can one know that someone else knows that *p* without knowing herself that *p*? And how can one know of everybody in the room that they are in *M*, while she is herself in the room, and fail to know of herself that she's in *M*? These would seem very straightforward and very difficult problems indeed, but Hacker's way with them all is penetrating and simple.

As Hacker sees things, these will seem particularly intractable problems:

> …only if one cleaves to a certain *form* of analysis in disregard of the use of the relevant sentences. It is, to be sure, very tempting for philosophers to seek for uniform and elegant analyses of concepts. But uniformity and elegance are worthless if distortion of our concepts ensues -- given that what we are doing is clarifying our existing conceptual structures. (Hacker, 2005, p. 271)

The way he shakes off the distorting forms of analysis is indeed impressively simple. About being able to remember that one was in pain yesterday, he writes:

> [I]t seems to follow that I now know that I was in pain yesterday, that I knew yesterday that I was in pain, and that I know now because I knew yesterday…[But] [a]ll that follows from my remembering that I was in pain is that I was previously in pain, that I now know that I was in pain, and that I know now because I was previously in pain. So memory cannot be given this elegant, simple and all-encompassing analysis. (ibid.)

The rest get similar treatments. One lies about one's pain simply when one is in pain and professes otherwise with the intent to deceive. One knows that everyone in the same room as oneself is in pain just when one is in pain, and when one knows of the others in the room that they are in pain. One knows another knows about one's own pain simply when one is in pain and knows the other knows it. And all that acting for the reason that *p* entails, when *p* is "I'm in pain", is that one is not *ignorant* of the fact

that *p*.[75] As Hacker puts it:

> There is, as it were, a *singularity* in the epistemology of the psychological, at the point of the first-person present tense of certain verbs and phrases -- just as there is a singularity in arithmetic at the number 0. (ibid.)

As far as I can tell, Hacker has given with this line the definitive expressivist response to an important set of would-be technical problems attending the grammatical line that knowledge and belief are excluded, because doubt, mistake, and insufficient justification are. And so, if one wants to argue against the simple line on the question of knowledge (and belief), then, one will either have to deny that ignorance is indeed ruled out in such cases, or deny that ruling out ignorance indeed rules out knowledge. In any case, one had better argue thus, because if it's granted both that ignorance is ruled out and that this rules out knowledge, the technical difficulties one might've thought were themselves decisive against following through on this are, as Hacker has shown, quite without sting. Now, I tried to argue in chapter one that, despite arguments to the contrary, ignorance seems as ruled out as ever when it comes to (at least certain of) one's own mental states. And, as I've said already, I don't see that there's any argument to be given that ruling out ignorance is *eo ipso* ruling out knowledge. I don't want to dwell on this any more now, though, since I'll return to the matter in the next chapter. So with that, I'll leave off with the discussion of Hacker, turning next to Dorit Bar-On's expressivism.[76]

---

[75] What Hacker probably means, if it were spelled out, would be that one's acting for the reason that *p* entails either that one *happens* not to be ignorant *or* that ignorance is *grammatically* ruled out. It's surprising, actually, that he pulls these both under the simple 'not being ignorant that *p*'.

[76] It was promised in chapter one to discuss how the expressivist would teach us to live without substantial self-knowledge. The discussion in this section, and then again in the next chapter, of Hacker's defense of the conservative response to (Q2) is meant, among other things, to make good on that promise.

Bar-On's Expressivism

I'm going to have a look now at Dorit Bar-On's expressivist view.[77] Hers is without a doubt the most thoroughly worked out expressivist view of first-person privilege on offer. Accordingly, in *Speaking My Mind* she sets out an octuplet of *desiderata* her expressivist view aims for, the most notable of which, for the purposes of this discussion, are the following: an expressivist account should (a) "present avowals as truth-assessable", (b) account for "epistemic asymmetry in its full scope", and (c) "allow for the possibility that avowals represent privileged self-knowledge". (Bar-On, 2004, p. 20). (b) actually breaks into four separate *desiderata*, which, for our purposes, can usefully be brought in as one. It's worth noting, however, that the "*full scope* of epistemic asymmetry" is ambiguous between, on the one hand, the whole range of *epistemic asymmetries* associated with avowals, and, on the other hand, the whole range of *avowals* distinguished by those epistemic asymmetries. The way (b) breaks up in her list of desiderata, it's clear that Bar-On intends the first disambiguation; it's clear from other things she says that she intends to capture the "full scope" in the second sense too. Now, it goes more or less without saying that expressivists are especially well placed to account for the "full scope of epistemic asymmetries" of whatever will *in fact* yield to expressivist treatment, the "full scope", that is, in the first sense; and so I'm not going to discuss Bar-On's argument to the effect that her expressivism is indeed so placed. I will, though, discuss her contention that expressivists can account for the "full scope" in the second sense; that is, I'll

[77] She calls her view Neo-Expressivism, but I'm going to refer to it simply as an expressivist view. Given that, as we've seen, an expressivist treatment of privilege (for at least some avowals) needn't deny truth-aptness to them, it will perhaps come to seem surprising that neither of Finkelstein or Hacker ever calls his view (or any part of it) an expressivist view.

discuss her contention that the whole range of privileged first-person mental state ascriptions *in fact* yields to expressivist treatment. This will be my discussion of the way she handles (Q3) and (Q4). First, however, I'll discuss briefly the way she handles my (Q1), and then I'll come to (Q3) and (Q4). I'll discuss her response to (Q2) last, since it will be of particular interest.

As will be clear from the first *desideratum* I've listed, Bar-On's view, like every view we've discussed besides the simple view (and the view of an erstwhile time-slice of Peter Hacker), admits truth-aptness to privileged (because expressive) first-person mental state attributions. The key distinction for Bar-On in this regard is between avowals as *acts* and avowals as *products*. (Bar-On, 2004, p. 251) *Acts* of avowing are taken to express one's 'first-order' mental states, while the *products* of avowing are taken to express truth-evaluable propositions.[78]

It's hard to say whether this distinction is significantly different from Finkelstein's distinction between expressive and assertoric dimensions. It's just as hard to say whether Bar-On takes the Augustinian view of the products of avowing. On the one hand, she says that her expressivism will invoke a "commonsense conception of mental states" which she sees as "neither Cartesian nor materialist *nor* behaviorist," but which is nonetheless such that expressive behavior is "not understood as merely symptomatic of the 'real' and hidden mental conditions; rather it is behavior in which, we take it, the conditions themselves can be perceived by

[78] These are obviously two different senses of "expression", and Bar-On has something to say about this. She, in fact, sets out three senses of "expression", two of which correspond to the acts of avowing and their products. For her three senses of "expression", see (Bar-On, 2004, p. 216), where she credits the threefold distinction to (Sellars, 1969, pp. 520, 1).

observers." [79] (Bar-On, 2004, pp. 423, 4) On the other hand, she writes, "I myself would welcome a reading of the Neo-Expressivist account which renders it consistent with materialist ontology." [80] (Bar-On, 2004, p. 417) Furthermore, she never explicitly addresses the question of the connection between the expressiveness of avowals-as-acts and the contentfullness of avowals-as-products. I won't try to get to the bottom of her view about this, though, since all that needs pointing out is that expressivists admitting truth-values to privileged utterances but taking other than a deflationary view with respect to "true" and "false", and so ruling out the possibility of 'fusing' together the expressive with the assertoric aspects of avowing, run precisely the risks already mentioned in the section of Finkelstein's expressivism. At any rate, Bar-On allows for the truth-aptness of privileged utterances to be handled by the avowal-as-product side of things, while letting the avowal-as-act side of things handle the story of those utterances' privilege. On (Q1), then, Bar-On's answer is, Yes.

On to (Q3) and (Q4), then. Bar-On, the only of my (not invented) expressivists to address these questions, takes more or less the same line as my invented, simple expressivist from chapter two. The simple line on these, remember, was that there was no need to relate the privilege of the non-expressive first-person psychological utterances to the expressive ones, because there was no such thing as a privileged non-expressive first-person psychological utterance. While the simple line didn't make mention of degrees of privilege that might map onto degrees of removal from the properly basically expressive, it seems clear to me that "privileged only

[79] Very Wittgensteinian indeed.

[80] Not so Wittgensteinian.

insofar as expressive" is just another way of saying "privileged only if expressive".

And what Bar-On says is "privileged only insofar as expressive":

> I will be suggesting that *insofar as* we think of all avowals as enjoying a special security, a security that goes beyond the epistemic security of well-grounded or highly reliable self-reports, it is *because, or to the extent that*, we regard them as (at least in part) serving to express the very conditions they ascribe. [81] (Bar-On, 2004, p. 264)

She indeed goes on to suggest just that:

> Having mastered the use of "I want the teddy" as a linguistically articulate expressive device, Jenny may put that device to partially reportive use. For instance, she may offer that kind of self-ascription in answer to such questions as "Why are you looking in that drawer?" or "What do you want most right now?". In such cases, the self-ascriptive utterance may seem to shade readily into ordinary reports. If it is still regarded as more secure than such reports, we may wonder why. I suggest that, if we regard non-evidential reportive avowals as more secure than other reports…this is still because, or *to the extent that*, we regard them as directly expressive of the self-ascribed state. (Bar-On, 2004, p. 301)

Fortunately for Bar-On's Neo-Expressivist view, merely trying to *tell* someone else of one's own mental state doesn't by itself take one away from the security of avowing proper to the less secure because less expressive. Indeed, for Bar-On it seems that telling someone else of one's own mental state is something that can go *right along side* expressing that state in words, not detracting at all from the "reportive avowal's" expressiveness (security):

> Given the context of the request ["Why are you looking in that drawer?"], it is not unreasonable to see [Jenny] as informing us about how things are with her mentally. But I submit that, insofar as we take Jenny to be avowing, we take it that she is **also** expressing a state she is in, and not **merely** presenting her findings about a state inside her…The presence of [a] communicative intention to inform of the condition does not by itself turn the self-ascription into a descriptive report. After all, the communicative intention could also accompany a non-verbal expression, such as an enthusiastic hug, yet the hug is not even a candidate for being a report, since it

[81] Note that *all* avowal are specifically in question here, as opposed to Hacker's similar point about avowals belonging to the "primitive language games".

> does not semantically express anything, let alone a self-judgment.[82] (Bar-On, 2004, pp. 302-4; my boldface.)

But then something is curious, because then it's not exactly clear what's going on with Bar-On's "to the extent that" claim. In the case of the enthusiastic hug, I take it that the communicative intention isn't supposed to detract one bit from the act's expressive character -- the point being, I take it, that communicative intention doesn't detract from the expressiveness and therefore the security of a "reportive avowal". So, on the one hand, it seems like intending-to-communicate and avowing-securely-because-expressively aren't two poles, somewhere between which self-ascriptions are supposed to settle. It seems like intending-to-communicate needn't be in competition with avowing-securely-because-expressively; they may both be operating, as it were, side-by-side. But on the other hand, in connection with Jenny's answering questions about what she wants, Bar-On is intent to point out that avowals are secure only to the extent that they are expressive (Bar-On, 2004, p. 302), making it seem as though Jenny's communicative intention would tend to compromise the extent to which her avowal was secure-because-expressive. So, are intending-to-communicate and avowing-securely-because-expressively in competition or aren't they? If so, then Bar-On runs into the problem that, on her to-the-extent-that view, almost all of what we take to be uttered with privilege is actually not as privileged as we'd thought. And if not, we're left wondering what *would* leave a self-report only *somewhat* expressive, and so only *somewhat* secure?

Actually, if it's not communicative intention compromising the extent to which an avowal is privileged-because-expressive, then I don't see what could fill

---

[82] It seems a "descriptive report" isn't the same thing as a "reportive avowal".

that role for Bar-On. One might suppose, for instance, that perhaps one moved to the less expressive as one moved to avowals about mental states for which there are no pre-linguistic, natural expressions. But that doesn't seem to be Bar-On's view. About a child's learning to avow the 'sensation' of smelling vanilla, for example, Bar-On writes:

> [I]t suffices if our child is in a position to offer a verbal characterization of the experience without naming her experience. We can imagine the child sniffing and saying, "Hmmm, vanilla!" Taking that to be the verbal expression of the child's olfactory experience, we may respond by saying, "Ah! You're smelling vanilla. That's right, Daddy is baking cookies." The child's subsequent articulate self-ascription, "Dad! I can smell vanilla!", can then be seen as expressing what had earlier received a non-self-ascriptive (linguistic) expression. (Bar-On, pp. 294, 5)

That the child's "subsequent articulate self-ascription" can be seen as expressing precisely "what had earlier received a non-self-ascriptive (linguistic) expression" indicates to me that, on Bar-On's view, this isn't the way to bargain out any expressiveness (security) from an avowal. Even Hacker's counter-examples from chapter two, "I see red", "I remember my third birthday", "I am thinking of Gödel's theorem" and "I believe that it is Tuesday today", would likely be given such a treatment by Bar-On.[83] And even Hacker, in 2005, ends up, after all, giving his stamp of approval to a treatment along much the same lines:

> Of course, both utterances such as 'I am in pain' and utterances of the form 'I think (expect, suspect, etc.) that *p*' *also have a use as statements or reports*. 'I think that *p*' is typically an avowal or expression of belief or opinion, but it can also be used as an autobiographical admission, confession or statement. However, the first-person statement that things are thus-and-so with me shares many of the logico-grammatical expressive features of the more primitive utterance from which it grows. So, for example, my statement

[83] About these Hacker said in the first edition of *Insight and Illusion*, p. 268, that they can't "by any stretch of the imagination, be thought of as having the logical status of outcries and facial expressions. Yet such sentences do share the central perplexing features of 'I am in pain'."

> that I believe that *p* is nevertheless still an expression of my belief that *p*, in as much as in stating that I so believe, I am still endorsing the proposition that *p* (which is why I cannot say 'I believe that *p*, but actually it is not the case that *p*' or 'I believe that *p*, but whether it is the case that *p* is an open question as far as I am concerned'). [84] (Hacker, 2005, p. 247)

Hacker's remark here that such reports share "many of the logico-grammatical expressive features of the more primitive utterances" from which they "grow" is just what I take Bar-On to be getting at with her story about learning to avow fine-grained olfactory experiences having no direct antecedent in non-linguistic behavior. So what is it, then, that would compromise the extent to which an avowal is privileged-because-expressive? Because we have on the one hand that Bar-On won't try to relate the privilege of far removed from expressive cases to thoroughly expressive cases, there being no such privilege. But on the other hand, it seems Bar-On is going to count a whole lot as thoroughly expressive.[85]

Whatever the resolution of that may be, there is anyway a potential fly in the ointment for Bar-On's expressivism, a seemingly quite privileged kind of first-person utterance that is, seemingly, quite unyielding to expressivist treatment. It's something I'm going to return to in the last chapter, but which I've already shown, back in chapter two, that Hacker was on to with his 1975 criticism of the expressivist thesis. Interestingly, Hacker seems to have forgotten this criticism between editions of *Insight and Illusion*, never to return to it himself. Also interestingly, it's something

---

[84] Since "I see red" and "I remember my third birthday" are, if privileged, highly plausibly construed as "I seem to see red" and "I seem to remember…", similar treatments of these are easily arrived at. Seeming to see red and seeming to remember something both have more nearly primitive, linguistic expressions in "Red!" and "It was…".

[85] Again, Bar-On does distinguish between "avowals proper" and "reportive avowals", the former being "spontaneous" and the latter being "communicative". It's just that, as I've been saying, the *expressiveness* crucial to explaining authority seems to be every bit as in force in the communicative as in the spontaneous case.

neither Wittgenstein nor Finkelstein address in their work. But Bar-On does, and the line she takes is quite surprising. When Hacker brought up "I see red", "I remember my third birthday" and the rest, it wasn't the assertion of just *these* that didn't "in any sense replace a primitive natural form of behavior", it was the "assertion of such sentences (*and of their negations*)". (Hacker, 1975, p. 268, my italics) What about "I *don't* see red", "I'm *not* in pain", "I *don't* want the Teddy"?[86] Aren't these every bit as privileged as *their* negations? Is there any smell-of-vanilla story we can tell about these? Can we have communicative intention running along side expressiveness when it comes to these? It certainly wouldn't seem so. And neither does Bar-On say so. What she does say is striking:

> So, for example, the subject is asked, "Do you believe *p*?", and her answer "No, I don't believe *p*" has to be understood neither as an affirmation of not-*p*, nor as expressing hesitation or uncertainty regarding *p*, nor yet as expressing the belief that either *p* or not-*p* is the case…I suspect that actual cases fitting this bill will not be easy to find. But if we find them, I see no compelling intuitive reasons to regard them as avowals…Such a self-report would *not* seem to share in the special security of avowals. It would seem open to doubt, questioning, requests for reasons/explanation etc…there is no reason for [the Neo-Expressivist account] to recognize the existence of genuinely negative avowals. (Bar-On, 2004, p. 335)

Bar-On exceedingly bravely concedes here that self-reports running "I don't have a headache anymore" *aren't* the privileged bunch their positive counterparts certainly seem to be. I won't say anything more about this now than that I think most will agree this is a quite a counterintuitive result for Bar-On's expressivism, and that perhaps it spells imminent doom for the Neo-Expressivist. I'll defer more comments on this point to the next chapter, wherein I assess the expressivist's chances in light of

[86] Where "I don't want the Teddy" isn't supposed to amount to "I *do* want the Teddy removed", which construal, actually, would capture its usual role in the mouths of most young children, but which role, obviously, is not the only one it plays.

everything that will have come before. For now, though, it can simply be pointed out that certain difficulties with Bar-On's privileged-to-the-extent-that-expressive line can perhaps be mitigated by counting, as she seems to, very much to be thoroughly expressive *while* communicative; but the privileged-to-the-extent-that-expressive line would seem to founder on the question of so-called negative avowals. It's also worth remarking once again that Bar-On's, like every other expressivist's account on offer, doesn't in any way try to relate the privilege of a class of statements *not* thoroughly (or even at all) expressive to the privilege of a class of thoroughly expressive first-person psychological statements. Her account simply denies any such privilege.

The last aspect of her expressivism I want to consider corresponds to the third of her *desiderata* set out above, her answer to (Q2) that Neo-Expressivism will "allow for the possibility that avowals represent privileged self-knowledge". Things get interesting here, and before I go on to set out Bar-On's view I want to make a few preliminary remarks. The first is that Bar-On separates, as all expressivists do, in some way or another, the question of first-person privilege (whence authority, transparency, and groundlessness?) from the question of privileged *self-knowledge* (whence privileged self-knowledge?). The expressivist project just is to account for (shed light on) the phenomena of privilege without appealing to any special first-person epistemic access; that is, without appealing to privileged self-knowledge. Now, as we've seen, an expressivist satisfied with his expressivistic account of authority tends to think the question of privileged self-knowledge won't get anything like a standard treatment in terms of epistemic justification. Bar-On seems to think that, on the contrary, we still need to show how one could arrive at privileged self-

knowledge and a kind of epistemic justification peculiar to it, even after an expressivist account of authority is satisfactorily in place. That is, Bar-On thinks we can still give an epistemic account of an extraordinarily secure justification accruing to self-ascriptive beliefs expressed in avowals (and so, an epistemic account of self-knowledge as privileged as that justification is extraordinary), even after giving an expressly *non*-epistemic answer to the question of the phenomena of privilege.

But what is this going to look like? As Matthew Chrisman points out, it would be strange if whatever explained privileged self-knowledge weren't exactly what explained first-person authority (Chrisman, 2009, p.8), so that if one of these explanations is inherently epistemic, the other will be too. If Bar-On can give a truly epistemic explanation of privileged self-knowledge, then she's in the danger that her non-epistemic account of special first-person authority simply won't be needed, given her epistemic account of privileged self-knowledge. Privileged self-knowledge would seem already to account for the phenomena of privilege; and so any expressivist account will simply be superfluous.

Now, Bar-On tries to stall this concern by stressing that her (suggested) account of privileged self-knowledge "makes direct use of key ingredients of the Neo-Expressivist account of avowals' security". Of course, just saying this in abstraction from saying what exactly those "key ingredients" are, wouldn't itself be guaranteeing that the Neo-Expressivist account of authority isn't superfluous, just that "key ingredients" of it aren't. But let's see what exactly her account of privileged self-knowledge is, and then let's judge whether her Neo-Expressivist account of authority really is rendered superfluous. My argument is going to be that, if there's a

way to look at her account of privilege so that, indeed, her account of authority isn't therewith rendered superfluous, then her reaction to (Q2) shares an important structural similarity to the reactions I've laid out coming from the other expressivists. That is, her answer to (Q2) either begins to look not so radically different from the rest, or else it begins to make her non-epistemic account of authority unnecessary.

We need to begin, though, by understanding her concession that avowals express self-directed beliefs *as well as* 'first-order' mental states. Bar-On's "Dual Expression" thesis is introduced thus:

> I see no general reason to maintain that one cannot express more than one state with a bit of behavior. Intuitively, one can express both one's fear and one's loathing with a single contortion of one's face. By the same token, one could in principle express both one's hope and one's judgment that one is hoping in a verbal (or mental) act…Let us dub the idea that, when I avow, unlike when I engage in naturally expressive behavior, I express not only the avowed condition but also my judgment that I am in that condition "the Dual Expression thesis".[87] (Bar-On, 2004, p. 307)

She continues, "Should this thesis be embraced or rejected?", and her answer is that it should be accepted for three reasons. The two most relevant to our discussion are, first, that otherwise we couldn't allow for privileged self-knowledge, and second, that "it is commonplace to think of the sincere utterance of an indicative sentence *p* as expressing the utterer's judgment or belief that *p*". (Bar-On, 2004, p. 308) Let's start with how she aims to accept and work with that commonplace and move to her treatment of privileged self-knowledge.

In the way of keeping with this commonplace, while keeping also with the expressivist view of things, Bar-On distinguishes two kinds of 'belief', an "opining" and a "holding-true" kind:

---

[87] I've dropped the subscripts on the occurrences of "expression" indicating its being the kind associated with avowal-as-act; I do so throughout.

> In what we may call the *opining* sense, one believes that *p* if one has entertained the thought that *p* and has formed the active judgment that *p* on some basis, where one has (and could offer) specific evidence or reasons for that judgment…But there is a second, more liberal, sense of belief, in which a subject believes that *p*, provided (roughly) that she would accept *p* upon considering it. This *holding-true* sense, as we may refer to it, is the one we apply when we say that people have beliefs concerning matters they have not yet even considered. (Bar-On, 2004, pp. 393, 4)

The difference between these two would seem just to consist in whether one has *actually* "formed the *active* judgment that *p*" or whether one *would* form such an active judgment. As Matthew Chrisman has pointed out, however, this wouldn't seem to the point: "When I avow 'I'm in pain,' it is not plausible to claim that I believe I am in pain but I have not yet considered whether I am in pain." (Chrisman, 2009, p. 7) Bar-On's response to this, though, shows that the point of her distinction isn't between actual versus counterfactual, but between having reasons and not. This comes out after she acknowledges just what Chrisman points out, and she says about the 'active' beliefs expressed in the issuing of an avowal that she:

> …would still want to deny that the mental self-judgments that are [thereby] expressed are formed on this or that basis, in which case it may still be inappropriate to regard them as beliefs in the opining sense, given our characterization above. (Bar-On, 2004, p. 366)

She suggests instead "a more liberal 'self-ascriptive' notion of belief" attaching to (actively issued) avowals:

> A subject may be said to express her self-belief simply in virtue of actively making the relevant self-ascription. *All there is* to having a self-belief in the case at hand, it might be held, is the (intentional) issuing of a self-ascription. (ibid.)

It becomes clear, then, that Bar-On's holdings-true are supposed to be understood, indeed, counterfactually, that is dispositionally; but it's what they are dispositions *for* that makes the difference, not simply that they're dispositional; and what they are

dispositions for is simply a kind of 'active believing' that *consists entirely in* issuing a self-ascription.

Chrisman, now, seizes on Bar-On's admission of 'active judgments', however thinly construed, as though that itself is going to bring down the expressivist project, which project should be meant, obviously, to oppose any kind of "introspectionism":

> [I]f we have active judgments and not mere holdings-true, there has to be some cognitive faculty or method by which they are formed…Bar-On's opponents call this "introspection", and so it looks like they at least have a name for what they are trying to explain. (Chrisman, 2009, pp. 7,8)

Now, Chrisman admits that he's working with a broad notion of "introspectionism" on which:

> [T]he introspectionist is just someone who thinks that there is some special cognitive faculty or method by which we come to have specially secure beliefs about our own mental states, beliefs that, when true, are articles of self-knowledge. (Chrisman, 2009, p. 8)

But I think his broad notion of "introspectionism" is indeed too broad, precisely because it seems to me insensitive to the thinness of the kind of 'belief' or 'judgment' that Bar-On is admitting in admitting that avowals express active self-judgments. Certainly, Chrisman wouldn't think one is adopting introspectionism just in admitting a "cognitive source" (whatever that means) to proclamations, in the course of a wedding ceremony, that one "does take this woman". But if we're working with Bar-On's super-thin notion of belief, whereby (actively) believing that *p* consists in nothing other than *saying* that *p* (here, upon being asked whether *p*), we'd have to count such ceremonial proclamations as indeed expressive of one's belief that one is, at the time of utterance, taking someone in holy matrimony. So I think Chrisman misses the mark here. As long as Bar-On maintains, as she does, that *all there is* to

believing that one is in pain, for instance, when one avows being in pain, is *simply* that one has avowed as much, then I just don't see that she's committed to anything one could call "introspectionism". Neither do I think anything rightly called "introspectionism" is invoked just in admitting that one believes, even when not issuing an avowal, that one is in pain, for instance, as long as that belief consists only in the disposition to say (avow) as much upon being asked.[88]

Therefore, I want to grant Bar-On her minimal sense of 'belief' and proceed to her view about how such self-beliefs may be justified, thus rendering them articles of genuine self-knowledge.[89]

Her (suggested) account of the epistemic grounding of self-ascriptive beliefs expressed in avowing is, importantly, austere:

> [T]he avowed sate itself, M, is also what provides the *epistemic reason* for my self-ascription, or what warrants me in avowing M. The state is not a justifier in the traditional sense, since it represents no epistemic effort on the subject's part. But the subject is still epistemically warranted--warranted simply through *being* in the state...[A]vowing subjects enjoy a special epistemic warrant, since their pronouncements…are epistemically grounded in the very states they ascribe to themselves, which states also serve as the reasons for their acts of avowing. (Bar-On, 2004, pp. 390, 405)

It needs to be pointed out immediately that I haven't made any mention of the way Bar-On thinks of the mental states that avowals express as "rational reasons" for those avowals *qua* expressive acts. It has to do with the fact that issuing an avowal is

---

[88] That is, if one wanted to answer truthfully, of course. Also, there are some puzzles, having to do with indexicality, about construing dispositional beliefs in terms of what one would offer on being asked (and being in the cooperative spirit). For instance, you might say everyone (probably) dispositionally believes she's being asked a question, since everyone would (probably) answer "Yes" to "Are you being asked a question?". But this is absurd, and the fix is that one dispositionally believes *p* if one would assent to "*p′*?", where *p′* is *p* with suitable adjustments made to tense indicators (*e.g.* "were" instead of "are").

[89] Bracketing Gettier-style concerns, which it would seem impossible to drum up in the cases at hand.

something (very often) within one's voluntary control (as opposed to the dilation of one's pupils, for example), but this isn't something crucial to the discussion of expressivism at the level I'm trying to conduct it in this dissertation.[90] So, for our purposes it suffices to reformulate Bar-On's account of the special epistemic relation (making for privileged self-knowledge), without this wrinkle thus: Avowing subjects enjoy special epistemic warrant because: the very same thing--one's being in M--is *expressed* by the avowal and grounds it epistemically[91], where one's avowal that one is in M is epistemically grounded *simply by one's being in M*.

This is highly curious. If the epistemic grounding of S's belief that S is in M *consists* simply in S's being in M, then why should an explanation of S's 'enjoying special epistemic warrant' in so believing have anything to do with anything besides just that, that she's in M? Specifically, Why should it have to do with additional facts about the expressive relation between M and any of S's actions (specifically S's action in avowing)? Bar-On says that "key ingredients" from the expressivist account of authority are to be drawn upon, and that is no doubt in reference to her insistence that what's special about this epistemic relation between M and the self-belief one expresses in avowing is that M *itself* is *also* expressed in avowing, which expression relation is indeed a key ingredient in the Neo-Expressivist account of authority. But what work is that doing for her?

In order to get where I want to go with this, I want to note that Bar-On herself discusses a worry that attaches to views, like the one she's putting forward, "that take

---

[90] Cf. (Bar-On, 2004, pp. 216f.).

[91] My "the very same thing…" construction is based on a sentence from (Bar-On, 2004, p 390).

the epistemic grounding of a belief to consist (at least in part) in the 'truth-maker' of the belief". (Bar-On, 2004, p. 391) The worry she discusses "has to do with the warrant of *false* beliefs".[92] (ibid.) She handles that worry to her satisfaction, and I'm not going press on anything in that connection. The problem I see with her discussion, though, is that there are other pressing worries with such views that are completely left out. The one I want to bring up is the following.[93] In general, one's warrant for believing that *p* won't simply consist in *p* itself (or in *p*'s obtaining), and it's hard to see that in any particular case that could happen either.[94] Richard Fumerton addresses this kind of view in connection with the contention that one's pain might itself be the justifier for one's belief that one is in pain:

> But what's the difference between my being in pain and Paris's being in France, the difference that makes it appropriate to identify the former as a justification for me to believe that I'm in pain, while the latter is no justification at all for my belief about a city in France? Following Russell and others, it seems to me that we should look for noninferential justification not in the truth maker (that fact that makes true the belief) *by itself*, but in a *relation* that the believer bears to the truth maker. (Fumerton, 2006, pp. 61, 2)

The difficulty this raises for Bar-On's view is plain: quite apart from the possibility of false beliefs, how could the truth-makers for your self-ascriptive beliefs *all by themselves* be the epistemic grounds for those beliefs? The example about Paris's being in France, as I see it, is perfectly to the point. The curious thing is, it seems

---

[92] It's interesting that Bar-On does allow for "false" avowals, because she allows for "expressive failures". The example she gives is of a person crying out with pain in anticipation of something painful. Cf. (Bar-On, 2004, pp. 320-35, 419-21). She takes this, obviously, as a strength of her view.

[93] Besides this one, the other main worry I have in mind is the general worry about bringing states of the world into the "space of reasons". Cf. (Bonjour, 1978) wherein he, rightly, credits the worry to (Sellars, 1997).

[94] I don't mean "always" by "in general", obviously, but I don't just mean "usually" either. What I mean by "in general" is clearer in the following use: "In general [(If *p* then *q*) and *q*] doesn't logically entail *p*, but in some particular cases it does."

Bar-On herself agrees, because, while on the one hand she says the epistemic ground for S's belief that S is in M consists simply in S's being in M, on the other hand she has it that S "enjoys a special epistemic warrant, since" *both* M is the "epistemic ground" of the belief *and* S expresses M. And just because Bar-On seems conflicted about the efficacy of this in-the-truth-maker-consisting "epistemic grounding", and just because she wants to supplement it with appeal to her expression relation in order to tell the whole story, it's looking like Bar-On is indeed "following Russell and others" by looking for noninferential justification in a *relation* that "the believer bears to the truth maker".

I think it's important that Bar-On is being pulled both ways here (in letting M itself be the epistemic ground, and then not letting that be the whole story of the justificational situation). The reason I think she finds herself thus conflicted is that, on the one hand, the expression relation seems to have nothing 'epistemic' about it whatsoever, and so the epistemic grounding will have to be located elsewhere (for her, simply in the truth-maker for the belief); and, on the other hand, that supremely weak kind of justification could only possibly seem less so if we immediately couple it with a reminder about one's self-ascriptive belief being issued only in the course of *expressing* the truth-maker, as it were. And this brings me to the main point I want to make about her view on privileged self-knowledge. Either (a) she can really commit to the justification lying simply with the truth-maker, making no mention of matters of expression, or else (b) she'll indeed need to rely on the expression relation in telling the story of "special epistemic warrant". If (a) then, first of all, she's in for an uphill battle given such a weak notion of justification, something it will be very hard

to argue applies only to certain mental state self-ascriptive beliefs. But also, whatever progress she makes in that regard will, as I see it, render her expressivist account of authority entirely unnecessary.[95] The special epistemic relation she will have devised will take care of any questions about authority. While, if (b), then it seems to me that what we have really is that the appearance of the extreme epistemic weakness of the relation between a self-ascriptive belief and its truth-maker is supposed to be *made up for*, or *explained away* by directing one's attention to the expressive relation between one's self-ascription and its truth-maker. But then, the only thing that lends any strength at all to the epistemic relation, could make it even begin to seem like something of an epistemically efficacious relation, would have to come from constantly reminding oneself that, after all, the very state (supposedly) grounding the self-ascriptive belief epistemically is *expressed* by the self-ascription. And so, just like with Finkelstein, the *prima facie* flimsiness of her response to the epistemic question can only possibly gain strength by seeing it in light of her expressivistic account of avowing. That is, at least, as I've been saying, if she doesn't want her expressivistic account of avowing with authority to be rendered superfluous.

I want to leave off there and pick up with this point in the next chapter, because, as I've said already, this is going to help to show up a deep structural similarity among all the expressivistic responses to the problem of self-knowledge we've covered, and so I want to retain that discussion for the next chapter wherein

[95] Consider the fact that Laurence Bonjour (Bonjour, 2001), certainly not an expressivist by any stretch of the imagination, tries to pull off very much the same trick, letting certain mental states (the 'conscious' ones) serve in and of themselves as the justifiers for one's claims to be in them. The tack is also very much redolent of Chisholm's proposal concerning "self-justifying" propositions; see (Chisholm, 1982, pp. 136-139).

stock is taken and prospects assessed. For the same reason, I won't bother with any kind of recapitulation or drawing of any moral in bringing this chapter to an end.

CHAPTER V
TAKING STOCK AND ASSESSING PROSPECTS

Introduction

In this final chapter I aim to take stock of what's come before and to assess the prospects for viable expressivist theories of first-person privilege. The structure will be simple. I'll first (a) discuss expressivist options with regard to the question of truth-aptness; then I'll (b) discuss expressivist options with regard to the question of self-knowledge; then I'll (c) discuss expressivist options with regard to the question of non-expressive while nonetheless privileged mental state self-ascriptions; then I'll (d) discuss expressivist options with regard to the possibility of, should such non-expressive while privileged cases be admitted, relating the privilege of the non-expressive cases to the privilege of the expressive cases in such a way that the resulting picture of things could warrant the label "expressivistic"; I'll then (e) very briefly tie up a handful of loose ends; and finally I'll (f) offer some remarks in the way of bringing to a close this investigation into the merits of expressivist theories of first-person privilege.

As a quick note, I want to remind the reader of a terminological point that will become important to keep in mind at various points along the way. A *conservative* expressivism with respect to some question is an expressivism that keeps in touch with the point of the expressivist insight by staying in step with the simple version of expressivism (set out in chapter two) on that question. A *progressive* expressivism is one that aims to keep in touch with the point of the expressivist insight while departing in some way or other from simple expressivism.

With those things in mind we're ready to begin the final movement of this investigation into the merits of expressivist theories of first-person privilege.

On the Question of Truth-Aptness

Each of the versions of expressivism about first-person privilege we've considered, except for the fictional simple expressivism set out in chapter two, admits truth-values to privileged, because expressive, mental state self-ascriptions. In this they seriously depart from expressivist (emotivist) views in ethics, which are concerned to handle certain difficulties (viz. difficulties to do with ontology and moral motivation) by strictly separating the "cognitive" from the "non-cognitive".[96] It was already remarked early on in chapter two, but it bears repeating that it was perhaps only due to a few ill-chosen remarks by Norman Malcolm[97] and to the familiarity of non-cognitive (truth-valueless) treatments of ethical discourse about the time of Wittgenstein's *Philosophical Investigations* that expressivist approaches to the problem of first-person privilege, pioneered by Wittgenstein in that work, seem to have been saddled until quite recently with the presumption that privileged because expressive mental state self-ascriptions ought to be devoid of truth-value. As we've seen, however, sophisticated expressivists are keen to renounce any such doctrine. Indeed, there doesn't seem any reason to think anyone working from the expressivist insight about first-person privilege ever denied truth-values to privileged because

[96] See (Stevenson, 1959), (Ayer, 1952) and (Blackburn, 1984) for classical non-cognitivist treatments of ethical discourse.

[97] E.g. that, according to Wittgenstein, avowals have the same "logical status" as "outcries and facial expressions" (Malcolm, 1966, p. 83)

expressive mental state self-ascriptions. [98] It's a good thing, too, because I don't see any hope for an expressivist theory of first-person privilege that doesn't accord truth and falsity to privileged because expressive utterances. The technical difficulties raised in chapter one are too much to handle, I think; thus I consider that option closed for the expressivist and in the following discussion it will remain to see whether any different difficulties arise for the truth-aptness-granting avowal expressivist.

The first thing to recall is that, having admitted truth into the picture, the so-called Augustinian conception of truth and truth-makers can be rather unfriendly to the expressivist project. For Wittgenstein, certainly, the struggle against the name-object grammar of sensation talk (and, presumably, of talk of a great deal of mental reality) seemed one and the same with the struggle against the inner-access conception of the processes issuing in privileged mental state self-ascriptions. Furthermore, the way he seemed to used it, it was precisely his likening of avowals to unlearned expressions of mentality that was to (at least help to) lay low the two-headed beast. On the one hand, he saw that we were tricked into thinking the linguistic practices in question "begin" with the appearance of a mental item that we

[98] A. J. Ayer says that the truth-valueless thesis of avowals "is alleged" by "philosophers" who go unnamed. (Ayer, 1963, p. 60) Bruce Aune says the thesis is "occasionally defended"; though by whom remains a mystery. (Aune, 1965, p. 36) James Tomberlin says that Wittgenstein was "sympathetic" to the doctrine. (Tomberlin, 1968, p. 91) Peter Hacker attributes to Norman Malcolm and Peter Strawson the attribution of the truth-valueless thesis to Wittgenstein. (Hacker, 1975, p. 258) Later he makes the attribution himself directly to Wittgenstein. (Hacker, 1986, p. 298) It's unclear whether Hacker was right about Malcolm or Strawson. (Cf. (Malcolm, 1966) and (Strawson, 1966)) Robert Fogelin attributes a non-assertoric thesis of avowals to Wittgenstein (Fogelin, 1987, p. 197), while John McDowell says that Wittgenstein "toys with" this thesis (McDowell, 1996, p. 22). Rockney Jacobsen even tries to reconcile Wittgenstein's deflationism about truth with a truth-valueless thesis he attributes to Wittgenstein. (Jacobsen, 1996). It seems, however, that besides those time-slices of Hacker that were defending their best reading of Wittgenstein, no philosopher has ever actually himself or herself defended the view that (even just certain) avowals were *merely* expressive, that is, without truth-value. It always seems to be other people who do this, and those other people all seem to be named "Wittgenstein", if ever they are named.

then proceed to "describe", when really they begin, so he says, with an *expression*. (Wittgenstein, 2009a, §290) And words like "toothache" (or sentences like "I've got this awful toothache"), just because they are (at least *sometimes*) "substitutes" for expressive behaviors like moaning, were seen to have grammars completely set apart from the grammars of words like "Watson" (or sentences like "Make sure Watson is there too"). (Wittgenstein, 1993, p. 298) On the other hand, he saw that it was just the fact that (at least certain) mental state self-ascriptions are "akin" to expressions that could help us to radically reconceive the immediacy with which they are often proffered. (Wittgenstein, 1980b, §63) And it was, as he saw things, our language-games of mental state self-ascription, bringing along as they do the *expression* of mentality in behavior, against which it becomes senseless to doubt or to be mistaken about whether, for instance, one is in pain. (Wittgenstein, 2009a, §288) These twin projects seem, as I've tried to bring out in a number of places along the way, inextricably intertwined for Wittgenstein.

We've also seen in the last chapter how Finkelstein, who seemed only explicitly to call on kinships with expressive behavior specifically in the way of making best sense of *our* warrant in taking somebody's *else's* word for it that she's happy, depressed, intending to bring it about that *p*, or what have you, we've seen how he too does well to have those kinships molding, as it were, the singular grammar of his truth-evaluable while authoritative, because expressive, mental state self-ascriptions. The reason was that it seemed he wouldn't have any hope of putting the question of self-knowledge to rest on the strength simply of his expressivistic account of our rightly taking people's words for it that they're in various mental states

if the expressive dimensions of avowals were, as it were, indifferent, ontologically speaking, to their truth-condition-setting dimensions, however much they could be expected to coincide. That is, if it weren't built into the very fabric of the fact, as it were, of a person's being in such-and-such a mental state that in avowing it she needn't *ascertain* anything but simply *express* the state, then no matter what expression-heeding account we could give of our rightly passing from her avowal to our own third-person judgment about her state of mind, we'd be left with the as yet unaddressed question of the epistemic credentials of her own truth-condition-setting claim.

I take it, actually, that these are not just peculiarities of Wittgenstein's and Finkelstein's expressivist hopes in particular, that the first self-consciously joined (certain) mental state self-ascriptions' expressive character with the very ontology of the mental reality they self-ascribe and that the second had better do the same; I take it that any successful expressivist theory of first-person privilege granting truth-aptitude to privileged because expressive self-ascriptions will follow suit. The general pressure seems to be this, once it's granted that a person is indeed making a truth-evaluable claim in avowing some mental state, the question immediately rushes in what *basis* the person has for making that claim. The expressivist answer, it seems to me, will have to be that the claim is made on *no basis at all* and that it's the *expressive* character of the claim that makes for this, perhaps, surprising absence. But the expressivist will have to say more than simply that such a claim is in fact without any basis because issued expressively; she'll have to say how anyone could *legitimately* make a genuine *claim* about how the world is absent any basis; and in

doing so, she'll want to rely, yet again, squarely on the expressive character of the claim; otherwise it will be hard to see how attention to the expressive character of such claims is supposed to be doing the real work in the expressivist's account of first-person privilege. But it seems to me she can only do this, she can only rely on the expressive character of avowals in the way of explaining their being genuine setters of truth-conditions absent any epistemic basis and this while being quite beyond epistemic reproach, if she casts those very truth-conditions, or let's say the very *meanings* of those claims, as inherently cut out for precisely that purpose. Otherwise, it will be hard to see how the *expressive* element of an act of avowing should itself account for that act's being a legitimate, baseless *claim*.

Putting it slightly differently, it's only when the expressivist has it that to make a claim in avowing, to *represent reality* with an avowal, just *is* to express the reality represented, only then can she have it that the expressiveness of an avowal handles all of our questions about its features as a claim, as a genuine representation of reality. But to do that, now, she'll seriously have to deflate the notion of *representing* reality, at least in the case of issuing an avowal; and then in the moment she does *that*, she also seriously deflates the notion of a *reality* thus represented.

This connection between deflating notions of truth, representation, etc., and therewith deflating the notion of a corresponding reality itself has not gone unnoticed since Wittgenstein, of course; in particular it's been the theme of much of Crispin Wright's work on how best to conceive realism/anti-realism disputes. His basic thought has been that, for any particular discourse satisfying certain minimal requirements (viz. bearing the syntactical hallmarks of assertion, and being subject to

norms of warranted assertion), we should retain but *deflate* the notion of 'representing reality' ('being true or false', 'referring to objects', *etc*.) appropriate to that discourse, and then we can go on to see whatever more "robust" content there is to be had by that (family of) notion(s) in application to that discourse as something to be *added*, as it were, *after the fact* of that discourse's sentences meriting such nominal honors as (aiming at) 'corresponding to reality', etc.[99] These additional factors, for Wright, go to determine whether we should adopt "a realist perspective on the discourse concerned". (Wright, 1992, p. 78) Specifically, for Wright, the way one *adds to*, makes more "robust", the notions of truth, correspondence, etc., as applied to a given discourse, is to locate those notions, for those discourses, on the right side, the robust side, of certain contrasts disconnected from the question of truth-aptness itself.[100] (Most notably, these contrasts include the Euthyphro contrast and the Cognitive Command contrast.[101])

Now, Wright makes it clear that he aims with this framework to capture Wittgenstein's insistence on *differences* between discourses (Wright, 2003a, p. vii),

---

[99] Wright's ideas in this connection were introduced in book form in (Wright, 1992), wherein (as elsewhere) he was anxious to distinguish his *minimalism* from standard forms of *deflationism* (for instance, that propounded in (Horwich, 1998, ch. 4)) on the basis of a subtle, and dubious (cf. (Miller, 2001)), argument to the effect that, *contra* standard deflationary views, 'truth' is a real property (however "non-robust" it may be as applied to this or that discourse). That argument notwithstanding, I'll here style his minimalist view a 'deflationary' one, keeping that peculiarity in mind.

[100] Thus, one of the main features of Wright's program for realist/anti-realist disputes is that anti-realism by way of non-cognitivism is a non-starter.

[101] A discourse's notion of truth (and the rest) falls on the 'realist' side of the Euthyphro contrast if it happens that one's best judgments within that discourse *track* (as opposed to *constitute*) truth in that discourse. It will fall on the 'realist' side of the Cognitive Command contrast if differences in individuals' different judgments within that discourse are bound to betray "cognitive shortcomings". See (Wright, 1992) and (Wright, 2003a) for further discussions of these and other of his realism/anti-realism deciding contrasts.

and there are no doubt deep affinities between his project and Wittgenstein's own.[102] For present purposes, the important Wittgensteinian thought I see at work in Wright's deflationary view, and what I see must have a place in an expressivist's account of first-person privilege, is precisely that just in admitting that the notions of truth, representation, etc., indeed apply straightforwardly to a given sentence (or discourse), we don't yet say *anything* of importance about what those notions *amount to* for that particular sentence (or discourse). In other words, an acquiescence in the name-object grammar of a given discourse needn't be automatically, as it were, brought along with an acquiescence in the applicability to that discourse of the notions of truth and its cohort.

Notably, Wright never explicitly considers, the way Wittgenstein does, attending specifically to the ways words (and sentences) are *taught* as in itself helpful in specifying the peculiarities of those words' own attendant notions of reference (and truth, etc.). (I think it isn't too much of an exaggeration to say that, even in such camps as Wright's deflationist camp, such Wittgensteinian attention to the ways one *learns* one's language (except for concerns about continuing a causal/referential chain, for instance) simply don't come into question for those who would ponder the here-and-now workings of the sentences one eventually does come to learn to use, who knows how.[103]) But if the expressivist about first-person privilege in particular

---

[102] There are certainly, possibly just as deep, differences too. For one, Wittgenstein, often thought of as a *quietist* about philosophical disputes, probably wouldn't have cast the differences between discourses as differences between discourses about what is more or less 'real', as Wright's "realism/anti-realism" terminology is meant to suggest. This point about Wittgensteinianism as quietism becomes relevant in the main body of this discussion.

[103] Quine, for instance, while he thought it was important to see in general how one makes one's way from baby talk to adult speak, didn't think that any particular route one should take would make any difference. He saw that "[b]eneath the uniformity that unites us in communication there is a chaotic personal diversity of connections, and, for each of us, the connections continue to evolve." (Quine,

is going to be able to pull off what I've been saying she needs to, viz. building avowals' expressive character into their assertoric (claim-making) character, it seems to me this is just what she'll have to do.

That is, the best chance, as I see it, the expressivist has of being able point to an avowal *qua expression* in the way of answering questions about the avowal *qua claim* is indeed to retain but deflate all talk of avowals' being claims, setting truth-conditions, having truth-makers, corresponding to reality, etc.; and then in order to see the right kind of life, as it were, added back into those notions as applied to avowals *qua* claims, the expressivist needs to tell a particular, crucially-expression-involving, story specifically about the ways we *teach* language learners to make avowals. That is, the expressivist, it seems to me, desperately needs to follow up on Wittgenstein's suggestion about the "one possibility" for setting up the connection between words and sensations; viz. that "words are connected with the primitive, natural, expressions of sensation and used in their place."[104] (Wittgenstein, 2009a, §244)

If such a story can really be told, then given a Wittgensteinian, anti-Augustinian conception of what 'referring', 'claiming', 'setting truth-conditions',

---

1960, p. 13) And it's this kind of view of the ultimate irrelevance of one's personal instruction in language that lies behind an important strand in his much lauded criticism of the linguistic doctrine of necessity, namely that "[l]egislative postulation contributes truths which become integral to the corpus of truths; the artificiality of their origin does not linger as a localized quality, but suffuses the corpus." (Quine, 1976, p. 119, 20) Of course, this view of things brings along with it that, strictly speaking, nobody "finish[es] learning [his language] while he lives" (Quine, 1960, p. 13) where 'learning one's language' comes to *nothing more than* revising one's beliefs.

[104] Pointing out simply that avowals are *criterionless*, for example, won't do unless they're criterionless *because* expressive; otherwise, there will be nothing to favor expressivism as a theory of first-person privilege over constitutivism. But, it seems to me, the project of showing how avowals' criterionlessness could be bound up with their expressiveness will itself turn to just these behavior-replacement language-learning considerations.

etc., really amount to in any given case, the expressivist will indeed be on the way toward securing the necessary connection between avowals as expressions and avowals as claims.[105] Of course, spelling out that replacement story and making a convincing case for the anti-Augustinian presumption are two rather difficult tasks. We'll return to the question of the first difficulty when we consider expressivist responses to (Q3). But as for the second, it can be pointed out that, even besides the fact that approaches to ontological questions *via* deflationism (minimalism) about truth, reference, and the rest don't enjoy anything like widespread acceptance, Wright himself has argued convincingly that that path (indeed, any path) to anti-realism about the psychological in particular is self-undermining. The crux of his argument, given our purposes, is that anti-realism about the psychological enforces anti-realism about the realism/anti-realism distinction itself, and that it's impossible to have a philosophical (that is, for Wright, an *a priori*) *argument* for one's claim that *p*, where the facts about *p* are to be construed anti-realistically.[106] On the other hand, insofar as Wittgenstein's anti-Augustinian sentiments would be *distinguished* from those informing Wright's project of sorting out the realist discourses from the anti-realist discourses, that is, insofar as Wittgenstein's not-a-something-but-not-a-nothing-either treatment of sensations isn't meant in any way as a compromise on their *reality*, then it seems a much loathed *quietism* about the significance of realist/anti-realist disputes

[105] Bar-On does probably the most of anyone toward spelling out such a story, though not explicitly with any anti-Augustinian concerns in mind. Cf. (Bar-On, 2004, pp. 285f.).

[106] See (Wright, 2003c) for the full argument. Paul Boghossian too argues against "irrealism" about intentional content, and in largely overlapping ways with Wright's argument; notably Boghossian's argument also brings in its wake an argument against any deflationary theory of truth which, unlike Wright's, denies that truth is a genuine property (cf. (Boghossian, 1990)). I won't argue against Wright's argument here, as it seems to me quite sound.

in philosophy is sure to be ushered in.[107] This bodes ill for the expressivists. Ill enough, I think, that we can stop there setting out what we need to in order to assess the prospects for expressivist views vis-à-vis their available responses to the question of truth-aptness.

Summing up, the main points have been these: (a) Expressivists had better admit truth-values to privileged because expressive utterances (as, it seems, every actual philosopher working from the expressivist insight indeed does), because the difficulties attending their denial seem simply insurmountable. (b) Expressivists had better admit truth-values to avowals in a way such that the attendant notions of an avowal's qualities such as *representing reality*, etc., are notions of qualities 'fused together' with an avowal's *expressive* qualities; otherwise it will be very hard to see how our attention to an avowal *qua* expression will serve to address our questions (principally epistemological) about the avowal *qua* reality-representing claim. (c) This, then, reaches through the, now, as it were, thinned-out notion of *representation* to the very notion of a *reality* represented, which now, it seems, will itself thin-out accordingly. (d) But a thus, as it were, thinned-out psychological reality will either be conceived in *opposition* to realities (and attendant notions of truth, etc.,) *not* thus thinned-out; or else psychological discourse, even though freed from problematic ontological commitments, will still be conceived, so to speak, on a par, ontologically speaking, with every other (non-problematically ontologically committed) discourse.

---

[107] See (Wright, 2003b) and (Blackburn, 1998) for compelling considerations against quietism about traditional, realist/anti-realist, philosophical disputes. Notice also that this problem about being forced into quietism isn't the above mentioned problem about being forced into anti-realism about realism/anti-realism questions themselves. *That* problem can be restated as the problem that it isn't an *objective* matter whether these or those truths are themselves objective (cf. (Wright, 2003c)); but relegating questions of objectivity themselves to the non-objective is not the way of the quietist, who would do away entirely with sorting the objective from the non-objective.

(e) If the first of these, the expressivist will run into Wright's, to my mind quite solid, argument against the tenability of anti-realism about the psychological; if the second, the expressivist runs into quietism about traditional realism/anti-realism disputes in general.

I'll cast the foregoing as a dilemma for the avowal expressivist: find some way to handle the technical difficulties with denying truth-aptness to avowals, or go quietist about traditional realism/anti-realism debates in philosophy. I've, of course, not even hinted at suggestions for how to do either; but I think it will be agreed that either in itself would be rather a significant undertaking, and with grim prospects of success.[108] The progressive avowal expressivist, then, who shuns the denial of truth-values to privileged because expressive mental state self-ascriptions is, it seems, trading one set of headaches for another.

On the Question of Knowledge

I said a little in the last chapter about how I saw that there was an important structural similarity among each of the expressivisms we've considered with respect to the question of privileged self-knowledge, whether the conservative Wittgensteinian and Hackerian expressivism(s), or the (seemingly) more progressive expressivisms of Finkelstein and Bar-On. It's now that I want to set that out quite explicitly, because it seems to me rather indicative of an essential feature of expressivist projects about first-person privilege *per se*.

Indeed, the non-simple expressivisms hitherto canvassed all seem to me comparable to variations on a theme, a theme set by the simple expressivist back in chapter two. Remember, the simple expressivist saw it that privileged because

[108] Though, see (Horwich, 2006) for a contemporary quietist's plea.

expressive self-ascriptions weren't in any sense representative of knowledge the privileged self-ascriber had attained. This, seemingly paradoxical, let's call it *nihilistic*, response to the question of self-knowledge only derived whatever support it could from the emphatic likening of avowals to unlearned expressions of mentality in behavior. The motivation, recall, for such an emphasis on that likeness, indeed the wholesale assimilation of avowals to the non-linguistic/expressive, was the quite straightforward thought that one shouldn't, as it were, borrow only certain features of non-linguistic/expressive acts (viz. the inappropriateness of the perceptual model of introspection in their characterization) without borrowing certain other features (viz. the inappropriateness of any term of epistemic praise or blame in their characterization) so obviously and intimately interwoven with those acts *qua* non-linguistic/expressive acts. In any case, it was, as mentioned, precisely the emphatic likening of avowals to non-linguistic behavioral expressions that was to soften the initial shock of the nihilistic proposal.

Equally as nihilistic about the question of self-knowledge (at least of certain basic cases) were my Wittgenstein and Hacker. In this they were certainly conservative expressivists, but in their motivations for their nihilism about self-knowledge they were somewhat progressive. For them, the question turned to questions not strictly about how very like natural behavioral expressions of mentality avowals could be made out to be, but to questions about the *impossibilities* of *producing grounds*, *being mistaken* or *being in doubt*, the very phenomena of first-person privilege themselves, and questions about the all-or-nothing applicability of terms and their antitheses. As noted, this departure from the simple considerations

can, indeed, be brought back into contact, as it were, with the expressivist insight, just so long as it's precisely an expressivistic account of the phenomena of privilege that will be finally adverted to. For Hacker, certainly, at least within the "primitive language-games out of which their use arises" (Hacker, 2005, pp. 245), avowals of feeling, intention and the like are privileged just because they're expressive. And for Wittgenstein, it was notable that it was precisely the impossibility of moaning by mistake that was likened to (though, not strictly identified with) the impossibility of going wrong in identifying oneself in such sentences as "I have a toothache". (Wittgenstein, 1969, p. 67) It was also notable, for Wittgenstein, that it was the language-game "*with*" behavioral expressions of sensations that made for impossibilities of doubt and error. (Wittgenstein, 2009a, §288) In these ways, Hacker's and Wittgenstein's treatments of the question of self-knowledge in terms of the impossibilities of mistake, *etc*., can certainly be brought, at least to some extent, back to the expressivist insight. What complicates matters, it was seen and will be returned to, was their admission of cases of privileged self-ascription 'far-removed' from the most nearly expressive ones.

In any case, it was seen that their nihilistic response to the question of self-knowledge was only made plausible in the least by an emphatic insistence, first, on the impossibilities mentioned (at least in some cases and in some ways connected with the expressivist insight), and second, on the 'grammatical' point running, roughly, that where there can be no failure, there can be no success.[109] Their

[109] This point has some intuitive appeal, even (especially) divorced from the terribly unpopular 'grammaticalism' (conventionalism) about necessity that spawned it; it was, remember, agreed to by Armstrong (Armstrong, 1963, p. 422), for whom it was, unsurprisingly, used as an argument against privilege.

variation, then, on the theme set by the simple expressivist in response to (Q2) was simply to bring in more abstract considerations to do with impossibilities and antithetical terms generally.[110]

Turning now to Finkelstein's reaction to (Q2), we saw that, while not as shockingly nihilistic as Hacker's and Wittgenstein's, his was suspiciously *evasive*. Remember, for Finkelstein, the idea seemed to be that once the right view was taken of the phenomena of privilege, the question of privileged self-knowledge wasn't one of "philosophical import".[111] (Finkelstein, 2003, p. 152) Again, it seems we have a variation on the theme set by the simple, expressivist. This time it's an evasive (instead of nihilistic) reaction to the question of self-knowledge, garnering whatever plausibility it can from the direction of our attention precisely to the expressivist treatment of the phenomena of privilege. For Finkelstein, turning our attention the expressivist treatment of privilege is supposed to make the self-knowledge question *go away* (instead of get answered paradoxically, for instance). Opposite to Hacker and Wittgenstein, then, Finkelstein's progressiveness comes in his evasiveness of (Q2), while his conservativeness comes in his turning directly to expressivistic considerations (that is, not to more general ones about possibilities and antitheses) in an effort to help with the knee-jerk suspicion and philosophical disappointment certain to be aroused by his initial evasiveness.

---

[110] Hacker's suggestion for dealing with the technical difficulties attending this view will be brought in only after showing up the hidden similarities across the board for the surveyed expressivist responses specifically to (Q2).

[111] For Finkelstein, actually, his expressivist account was strictly a matter of accounting for someone's being the "best person to ask" (Finkelstein, 2003, p. 100) about what she was thinking, feeling, etc., and not a matter of accounting for the serious kind of privilege defined in chapter one. This difference, however, doesn't affect the point made above.

Dorit Bar-On too, we saw, wanted to direct our attention to the expressiveness of an avowal in order to help with the palatability, so to speak, of her response to the question of self-knowledge in avowing. While her response to that question was neither nihilistic nor evasive, it was nonetheless, we saw, reliant on conspicuously *weak* epistemic notions, notions of the "epistemic grounding" of avowals in the very states avowed and of the "epistemic warrant" enjoyed by an avowing subject simply by *being in* the state avowed. (Bar-On, 2004, pp. 390, 405) In this, no doubt, she is progressive, but again conservativeness creeps in and shows her response too is a variation on the theme set by the simple expressivist. Her conservativeness, just like Finkelstein's, comes in to the extent that, somehow, it seems, attention to the fact that avowals are *expressive* of the very states grounding them epistemically is supposed to make up for the weakness of that notion of epistemic grounding. Actually, it was seen that we first had to rule out that she really had hopes of letting that, apparently quite thin, notion of epistemic warrant carry the day, as it were, vis-à-vis the question of self-knowledge; we had to rule this out because otherwise she simply had a completely non-expressivistic (and highly implausible[112], though, indeed, not without adherents![113]) response to the question of privileged self-knowledge, and then her expressivistic account of the phenomena of privilege[114] would be rendered superfluous, and indeed rather puzzling.[115] But once this was ruled out, what we

[112] Cf. (Fumerton, 2006, pp. 61, 2).

[113] Notably, Bonjour (Bonjour, 2001) and Peacocke (Peacocke, 1998).

[114] Again, she doesn't characterize first-person privilege in the serious way I have, but that's unimportant to the point made above.

[115] As remarked by Chrisman in (Chrisman, 2009, p.8).

ended up with was something, as I've been trying to draw out, quite structurally similar to all the versions of expressivism investigated, from the most simple and conservative, to the seemingly most progressive.

Indeed, we can use Finkelstein's expressivism as a kind of bridge between Bar-On's expressivism, on the one side, and Hacker's/Wittgenstein's expressivism on the other. Remember that part of Finkelstein's response to the question of self-knowledge was that we could grant self-knowledge so long as we gave up that "knowledge always requires epistemic justification" (Finkelstein, 2003, p. 151) Furthermore, it seemed this only recommended itself, if at all, in light of Finkelstein's expressivist treatment of avowing. But why not bring Finkelstein closer to Bar-On? Why not have him granting that we can say what we want about *epistemic justification* (that is, we're free to grant it) so long as we give up that epistemic justification always involves more than simply being in the state one is justified in claiming to be in? That could only be as unintuitive as Finkelstein's actual line that we should give up that knowledge always requires epistemic justification. This modification of Finkelstein's response, just like his real response, would again, it seems to me, recommend itself only in light of the expressivist treatment of the phenomena of privilege. But then, it looks like what we end up with is something very like Bar-On's view, except that Finkelstein* (this modified Finkelstein) is only *allowing* that we should speak of epistemic justification in such cases, instead of seeing, as Bar-On does, that we indeed *ought* to. But, presumably, Bar-On only has it that we ought to do that in order not to jettison completely our pre-theoretical intuitions about privileged self-knowledge and the kind of justification peculiar to it;

but then, presumably, that's precisely what Finkelstein (Finkelstein*) is doing in merely allowing for talk of privileged self-knowledge (and the kind of justification peculiar to it).

So Finkelstein's* expressivism is, for all practical purposes, it seems, indistinguishable from Bar-On's. But then why not construct a Bar-On* in the reverse direction? That is, since Bar-On's conspicuously weak notion of epistemic justification in the case of avowing, just like Finkelstein's*, only gets forgiven its weakness in light of the expressiveness of acts of avowing, we may as well see Bar-On turning into Bar-On*, and claiming as well that, in light of the expressiveness of an avowal, *we can say what we want* about the questions of self-knowledge and its peculiar kind of justification, it being *no matter of philosophical import*; this is precisely what she seems to be doing anyway with such a weak epistemic relation. So then Bar-On's expressivism can be transferred without loss, it seems, to Finkelstein's, just as his can be transferred without loss to hers.

But then, we also saw how Finkelstein's say-what-you-want strategy made best sense as a reworking of Wittgenstein's/Hacker's own allowance for a *grammatical* use of sentences like "I (must) know I'm whether I'm in pain", a use intended precisely as a way of showing up the (*grammatical*) *exclusion* of ignorance, doubt and insufficient justification. And so, why not see Bar-On* (and therewith Bar-On herself) as very like Wittgenstein/Hacker; that is, of course, only insofar as we take the Wittgenstein/Hacker line to advert to an expressivist account of those grammatical exclusions?[116] Essentially, my suggestion is that Bar-On's proposal in

[116] Bar-On is keen to point out that her expressivism allows for *false* avowals, since it allows for "expressive failures". These are, for her, *not* errors of misidentification, or misascription, however.

the way of offering an expressivist inspired, positive answer to the question of self-knowledge, because it's so *weak* in the epistemic uhmph department, so to speak, is perhaps just as well recast as the seemingly distant proposal that talk of self-knowledge (and justification) just doesn't apply in the case of avowing.[117]

Therefore, I think that Bar-On, Finkelstein and the simple expressivist all do well to pay very much interest to Hacker's defense against the technical difficulties attending the view it seems the rest of them (through plausible transformations) can be seen as 'really' operating with. It was that line that was to teach us how to follow through with the expressivist response to Armstrong's argument against the very phenomena of privilege themselves, the argument, that is, that rested on the presumption that there's got to be substantial self-knowledge. While Armstrong didn't give an argument for *that* presumption, anything one could hope to muster toward that end was confronted head on by Hacker in his 2005 article. The basics, remember, of the barrage confronted by Hacker were simply that, besides simply *knowing* what mental states we're in, we take ourselves to do a lot else besides with respect to them, which lot else severally *entail* that, *contra* Hacker/Wittgenstein, we must indeed (have) know(n) ourselves to be in all those sorts of mental states with

---

And so, in any case, we could at least take it that *these* are ruled out ('grammatically excluded'). (Bar-On, 2004, 320f.) But even more than that, her motivating example for the possibility of expressive failures is just that familiar one of someone expecting something painful and so yelping even though completely unharmed. Now, *certainly*, Wittgenstein and Hacker can allow for *that*; so whatever kind of failure that amounts to, both Wittgenstein and Hacker can grant such 'failures' while preserving the grammatical exclusion, then, of a 'certain sort' of ignorance, doubt, etc. (probably, precisely the kind of ignorance Bar-On describes as resulting from errors of misidentification and misascription).

[117] As hinted at before, this should remind us of the situation Chisholm sees cropping up with his talk of "self-justifying" propositions: "Paradoxically, these things I have described by saying that they 'justify themselves' may *also* be described by saying that *they are* '*neither justified nor unjustified*.' The two modes of description are two different ways of saying the same thing." (Chisholm, 1982, p 137, my italics)

respect to which we want a proper account of first-person privilege. These were: lying about them, remembering them, acting for the reason that we're in them, knowing that other people know we're in them, and knowing that everybody in a group is in one of them while being ourselves in that group. The Hackerian response was simple: "all that follows from my remembering that I was in pain is that I was previously in pain, that I now know that I was in pain, and that I know now because I was previously in pain." (Hacker, 2005, p. 271) Similar treatment with adjustments were made for the rest; the only odd one out being acting for the reason that one is in a state (taking aspirin because you've got a headache, e.g.), which, on the Hackerian line, only entails that one is *not ignorant* that one is in the state (remember that, on this line, knowledge and ignorance *both* are ruled out). Now, it's not my concern right at the moment so much to evaluate this line as it is to show that, insofar as expressivists about first-person privilege one and all eventually confront the question of self-knowledge, they do well to pay special interest to this defense of the nihilistic response to (Q2). And that's precisely because all the expressivist responses to that question on offer seem, in the final analysis, to be different ways of making the same kinds of moves; and so there's hope that, should the Hackerian defense of the traditional Wittgensteinian version indeed prove quite secure[118], the other versions of expressivism will be able either simply to rework their own versions so as more obviously to align in specifics with the Hackerian line on this question, or else they may perhaps even be able to devise some structural equivalent within their own scheme. Thus, it seems to me, the prospects for a successful expressivism vis-à-vis

[118] There are, as yet, no published critical responses that I know of.

the question of self-knowledge are, from where I can see, best pinned onto the prospects of this, rather simple, Hackerian line.

If there's any problem I can foresee with it, it will perhaps come as an objection to the nonchalance with which it seems Hacker is happy to dismiss *other* philosophers' "uniform and elegant analyses of concepts" (Hacker, 2005, p. 271) (e.g. of the concepts of 'remembering' and 'lying') in favor of his own analyses (bearing their own marks of elegance and inelegance), which he seems to think he's entitled to just in being the expert English user he most certainly is. This can tend to get people rather upset.[119] Then again, if the people he thus upsets are, after all, reacting against the possibility of anything resembling *a priori* philosophy anyway, people who've given up on anything resembling conceptual analysis anyway, then their problem with the Hackerian line will be nothing other than their problem with anything like traditional philosophy itself. And, on the other hand, if the people he thus upsets are indeed quite content with conceptual analysis as the method of philosophizing, then I don't see that they'll have anything in his dismissiveness in particular to complain about, besides his coming off as somewhat impolite. In any case, any *philosophical* disagreement between them will come down to disagreements about *conceptual matters*. For instance, whether a given philosophical position is literally nonsensical (as Hacker is apt to think about many other philosophers' positions) will be both for Hacker and his rival traditional philosopher a *conceptual matter*, and the arguments for and against will have to proceed at the level of conceptual analysis for *both parties*; even though some people will perhaps see that it counts against Hacker that

[119] See (Dennett, 2007) for a paradigmatic example of the kind of offense that's apt to be taken to Hacker's seemingly unilateral, as it were, grammatical insights.

it's impolite to call someone confused to the point of uttering nonsense, such considerations are nothing to take with you into a show and tell battle of conceptual analyses.

So, barring ruffled feathers and Quinean naturalism, there doesn't seem to me to be much else that might be brought against Hacker's defense of the Wittgensteinian line on the question of self-knowledge (in cases, that is, where first-person privilege is in sway, as it were). That is, I don't myself find anything conceptually impermissible with his defense. And so, as I see it, the prospects for a viable expressivist treatment of (Q2) are quite good. Quite a good deal better, anyway, than the prospects for a viable treatment of (Q1) arrived at last section.

On the Question of Far-Removed Cases

The main approaches to (Q3), we saw, were two: the simple/Bar-Onian approach and the Hackerian/Wittgensteinian approach.[120] The first of these was to deny that there were any privileged mental state self-ascriptions that were, in any sense, *distant from* a set of basic privileged self-ascriptions most near to and happily compared with non-linguistic, natural reaches, cries, grimaces, *etc*. Bar-On, actually, had it that a mental state self-ascription was privileged "only to the extent that" it was expressive of the very state self-ascribed (Bar-On, 2004, p. 302); but this seemed just another way of saying "privileged only if expressive", and anyway it wasn't clear what in the world it was for Bar-On that was supposed to *compromise* (without entirely cancelling) the 'extent to which' a self-ascription was privileged-because-

[120] Finkelstein, we saw, didn't address the question.

expressive. (Most notably, it was seen that communicative intention made for no such compromise.)[121]

The approach we saw at work in Hacker and Wittgenstein, on the other hand, was simply to grant that indeed there are non-expressive, though privileged, mental state self-ascriptions; that is, we saw them granting, as Hacker put it, a "spectrum of cases" (Hacker, 1990, p. 196) of privileged self-ascriptions, one end very near to natural expressive behaviors, and the other end quite removed from them. The interest in this section, then, will be to assess the prospects specifically for the approach adopted by Bar-On, since whatever difficulty there is with the Hackerian/Wittgensteinian approach can only come out in the next section, where it's asked how the expressivist insight is supposed to help with the privilege attaching to the 'distant' cases.

The basic dilemma for Bar-On was that she has to cast all the *prima facie* privileged while non-expressive self-ascriptions either as expressive after all, or as non-privileged after all. Most of her approach consisted in grasping that first horn. The way she managed this was by telling the kind of 'replacement story' we saw Wittgenstein suggesting in chapter two, and which, as we saw last section, expressivists generally had better be at pains to spell out. The general form of her story is to work from moans, reaches, *etc*., to one-word interjections ("Teddy!"), to certain basic *spontaneously offered* sentences/exclamations ("I want Teddy!") and certain less basic ones ("I smell vanilla!"), and finally to so-called *reportive* avowals ("What is it you want, Sweetie?"; "I want the Teddy, Daddy.").

[121] Cf. (Bar-On, 2004, pp. 301-4).

One thing worth pointing out is how superficially similar this is going to look to something like a Quinean story of the 'roots of reference' in holophrastic stimulus-response verbal offerings.[122] But Bar-On wants to cast, as she should, the increasingly complex linguistic tools as "new forms of expressive behavior" (Bar-On, 2004, p. 288); thus, to tell her story, she won't want to have, like Quine does, that "'Ouch' is a one-word sentence which a man may volunteer from time to time by way of laconic comment on the passing show." (Quine, 1960, p. 5) Bar-On wants to, and needs to, deny that "a prerequisite of [one's] issuing a genuine self-ascription is that [one] form a recognitional judgment about the presence and character of the relevant condition" (Bar-On, 2004, p. 289), while admitting that the terms of genuine self-ascriptions "have currency in the language", and are "governed by various linguistic conventions" which one must "master" in order "genuinely to ascribe to [oneself] a desire", or any other mental state yielding to privileged self-ascription. (ibid.)

Now, the fact that, according to Bar-On, one's *mastery* of the *conventions* governing genuinely *referring* terms should account for their being used genuinely referentially while *at the same time* used, with respect to 'cognitive effort', simply as *tools of expression* (for states like *pain*) seems to put her very much in the Wittgensteinian spirit. (Cf. (Wittgenstein, 2009a)) Indeed, I don't see how the sort of replacement story Bar-On is trying to tell here could be understood in any other spirit. In any case, given the pressures set out two sections ago, it would seem that it's in just that spirit that expressivists should be telling their replacement stories anyway. And because we've already noted in that section the burdens that come with adopting the Wittgensteinian conception of language use, I'll just let those observations stand

[122] Cf. (Quine, 1990).

as marking the burdens weighing on Bar-On's replacement story, indeed on any expressivist's replacement story. It's really not the way Bar-On conservatively (that is, in step with simple expressivism) grasps the first horn of the dilemma above (for one range of mental state self-ascriptions) that's of particular interest in showing up the deep difficulty for expressivist views that I want to show up in this section. It's the way she conservatively grasps the second horn (for another range of mental state self-ascriptions).

That other range of mental state self-ascriptions comprises the *negative avowals*, avowals like, e.g., "I *don't* have a headache", "I have *no* desire to move my left thumb at the moment" and "I *don't* believe Goldbach's conjecture". Bar-On's line, instead of granting that while such self-ascriptions aren't at all plausibly cast as learned extensions of expressive behavior they're nonetheless privileged, was indeed denying that they "share in the special security of avowals". (Bar-On, 2004, p. 335) The term "negative avowals", on Bar-On's view, is in fact a *misnomer*, since such self-ascriptions aren't *avowals* at all. For Bar-On, such sincerely uttered claims are indeed "open to doubt, questioning, requests for reasons/explanation, etc.". (ibid.) This, however, seems patently false, so I'd like to examine the moves Bar-On makes in the way of making it seem plausible, if only for the sake of seeing how desperate the situation is for the conservative expressivist's response to (Q3), once it's granted that negative avowals won't fit into any replacement-of-expressive-behavior story.

The first move was already present in a way last chapter, though unremarked upon, in the passage of hers I quoted while setting out her response to (Q3). It consists in pointing out that negative avowals (as I call them, *pace* Bar-On) can often

be stand-ins for corresponding ("complementary") *positive* avowals.[123] She points out, for instance, that the merely grammatically negative self-ascription "I don't want to go" is often really the positive self-ascription "I'd like to stay/not to go", and that "I don't like this" really says "I dislike this". (Bar-On, 2004, p. 334) Examples abound, no doubt, but it seem to me the serious opponent won't be satisfied; after all, there are certainly at least *some* genuinely negative self-ascriptions (e.g. "I'm *not* in excruciating pain"), she will insist, that are most certainly ascribed with at least as much privilege ("special security") as their positive counterparts (e.g. "I *am* in excruciating pain"). Bar-On remarks that the genuinely negative ones, the ones not standing-in for any positive ones, "will not be easy to find".[124] (Bar-On, 2004, p. 335) But, it seems to me, it doesn't matter how easy or hard such cases are to find; what matters is whether there are any.

---

[123] Her notion of 'standing-in' (she doesn't use that term) is only as drawn out as is her flat assertion that "many [merely] grammatically negative self-ascriptions are not self-ascriptions of absence but rather ascriptions of 'complementary' mental states" (Bar-On, 2004, p. 334); that is, not at all. This shouldn't pose any problem, however, since, as I've been saying, viable expressivisms about privilege will have to, as it were, go Wittgensteinian about language use anyway. And the Wittgensteinian view is truly, you might say, all about 'stand-ins': superficial notational forms concealing what work is really being done by 'moves' in 'language-games'.

[124] One way she seems to over-reach in establishing the rarity of genuinely negative mental state self-ascriptions is in her mini-treatment of a subject's possible responses to the question "Do you want/hope for *q*?" For a subject genuinely to respond in the negative, says Bar-On, her response "is to be understood simply as meaning that she has no desire or hope one way or the other (not, mind you, as expressing indifference)". (Bar-On, 2004, p. 335) Indifference, we see for Bar-On, is supposed to be one of those complementary positive states that merely grammatically negative self-ascriptions sometimes really (below the surface) self-ascribe. But indifference is not ambivalence, the genuine *presence* of two (conflicting) states; indifference is the genuine *absence* of each of a pair of conflicting states. Of course, one can *act* indifferently, just as one can *act* freely from pain, and freely from any mental state, or pairs of them (besides, perhaps, the state of being awake enough to act all, if we're calling that a mental state). And if *this* is called 'expressing indifference', then so can smiling be called 'expressing freedom from pain', and so can every would-be mental absence be seen as indeed positively expressible in behavior, so that genuinely negative mental state self-ascriptions can turn out entirely fictional, for the expressivist, and not simply infrequent. In that case, of course, the expressivist needn't let go of the first horn of the dilemma: what seemed impossibly taught as a replacement for expressive behavior, she will argue, is indeed possibly so taught. I return to this briefly in the conclusion of this section.

Bar-On's second move is to suggest invoking *introspection*, very much in the traditional *scanning* sense (that is, not the completely thin sense, whereby any and every avowal, even expressivistically construed, is counted the result of 'introspection'). She suggests that, in the cases at hand, a subject "surveys her present state of mind, and finds no belief that *p* or desire/hope for *q*", or sensation *X*, etc. (ibid.) An expressivist, she says, doesn't have to deny that "we do sometimes employ genuine introspection as a mode of access to our present state of mind." (ibid.) But then, on the strength, no doubt, of previous (good) arguments in her book[125] she points out that introspection shouldn't be expected to deliver anything like the security (freedom from doubt, questioning, requests for reasons/explanation, etc.) it was the point of her Neo-Expressivism to account for in the first place. And thus, it's not supposed to come as any surprise, on the suggested invocation of genuine introspection, that genuinely negative self-ascriptions should end up looking not at all like the privileged bunch their positive counterparts were and remain.

This second move too, however, will be unconvincing to the serious opponent. The suggested invocation of less-than-privileged introspective processes seems to ignore the real source of dissatisfaction with the concession itself that, for example, "I have *no* desire to move my left thumb at the moment" should be denied the possibility of being uttered with *every bit* the privilege an utterance of "I *do* desire to move my left thumb now" can be uttered with. It can seem rather as if someone should claim that, for example, there are no cats in America, and then if that someone were to defend the claim by suggesting that if there *were* no cats in America, this might've happened because they all got sick and died. Perhaps we can allow the

[125] Cf. (Bar-On, 2004, pp. 95-104, 122-128).

expressivist the actuality of a genuine (scanning model) kind of introspection, as long as this doesn't have anything to do with her account of privileged self-ascription.[126] But that's just like allowing that things do get sick and die; and that certainly doesn't mean allowing that that could be why there aren't any cats in America. We aren't looking for an explanation for that, because it's patently false. Similarly, allowing the expressivist the actuality of genuine efforts to scan the contents of her mind doesn't mean allowing that that could figure into an explanation for why "I have *no* desire…" isn't possibly every bit as privileged as "I *do* desire…". Why propose explanations for what's patently false?

It seems to me Bar-On's moves in the way of making plausible her denial that there even *are* any negative avowals simply don't speak to the *from-the-outset* implausibility of that denial. And I don't see any other moves available, either. I think, therefore, that there isn't much hope for a Bar-Onian expressivist's response to (Q3).

But what about backtracking some in order to preserve the conservative response? Specifically, What about trying to work with the first horn of our dilemma exclusively, subjecting even the problematic negative avowals to the replacement-of-expressive-behavior treatment? More specifically, What about the suggestion that would-be mental absences are indeed positively expressible in behavior? For instance, it might be claimed that smiling while laughing and playing, *etc*., as a matter of fact positively expresses *freedom* from sadness (or pain, or what have you). Or it might be claimed that sitting, thinking, completely undisturbed by the periodic

[126] Though, to my mind, this seems so contrary to the expressivist spirit that I can't think of the resulting picture of things as anything less than bizarre.

departure and return, say, of a squirrel did, as a matter of fact, positively express one's *freedom* from care about whether the squirrel stayed or went. And it might be suggested that it was during such moments as these that our elders taught us to replace our natural expressions of these states of freedom-from-*x* with new, linguistic means of expressing them.[127]

One difficulty with this is that for any actual laughing, playing child, for example, she'd have to be positively expressing *infinitely many* freedom-from-*x* states of mind (not to mention her happiness too) while going about her playful activity, so that it would seem impossible to say which freedom-from-*x* state exactly (if any--it could, again, be her happiness) we were replacing with the sentence "I'm not in pain", for example, should we take the occasion of her being free from pain as an opportunity to inculcate the use of that particular negative avowal. This difficulty, of course, generalizes beyond the case of "I'm not in pain" to every other negative avowal.

But the real problem with this view, it seems to me, is with the very notion of *expressing* an *absence*. Certainly, one can't express a state one isn't in, but it seems to me one doesn't express *not* being in a state either. What one expresses are precisely those positive, 'complementary' states Bar-On was keen to point out many seemingly negative avowals are really stand-ins for, which positive states, no doubt, *exclude* their complements; but the *exclusion itself* isn't something that gets expressed. The absence of pain behavior, for instance, isn't to be thought of as the *presence* of the-absence-of-pain behavior; thinking this is too much like thinking that if something came from nothing, then there must be something (namely nothing)

[127] This kind of move was suggested to me in conversation by Jeremy Shipley.

from which it came. So it's not just that you run the risk, in the teaching scenario envisaged, of accidentally replacing a child's expressions of (positive) sadness-excluding states (e.g. happiness) with the new expressive tool (e.g. "I'm not sad"), it's that there's no expression precisely of the-absence-of-sadness one could even hope to replace to begin with. (One doesn't *replace* a hole with a plug.)

Of course, one might well take the occasion of a child's clearly being free of some state *m* as an opportunity to get that child to say (roughly) "I'm not in *m*", thus instructing him in the use of that sentence; but this isn't getting the child to replace occurrences of his non-occurrence-of-*m* behavior; there are no such occurrences to be replaced. (Again, if we're talking about replacing his *m*-excluding behavior, that's quite a different thing.) So, if the child as a matter of fact learns from such instruction (that is, from instruction in the use of "I'm not in *m*" that consists (at least in part) in being brought to say so precisely when it's true), it simply won't be due to any kind of replacement technique whose immense importance when actually applicable expressivists like Bar-On do well indeed to point out.

Making sense of what *could be* going on in such learning scenarios, then, is a serious problem for the expressivist. And so conservative expressivism in response to (Q3) faces rather serious difficulties. Neither horn of the dilemma 'replacement story or non-privileged' seems workable when it comes to negative avowals. A viable expressivism, I believe, will have to respond progressively to (Q3), thus opening itself up to (Q4), the subject of our next discussion.

On the Question of Relating Privilege Across the Spectrum of Cases

The fourth of my questions for expressivists about first-person privilege, we saw, was never quite fully engaged by any of the expressivist views on offer; neither the simple (fictional) expressivism I constructed in chapter two, nor any of the other expressivist treatments actually on offer, had it *both* that there were privileged mental state self-ascriptions farther removed than others from the behavioral expressions of mentality that are agreed, among expressivists, to be 'extended' in the acquiring of certain basic, privileged means of self-ascription, *and* that there was some important and expressivist-insight-involving way of relating that privilege across the 'spectrum' of privileged psychological self-ascriptions. Bar-On and my simple expressivist both, we saw, denied there were privileged while non-expressive self-ascriptions, and so, in their way, they certainly do acknowledge the question. Theirs were, then, the most fully worked out expressivist theories encountered. Theirs also, we saw last section, suffered extreme difficulty, particularly on the question of negative avowals. Hacker and Wittgenstein, on the other hand, granted a 'spectrum of cases' of privileged self-ascriptions, but never bothered to relate first-person privilege across that spectrum in a particularly expressivist-friendly way, or even to acknowledge the question whether that's something worth doing. Finkelstein, we saw, never addressed even the possibility of a spectrum of cases. Therefore, in trying to assess the prospects for expressivist responses to (Q4), I won't have any actual responses to that question to work with. And so I'll have to come up with the best response I can myself, on behalf of the expressivist.

What I'd like to do first, however, is to remind the reader of just how the Hackerian and Wittgensteinian concessions came out that indeed there is such a spectrum. Then I'll remind the reader why this concession should pose a problem for anyone hoping for a thoroughly *expressivistic* treatment of the problem of privilege before turning to set out my own suggestion for an expressivistic response. (It should be remembered, of course, that neither Hacker nor Wittgenstein ever claimed to be expressivists; neither ever claimed to have fully worked out an expressivist account of privilege. Still, it seems to me, anyone who *is* trying to work out such an account does well to follow their lead in admitting the 'spectrum of cases'; otherwise the difficulties set out last section will be inevitable.)

We'll begin with Wittgenstein. The crucial passage came from *PPF*:

> I say "I'm afraid"; someone else asks me: "What was that? A cry of fear; or did you want to tell me how you feel; or was it an observation on your present state? -- Could I always give him a clear answer? Could I never give him one? (Wittgenstein, 2009b, §73)

His follow up to this was that "[o]ne can imagine all sorts of things here" as ways of rendering what was 'done' with that sentence, and the most important one for our purposes was: "I'm afraid. I am sorry to have to admit it". (Wittgenstein, 2009b, §74) This, it seemed clear, was supposed to be a matter of telling how one feels *as opposed to* crying with fear, as in his other example: "No, no! I'm afraid!". (ibid.) And just a few remarks later we find that "[t]he words 'I am afraid' may approximate more, or less, to being a cry. They may come very close to one, and also be *very* far removed from it." (Wittgenstein, 2009b, §83)

This notion of coming closer to and retreating farther from a cry certainly seems an important one for the expressivist, but just as important for Wittgenstein's

expressivism (such as it is) is the thought that simply in *telling* how one feels one should retreat from the basic, expressive case, that one should so swiftly transcend, that is, the blissfully tractable cases at the 'low-end' of the spectrum.

We saw that Hacker's view, an explication and defense of Wittgenstein's own, comes to very much the same thing:

> The affinity between spontaneous avowals and natural expressive behavior must not mask the fact that the uses of first-person psychological sentences are heterogeneous. Some approximate to primitive cries and gestures, and others are far removed from those paradigms. Wittgenstein was not suggesting that there is no such thing as reporting, informing, telling others how things are subjectively with one. But what is called 'telling someone what one feels', 'describing one's state of mind', or even 'observing one's emotional state' are much more unlike reports, descriptions, and observations of the physical world than one thinks. (Hacker, 1990, p. 196)

It's seems clear that, for Hacker, what makes first-person reporting, informing, telling, *etc*., even though removed from the paradigms of primitive cries and gestures, so unlike reports, *etc*., of the physical world, is precisely that the former are *privileged* in a way the latter aren't. For instance, when someone lies about how she feels, she is thereby removed from the 'low end' of the spectrum because, Hacker grants, she is *telling* someone (an untruth) about how she feels (with the intent to deceive); but lying about one's feelings is supposed to be radically unlike lying about, for instance, someone else's feelings, precisely because "I lie about my feelings or thoughts [merely] if I feel or think such-and-such and deny that I do, with the intent to deceive." (Hacker, 1990, p. 199) This, we saw, was exactly the analysis of lying about one's own mental life that Hacker uses to combat the difficulties for the view that, because there can be no self-ignorance when it comes to so much of one's mental life, there can be no self-knowledge either. Also, for Hacker, when one,

removed from the low end of the spectrum, "confesses" what one thinks or feels, "[one's] sincere confession is a criterion for [one's] thoughts and feelings", whereas, "when [one] say[s], in all honesty, that A is upset, that does not guarantee (*ceteris paribus*) the truth of what [one] say[s]". (ibid.) Furthermore, for Hacker, cases of "confessing, reporting, or telling what one thinks, intends, or feels" are cases wherein one does so without grounds. (ibid.) Descriptions of one's mental life, now, the 'high end' of the spectrum of psychological self-ascriptions, are, unlike reports, confessions, *etc*., reserved for the purposes of describing aspects of one's motivations and emotional life on Hacker's view; and these Hacker grants are indeed "run through with the distorting influence of the will and fantasy". (Hacker, 1990, p. 200) But defects of the will, "rather than of the intellect", have nothing whatever to do with "mistaken observation" on the Hackerian view. (ibid.)

It's hard to know how much of all of this detail Wittgenstein would have agreed to, though it's plain that in every way it's very much inspired by Wittgenstein's writings; and certain surprising details (e.g. that "to say what one intends, believes, or thinks is never to describe one's mental state" (Hacker, 1990, p. 199)), which I can't help but think other details are contrived to bend around, come direct from Wittgenstein.[128] In any case the picture that emerges is clear enough to state a puzzle for the project of understanding the phenomena of first-person privilege wholly, or even substantially, on the strength of the expressivist insight. We have only at one far end of the spectrum of psychological self-ascriptions the appropriateness of the model of replacements of expressive behavior, while we have

[128] Cf. (Wittgenstein, 1980a, §599). It should be remembered, of course, that while it seems obvious Hacker gives his full endorsement to these views, they do appear in a *commentary* on Wittgenstein's own views.

shared throughout the spectrum (though slightly less so, and only 'less so' in a qualified sense, at the highest end) "the central perplexing features of 'I am in pain'". (Hacker, 1975, p. 268) The question for the expressivist is this: What has the privilege attaching to the *non*-expressive cases *got to do with* the privilege attaching to the expressive cases? The worry is that if the privilege attaching to the non-expressive mental state self-ascriptions doesn't in any way *derive from* the privilege attaching to the privileged-because-expressive self-ascriptions, then the importance of the expressivist insight for understanding the full range, or even most, of the phenomena of privilege is put in serious doubt, especially since we bid farewell to the expressive cases, on Hacker's and Wittgenstein's view, just as soon as we move on simply to *telling* others what we think, feel, *etc*. The puzzle for the expressivist is how to meet this worry. What's the crucial, importance-of-the-expressivist-insight-saving relation supposed to be between the non-expressive cases and the expressive ones?

That there even should be a continuum itself already seems to suggest there may be an important relation, a somehow 'downward' pointing relation, among the full range of cases of privileged self-ascriptions; but that the expressivistic end of the spectrum should be so narrowly restricted, that simply in *telling* about one's fear, for example, one should have compromised the expressiveness of his self-ascription, this may perhaps seem to count against the hope of grounding one's understanding of the phenomena of privilege on the likening of privileged self-ascriptions to expressions of mentality in behavior, which likening was the liberating thought for at least all those 'bottom-most' cases, which liberating thought the expressivist is, *qua*

expressivist, out to milk for all it's worth, so to speak. But if the liberating thought applies in the first instance only to *such* exceptional cases, How could one's understanding of the full range of privilege be grounded in that thought? Certainly, one would think, not by those cases being the *model* for the rest. (The working out of that strategy is just the work seen last section, whose otherwise valiant execution we saw stopped in its tracks on the question of negative avowals.) The model-model won't do.

Minimally, if there's to be the dependence relation the expressivist needs, she'll have to insist that the existence of privileged-because-expressive cases (p.b.e.'s) is a *necessary condition*[129] for the existence of privileged-while-non-expressive cases (p.n.e.'s).[130] But, importantly, more than simply that if-then relationship will have to emerge if the p.n.e.'s are to be seen in any significant sense to derive their puzzling privilege from the non-puzzling privilege of the p.b.e.'s. The reason is, we could simply grant the expressivist that the p.n.e.'s demand the p.b.e.'s, but that would only be to grant that a puzzling set of facts demands a non-puzzling set of facts; and that's not to remove any puzzle. I will (eventually) return to this point.

So, just having that if-then relationship won't be enough, but it will need to be in place. What more needs to be in place if the expressivist admitting p.n.e.'s is going to make sense of the privilege attaching to them, but only by way of the sense she *already* makes of the privilege attaching to p.b.e.'s.? It seems to me that what more needs to be in place, actually, is a story about how p.b.e.'s could figure in a *sufficient*

---

[129] A necessarily necessary condition, that is; not simply materially necessary. Please read the modal operator throughout as appropriate.

[130] "p.b.e's" will stand either for "privileged-because-expressive *cases*" or "privileged-because-expressive *self-ascriptions*" as appropriate; *mutatis mutandis* for "p.n.e.'s".

condition for p.n.e.'s. The idea would be to work *from* the non-puzzling cases of privilege, making it clear how they could *result in* the initially puzzling cases of privilege. Indeed, the present kind of expressivist wants to make the privilege attaching to the p.b.e.'s out to be the *source* of the privilege attaching to the p.n.e.'s. She wants to answer the question, "Where's all this privilege coming from?" by pointing to the p.b.e.'s and saying, "There". What it seems the expressivist needs then, if she admits p.n.e.'s, is to work out the *mechanism of transfer* of privilege from p.b.e.'s to p.n.e.'s.

But now it seems something important is happening, namely that the task is shifting away from that of making sense of the problematic privilege attaching to p.n.e.'s by way of working out how such privilege is only a privilege conferred on them by the p.b.e.'s, whose privilege the expressivist sees as unproblematic. And what the task is shifting to is precisely that of *making sense of* a 'mechanism of transfer' of *privilege*. It will seem to the expressivist admitting p.n.e.'s that there *must* be such a mechanism; that's why she styles herself an *expressivist* about privilege, even though admitting p.n.e.'s. But there seems to be a problem in that she has yet even to *make sense of* how it could be. Even if the expressivist admits that some of these metaphors (e.g., a 'source' and a 'mechanism of transfer') are a bit picturesque and that all she hopes, putting it least picturesquely, is to make sense of *some* kind of dependence relation, the acknowledgment of which would go toward one's grounding one's understanding of the p.n.e.'s in one's understanding of the p.b.e.'s, still it's a dependence relation she's yet to *make sense of*.

And so we can ask, Hasn't the expressivist simply let go of one seemingly intractable "How could it be?" question only to grab hold of another? And aren't such questions just the sort Wittgenstein warned against, when he called the question "How can one think what is not the case?" a "beautiful example of a philosophical question"? (Wittgenstein, 1969, p. 30) And if the expressivist insists that there *must* be something like a mechanism of transfer, some kind of dependence relation that she can't yet quite pin down, isn't she open to the sort of complaint that Wittgenstein makes in another connection, viz. that she has "no model of this inordinate fact"?[131] (Wittgenstein, 2009a, §191)

This is perhaps just the place to insert Wright's anti-explanatory reading of Wittgenstein on the question of privilege that we saw in chapter three. The idea was that views like expressivism about first-person privilege are simply wrong-headed, as Wittgenstein sees them, precisely because they don't take their would-be *explananda* simply as matters of 'autonomous grammatical fact'. As Wright reads Wittgenstein, the point is just that "the first-third-person asymmetries that pose our question [Whence privileged self-ascription at all?] belong primitively to the 'grammar' of the language-game of ordinary psychology" (Wright, 1998, p. 39), and so expressivist accounts are simply out of the question, if Wright's Wittgenstein is right, *even with respect to* the most nearly expressive avowals. I urged in chapter three that there's a kind of 'explanation' that, even in Wittgenstein's own work, isn't the *bad* kind of philosophical explanation he denounced, but which amounts simply to *shedding light on* a difficulty. And the point there was that the expressivist only hopes to *shed light*

[131] See (Wittgenstein, 2009a, §§191, 2), where the "inordinate fact" was the mysterious grasping "in a flash" of the "whole use of [a] word at a stroke".

on the problem of privilege, even the privilege attaching to p.n.e.'s, by way of the comparison (of p.b.e.'s) with expressive behavior.

Perhaps now, however, we're seeing that Wright was right after all, because it's beginning to seem that, whatever puzzlement there was arising around the phenomena of privilege in the first place (before any comparison with expressive behavior was made), there's the same kind of puzzlement arising for the expressivist now around her mysterious mechanism of transfer. Now, the expressivist may, somewhat ironically perhaps, grab hold here of just Wright's point, claiming that it's simply 'grammatical' (not to be explained in any way) that the privilege attaching to p.n.e.'s indeed owes itself to the privilege attaching to p.b.e.'s. One response to this is, of course, that the expressivist could've saved herself a step and made this move right from the outset, *per* Wright's suggestion. However, the expressivist might respond that it's not a matter of saving steps, but of seeing things rightly; and seeing that the relevant (not to be explained in any way) grammatical fact is precisely that the source of privilege across the spectrum of psychological self-ascriptions is in the kinship between p.b.e.'s and expressive behavior, that's seeing things rightly. This response is, however, no good as I see it.

First of all, as a reading of Wittgenstein, this not-to-be-explained-in-any-way idea is simply bogus. Sure, sometimes Wittgenstein's responses to philosophical difficulties like "How can one think what is not the case? Isn't that like trying to hang a thief that doesn't exist?" were, you might say, grammatically curt: "Our answer could be put in this form: 'I can't hang him when he doesn't exist; but I can

look for him when he doesn't exist'".[132] (Wittgenstein, 1969, p. 31) But, as he says, this is just one *form* of responding to that difficulty. And nothing is clearer than that, notwithstanding the occasional "drop of grammar" (Wittgenstein, 2009b, §315), Wittgenstein worked doggedly to explain in great detail how to shake loose of philosophical difficulties (mainly by explaining how we get into them); this is hardly the simplistic grammaticalism Wright seems to think sums up Wittgenstein's approach to the problems of philosophy.

But quite apart from its being a poor reading of Wittgenstein, such an expressivist's response to the current difficulty is simply philosophically uninspired and uninspiring, not to mention completely unsatisfying. True, it's *slightly* more inspired than the constitutivist line Wright attributes to Wittgenstein; it does, after all, invoke an insightful comparison with expressive behavior. But I don't see that it should count as any more thoroughly *expressivistic* a line on the problem of privilege than the line that would, indeed, make the same insightful comparison between p.b.e.'s and expressive behavior, but that would adopt a strictly Wrightean constitutivism with respect to the p.n.e.'s, not minding about any extra, purely grammatical and not to be discussed, rule connecting these with the p.b.e.'s. And obviously *that* line could hardly be called an expressivist line on first-person privilege. That line would have to be called a hybrid expressivist/constitutivist line, containing, however, only as much helpful philosophical insight as is contained in its comparison of the p.b.e.'s with expressive behavior. And in this last respect it seems to me indistinguishable from the so-called expressivist line now under consideration.

---

[132] Or, "How do I know that this color is red?--It would be an answer to say: 'I have learnt English'". (Wittgenstein, 2009a, §381) This would be *an* answer, not *the* answer, and presumably one that only makes sense as such at all in relation to *other* (less playful) answers.

Can the expressivist do better, then? What about the problems that led us even to consider that Wright's constitutivism was muscling its way onto the scene? There was, first, the problem that the expressivist's "How could it be?" about relating the p.n.e.'s appropriately to the p.b.e.'s seemed just as intractable as the initial "How could it be?" about the very phenomena of privilege themselves, which initial "How could it be?" made it's appearance well before any expressivist insight. And second, there was the problem that it seemed the expressivist had no "model" for the "inordinate fact" of her, as yet uncomprehended, appropriately dependence-exhibiting relationship between the p.n.e.'s and p.b.e.'s. Can the expressivist address these afresh? I think so. These problems seem to me pseudo-problems, pseudo-Wittgensteinian scare tactics that ought not to put off the expressivist. However, they are welcome worries insofar as, it seems to me, quelling them shows the way to what I take to be the best hope for an expressivist's response to (Q4).

About the worry that there's something wrong with these "How can it be?" questions, one just as intractable as the other, it needs to be pointed out that, for Wittgenstein, such questions were "beautifully" bad only insofar as they expressed puzzlement over some *mundane* fact that in ordinary life doesn't puzzle. One of his favorite examples of this was Augustine's puzzlement about time.[133] And the main thing about that puzzlement for our purposes was that, famously, Augustine felt he knew what time was only when he wasn't *asked*. Such puzzlement (about what we take for granted in ordinary life) seemed indeed to be definitive of just the kind of philosophical puzzlement Wittgenstein saw himself as combating throughout his later

[133] See (Wittgenstein, 1969, p. 26) and (Wittgenstein, 2009a, §436) for just two of the explicit references to Augustine on time. Discussions of the Augustinian puzzles come up in many more places in Wittgenstein's later writings.

work. And it seems perfectly true that it's just this kind of puzzlement that characterizes our *initial* philosophical unrest in the face of the phenomena of privilege. But is the question "How could the privilege attaching to non-expressive-while-privileged mental state self-ascriptions be seen to depend on the privilege attaching to privileged-because-expressive mental state self-ascriptions?" expressive of a puzzlement over something we take ourselves ordinarily to take for granted? Something tells me, No.

And anyway, are we to take it that *all* puzzlement in the course of thinking through a philosophical problem was seen as so ill-begotten for Wittgenstein as Augustine's puzzlement about time? A moment of consideration (and some imagination) will show that it can't be so. Take Wittgenstein's contention that "if there were, for instance, no characteristic expression of pain, of fear, of joy…our normal language-games would lose their point."[134] (Wittgenstein, 2009a, 142) Now, I don't doubt that very many people have experienced a sense of "How could it be?" puzzlement in response to that contention.[135] But have they been, like Augustine, inclined to say they ordinarily knew how it could be, only not when they were asked? I shouldn't think so. Now imagine we went back in time and caught up with Wittgenstein himself in Cambridge. Imagine we met him only a couple days before he first had this very thought. Imagine we suggested it to him ourselves. Is it necessarily the case that in the moment he heard the suggestion he would've snapped

---

[134] This idea in general, of language-games 'losing their point', is paramount in Wittgenstein's later writing. One important discussion of the idea in explicit terms appears in (Wittgenstein, 1970, §350).

[135] Putnam certainly would've "come to bury…not to praise" such a view in 1961 when he set out to show that "even the weakened forms of the logical behaviorist doctrine are incorrect". (Putnam, 1980, p. 26)

it up? When Wittgenstein sat in his lectures "shaking his bent head slowly" and giving the "impression that thinking was a difficult business" (Mays, 1967, p. 81), was it that he was simply waiting for thoughts like this to come to him, which he then took up on sight, as it were? That can't be right. It's obvious to me anyway that he could very well have been presented with this very thought about our language-games losing their point and puzzled over it, wondering what there was to it, wondering whether and how it could be. And most definitely he would've thought the same could happen for his readers. I can't believe for a second he would've discouraged difficult bouts of thought over, for instance, how it could be that he was right to say what he did about language-games losing their point. I can't believe for a second either that he would've discouraged people from holding onto what they saw as a great deal of promise in this suggestion of his, while nonetheless not quite having put it all together. And I refuse to believe that it should've mattered to him who suggested what to whom, or whether the who and the whom were the same person or different people. Furthermore, I completely side with my vision of Wittgenstein on these points, and I thus reject the first pseudo-Wittgensteinian worry for the expressivist hope now under consideration. ("But how does one know which are the good puzzles and which the bad?" One does one's best.)[136]

[136] Take the question, How do we know when we're bringing preconceptions to a question, preconceptions about what we *must* find, instead of simply acknowledging what we *do* find without prejudice? That is, How do we know if we're abiding by Wittgenstein's "Don't think, but look!"? (Wittgenstein, 2009a, §66). It seems to me we should take such recommendations ("Don't think, but look!") rather like a master chess player's recommendation, for instance, never to cede the initiative. It only begins to be any help after much chess-playing and after the master has pointed out in a number of cases that, quite to one's surprise, *that* was ceding the initiative. But, hopefully, one does eventually begin to catch on. (Not if one hates chess, of course.)

The second pseudo-Wittgensteinian worry was that it seemed the expressivist's hope of understanding how the privilege of p.n.e.'s should depend on the privilege of p.b.e.'s was a hope of understanding an "inordinate fact" for which she had no model. This worry too, I believe, is ill-founded. For one thing, the same as before, as it was used in Wittgenstein's writing, the charge of trying to understand some inordinate fact on the basis of some missing model was a charge against trying to understand some over-blown version, as it were, of something quite ordinary.[137] Again, I see it that the expressivist will take the (to her) promising suggestion "the p.n.e.'s depend for their privilege on the p.b.e.'s" along the lines Wittgenstein himself might've taken the suggestion "without ordinary behavior our language-games of the psychological would lose their point" before he came fully to endorse it. That is, again, she won't take it as something she'd freely admit she ordinarily assumes to be the case but that sometimes (when 'doing philosophy' as Wittgenstein would put it) comes to seem highly problematic and mysterious.

But what's more important than any of that, I believe the expressivist has available to her the following response: she can say that she *does* have a model of the kind of dependence she only, as yet, perceives dimly, and that it's precisely the kind of dependence exhibited in that our language-games with psychological vocabulary should *lose their point* absent the actual and usual expressions of mentality in behavior that are their background. I'm going to say more shortly about how the expressivist can possibly hope to follow through with this response, but right now I want to return briefly to something left unfinished a little while back.

---

[137] Namely, in the passage this particular phrase comes from, some "much more direct sense" than the ordinary sense in which one can "grasp the whole use of [a] word at a stroke". (Wittgenstein, 2009a, §191).

It was said before that we could even grant the expressivist her dependence relation if all that came to was that the existence of the non-puzzling p.b.e.'s was a necessary condition for the existence of the puzzling p.n.e.'s; the thought was that simply pointing out that something puzzling requires something non-puzzling isn't pointing the way out of puzzlement. And it seems much the same thing can be said here: simply pointing out that something puzzling would lose its point absent something non-puzzling isn't yet to point the way out of puzzlement. Last we visited this line of thought, I just left it at that; now I'd like to respond to it.

It seems to me this is perfectly right, actually, but that the idea of just being *granted* one's is-a-necessary-condition-for relation or one's loses-its-point-in-absence-of relation is rather far-fetched and indeed subversive to the expressivist's purposes. One is never simply granted such things, and whatever there is disappointing about the thought of that happening shouldn't be taken to show that it could only be just as disappointing should one come by them honestly. In particular, the expressivist doesn't want merely to *assume* her loses-its-point-in-absence-of relation, but she wants to *establish* that there's such a relation between the use of p.n.e.'s and the use of p.b.e.'s. The problem has been all along which relation the expressivist could possibly hope to *establish* that would do the work she perceives dimly can be done, not which relation the expressivist could hope would *fall out of the sky*. But if the expressivist indeed establishes the relation in question honestly, that is, if she has a firm grasp of how and why the relation holds, then this reason for thinking that turning her thoughts to that relationship *couldn't* be the key to removing her puzzle will fall away.

Consider the following (in obvious, and no doubt significant, ways disanalogous) analogy. Suppose we just knew *that* if five-year-old Johnny has fifty dollars in his pocket then his mom had fifty dollars in her pocket, but we had no idea *why* this conditional should hold. It was given to us out of the clear blue sky, let's say. In that case we certainly wouldn't and shouldn't be satisfied by any appeal simply to that the puzzling fact (five-year-old-Johnny's having fifty dollars) demands the less puzzling fact (his mom's having had fifty dollars) as an attempt to address our puzzlement. "Ok," one might say, "if he's got fifty dollars then she had fifty dollars, but I'm still puzzled how a five-year-old ends up with fifty dollars". On the other hand, though, if we should understand better *why* this conditional holds (let's say, it was because whenever Johnny has fifty dollars he *got it from* his mom), then there's a much greater chance we're approaching our solution. In general, the more you know about the special reasons *why* puzzling x's demand non-puzzling y's, the likelier it is you'll be well placed to see how, starting with y's, one should end up with x's. The likelier, that is, that the puzzle of the x's will fade in light of their dependence on the non-puzzling y's.

So, what's it going to be, then, to understand the how's and why's of a loses-its-point-in-absence-of relation? It's going to be different, certainly, from understanding why a child should have fifty dollars only when his mom had fifty dollars. It seems to me it will be radically different, in fact. Consider again Wittgenstein's claim that talk of the mental loses it point in absence of expressive behavior. How does one arrive at an understanding of this rather singular relation of

dependence? What I want to suggest is rather sketchy, and the sketchiness of it only goes to show, I'm afraid, the plight of the expressivist.

My suggestion is this: the kind of understanding of a relation like this that an expressivist will be after comes as a matter of locating it within a broader framework of such relations. It needs to be asked just which *other* language-games are related to just which "extremely general facts" (Wittgenstein, 2009a, §142) in anything like this kind of way. We need to ask whether there's any discernible pattern among the various instances of this relationship we uncover, or whether each new instance is more like, as Wittgenstein said about Euclidean proofs, it's own sort of "trick". (Wittgenstein, 1979, p. 116) The suggestion is that understanding a relation like this one involves rather a lot of beating around the bush. We don't focus intently on what puzzles us, the way we might focus intently on every movement of the sleight-of-hand magician's hands. Rather, we stand back, as it were, and take in all the cases, each of which is individually troubling. But this isn't so that somehow we should thereby devise a method we can take back with us to any individually troubling case and with it spot the before unseen card-in-palm, as it were. Switching metaphors, our puzzlement is not like that of the person completely unfamiliar with wireless technology stunned at his first encounter with a remote controlled toy car. The dependence relation he sees baffles him because he has no clue about (no theory of) the unseen (causal) connections between the remote and the car. And my suggestion is not like the suggestion that he surround himself with wireless technology so as simply to *get used to it*. Doing that really would be suggesting he give up on the

quite legitimate search for hidden connections.[138] But there are no *hidden connections* whose discovery would explain the kind of loses-its-point-in-absence-of relation I'm suggesting the expressivist should be emphasizing. Coming to understand the dependence of the privilege of non-expressive avowals on expressive ones as a matter of sustaining-by-conferring-a-point-on is simply not itself a matter of discovery *in that sense*. It's only a matter of discovery in a much more frustrating sense, I suppose, for some philosophical theorists.

Such is my sketch for the strategy an expressivist should pursue in arriving at an explanation of non-expressive avowals' deriving their privilege from expressive avowals' privilege. And before going on to discuss, only briefly of course, the obvious implausibility of pulling off anything so fanciful, I want to address one minor concern. It will perhaps be thought that even if, instead of sprouting wings and flying to the moon, the expressivist were to pull off anything like what I've suggested, it would still only have been shown that, putting it somewhat abstractly, removing the puzzle from puzzlingly privileged x's was a matter of showing a special relation between x's and non-puzzlingly privileged y's; but that doesn't show that the privilege of the x's derives from the privilege of the y's. Now, while I've already warned against being disappointed in advance about the usefulness of an as yet merely notional result, I think a more direct response is in order. The response is simply that what this kind of thing *does* show is that our *understanding* of the privilege of the x's derives from our *understanding* of the privilege of the y's. And if our understanding of the privilege of the y's can be called Φ'istic, and if our

[138] Of course, such 'being used to it' is more or less the extent to which most people 'understand' how wireless technology works. "How are you controlling that car with that thing?" "Simple, you just push these buttons and move these knobs."

understanding of the privilege of the x's derives from our Φ'istic understanding of the privilege of the y's, then I don't see why we shouldn't say we have a Φ'istic understanding of the privilege of the x's. And it's only in this sense that, I should think, an expressivist would ever want to or need to say that, on an expressivistic understanding, the privilege *itself* of the x's (privileged-while-non-expressive self-ascriptions) derives from the privilege of the y's (privileged-because-expressive self-ascriptions).

So, finally, what about the plausibility of actually achieving any of this? Needless to say, I think chances are grim. We've already seen (in discussing the question of truth-values) that the expressivist's hopes are held hostage to the fates of a handful of certain highly unpopular, but at least pretty well familiar, Wittgensteinian views about the workings of language; now we see them held hostage to the fates of, let's be modest, a dozen more as yet undreamt of Wittgensteinian insights. Add to that the fact that Wittgenstein studies seem to have everything to do with getting straight on his method (quite rightly, of course), but precious little to do with continuing in it, and I think it's safe to say that no plausible expressivist line on the problem of first-person privilege that countenances privileged while non-expressive self-ascriptions will emerge in our lifetime. Perhaps there's some significantly more straightforward way than I can imagine of showing non-expressive privilege to derive from expressive privilege; but if not, as I say, an expressivism admitting privileged while non-expressive self-ascriptions has barely just begun to take shape as a viable option for a solution to the problem of the general phenomena of first-person privilege.

Loose Ends

Before turning at last to bring this investigation to a conclusion, I want to spend just a little space making good on some promises from chapter one. In the course of arguing that the phenomena of privilege are very real and so the need for a satisfying account of them just as real, I deferred discussion of a few questions that should now be returned to. First, there was the promise to discuss the compatibility of expressivist theories of privilege with available solutions to the mind-body problem. Second, there was a promise to discuss how concerns to do with the relational individuation of contentful mental states can be accommodated by an expressivist's understanding of privilege. Last there was the promise to discuss the implications of expressivism for foundationalist projects in epistemology. There won't be too much to say about any of these, but it's worth taking just a little space to see how expressivism as I see it will connect up with these classic concerns. I'll take them in order.

About the mind-body problem, if what I've said about the necessity for a Wittgensteinian anti-Augustinianism is correct, it should be clear that either some kind of anti-realism or some kind of quietism about the reality or otherwise of the mental will be enforced. Bar-On, who actually wanted to provide for privileged self-knowledge, of course isn't concerned that if she doesn't adopt some kind of deflationism about reference to the mental (and thus, of the ontology of the mental) then she'll be forced to deal with the question of self-knowledge as something substantial. On the question of the ontology of the mental, then, she wants to retain a materialist "ontology" while giving up on a materialist "ideology". (Bar-On, 2004, p.

425) This is pretty well in line with standard materialist solutions to the mind-body problem, certainly, and it will probably be tempting to take the Wittgensteinian line I see the expressivist must adopt as itself, if ever plausibly worked out, coming to something like the same thing. In any case, it should be clear that dualism is anathema to the expressivist. Indeed the driving motivation for the expressivist has to be just that one can do full justice to the phenomena of privilege without accepting a dualistic ontology of mind and body.

About the relational individuation of mental states and how this could possibly be compatible with even a quite fallible means of introspecting them, since Donald Davidson (Davidson, 1987) and Tyler Burge (Burge, 1988) weighed in on the subject there's been a strong suspicion that it's precisely the *perceptual* notion of introspection, and thus what introspective knowledge amounts to, that causes the trouble.[139] Paul Boghossian (Boghossian, 1989), brought the question down explicitly to the question of the substantiality of self-knowledge and argued that, given substantiality, there is indeed a tension between self-knowledge and the relational individuation of content. We've seen that the expressivist has available to her to deny the substantiality of self-knowledge, however, and that she despises the perceptual notion of introspection in particular. If Boghossian, Davidson and Burge are right, then, the expressivist is perhaps rather well placed to reconcile content externalism with one's ability to speak authoritatively on his own states of mind. But

[139] Specifically, the contemporary concern is with the sort of externalist theories of the individuation of content set forth in, for instance, (Putnam, 1981, ch. 1) and (Burge, 1979), and not specifically with the behavioristic theories of mind that were under discussion in chapter one. The concern, though, is general, of course, and has to do with any theory that individuates the contents of mental states in terms of their relations to things one wouldn't suspect should be available to introspection.

the only point that really needs to made is this: if in the final analysis the externalist really is committed to one's being able to infer from the fact that one is thinking of water, for instance, that there are molecules of hydrogen and oxygen combined in such-and-such ways, then *obviously* externalism is confused.[140] I hope for the externalist's sake, this is not a consequence of her view; but it's not the job of the avowal expressivist to sort out whether it is or isn't.

Lastly, about the implications of expressivism about privilege for foundationalist projects in epistemology, I think there is perhaps a misconception in the readiness with which some will tend to claim that expressivism is sure to undermine "the endeavor to provide foundations for our empirical knowledge in our knowledge of how things subjectively appear to us to be." (Hacker, 1975, p. 253) The reason is that, while I think there is a fundamental tension between avowal expressivism and foundationalist projects, I don't think this is coming from the simple fact that expressivists would, as it were, leave self-knowledge something quite hollow. The reason I don't think it's coming from there is that, it seems to me the real project of (classical) foundationalism is to found our empirical knowledge on something one can speak about *free from doubt and error*; that is, if the foundationalist had any hope of tracing one's knowledge that the door is closed, for instance, back to one's *freedom from doubt and error* concerning a host of facts about things like how brownly and blackly one is appeared to, I don't think the foundationalist would much mind whether an expressivistic account of that freedom

[140] To that extent I'd be in agreement with an internalist like Richard Fumerton who rejects externalism on such grounds. (Fumerton, 2003) But it doesn't seem to me anyone but a self-styled internalist would see externalists as committed to such a thesis. If it's complained that their not being so committed means that it simply isn't *clear* what 'content externalism' amounts to, then I suppose I can accept that point.

were appropriate or weren't. This is just the sort of thing I take Chisholm to be getting at when he says that "these things I have described by saying that they 'justify themselves' may also be described by saying that they are 'neither justified nor unjustified.'" (Chisholm, 1982, p 137) The point is, there is, as Hacker put it, a "singularity in the epistemology of the psychological at the point of the first-person present tense of certain verbs and phrases" (Hacker, 2005, p. 271) no matter how you slice it. It seems to me the tension between expressivism and foundationalism just isn't that the foundationalist thought there was knowledge at the foundations while the expressivist insists there isn't. The tension is rather that the expressivist *takes for granted* certain things like *behavioral expressions* of mentality, statements about which the foundationalist would only *derive from* statements about things like how brownly or blackly she's appeared to. This is where the tension lies, and it's a real one.

Concluding Remarks

In chapter two I said the task of this dissertation was to evaluate expressivist theories of privilege on two scores. An expressivist theory was to be evaluated, first, for whether it's plausible in its own right; and second, for whether it resembles simple expressivism enough to be considered the radically non-Cartesian sort of theory that makes the expressivist insight attractive in the first place. I now want to draw from the results of the foregoing discussions in order to bring to completion the evaluative task of this dissertation.

On the question of truth-aptness, a conservative expressivism, which denies truth-aptness, is highly implausible, though, of course, very much within the

expressivist spirit. A progressive expressivism, which admits truth-aptness, is, just for doing that, immediately highly more plausible. But just for doing that, we saw, a progressive expressivism incurs the debt of rendering plausible some theory of the world-representing aspects of avowals as themselves 'fused with' their expressive aspects; otherwise, we saw, it would be too difficult not to inquire after their epistemic credentials. Thus, progressive expressivisms are in this regard as implausible as deflationary views of world-representation in general, and bear as well the added implausibility of such views with respect to the psychological in particular. As for keeping with the expressivist insight, a truth-aptness admitting expressivism seems to me quite faithful indeed.

On the question of self-knowledge, we saw that a conservative expressivism is only as plausible as Hacker's defense against the battery of considerations that would make such a view seem absurd. It didn't seem to me there was anything wrong with this defense that wasn't wrong with armchair philosophy in the first place. Such a view is, obviously, quite in keeping with the expressivist spirit. The progressive expressivisms considered were two: a say-what-you-want expressivism and an avowed-state-justifies-avowal expressivism. We saw that both were meant to accrue plausibility only insofar as attention was directed to an expressivistic treatment of the phenomena of privilege, which are, after all, quite discernible from the third-person point of view. We saw in this their structural similarity with the conservative approach. And in this respect they seem themselves rather within the spirit of the initial expressivist insight. However, we saw for the last of these, the explicitly epistemic view, that if it should really be, unlike the others, a distinctively progressive

expressivism because genuinely epistemic, then it should to that extent undermine the point of expressivism about the phenomena of privilege and therefore depart quite drastically from the spirit informing the expressivist insight in the first place. Furthermore, its plausibility as a genuinely epistemic view of privileged self-knowledge was, we saw, seriously suspect anyway.

On the question of non-expressive though privileged mental state self-ascriptions, we saw that conservative expressivism, even if it should extend the scope of what could be considered expressive to its farthest conceivable limit, still stopped short, specifically on the question of negative avowals. In this it was judged unacceptable as a theory of first-person privilege; though, certainly, its fidelity to the spirit of the expressivist insight is beyond question. Progressive expressivism on this question can only be judged with respect to its fidelity to the expressivist insight, because its plausibility or otherwise only comes out as we see it dealing with the last (the next) of the four questions anchoring our investigation. And with respect to fidelity to the expressivist insight, it seems to me progressive expressivism on this third question only fares as well as it can plausibly keep with the expressivist insight in answering the fourth question.

The fourth question, on whether the expressivist could relate the privilege across the 'spectrum of cases' ultimately 'downward' to and deriving from the privilege attaching to the most nearly expressive cases, we saw was a particularly difficult question for the expressivist. The only expressivist option was to seek to arrive at the right kind of relation, it was clear, otherwise expressivism simply wouldn't be a theory of the full range of privileged phenomena. And since it seems

to me highly unlikely that one should end up with a different theory for the privilege of negative avowals than the theory for their positive counterparts, and since it seems to me highly unlikely that a theory of privilege for just those cases at the lowest end of the spectrum of privilege should be all that interesting, it seems to me a theory of the full range of privileged phenomena is indeed what the expressivist ought to be out provide. In doing so, it should be unquestionably in keeping with the expressivist insight. Whether it can plausibly do so, we saw, depended on whether it could muster a dozen or so Wittgensteinian insights that would help us to see how, because the cases of privilege at the high end of the spectrum lose their point in absence of the cases of privilege at the low end, therefore the privilege at the high end itself derives from the privilege at the low end, and thus the expressivist insight would play enough of a role in understanding the full range of privileged phenomena that the resulting understanding could indeed be called an expressivistic one. This seemed rather hopeless, however, since expressivism should then be held hostage not just to the plausibility of the most basic Wittgensteinian ideas about language use, but to the eventual uncovering of so many as yet undreamt of Wittgensteinian insights.

Such are the prospects for expressivist theories of first-person privilege. Without space in this work to evaluate the prospects for rival theories of privilege, I'll just say that it seems to me expressivism stands out as the most promising.

BIBLIOGRAPHY

Alston, William. 1965. Expressing. In *Philosophy in America*, (Ed.) M. Black, pp. 15-34. Ithaca: Cornell University Press.

Alston, William. 1971. Varieties of Privileged Access. *American Philosophical Quarterly*, pp. 223-241.

Armstrong, David. 1963. Is Introspective Knowledge Incorrigible. *The Philosophical Review*, pp. 417-432.

Aune, Bruce. 1965. On the Complexity of Avowals. In *Philosophy in America*, (Ed.) M. Black, pp. 35-54. Ithaca: Cornell University Press.

Aune, Bruce. 1967. *Knowledge, Mind, and Nature.* New York: Random House.

Austin, John. 1962. *How to do things with Words*, (Eds.) J. Urmson, & M. Sbisa. Oxford: Clarendon Press.

Ayer, Alfred. 1940. *The Foundations of Empirical Knowledge.* New York: Macmillam & Co.

Ayer, Alfred. 1952. *Language Truth and Logic* (2nd Edition). Mineola: Dover Publications.

Ayer, Alfred. 1963. Privacy. In *The Concept of a Person* (pp. 52-81). London: Mamillan.

Baker, Gordon, and Peter Hacker. 1980. *Wittgenstein Meaning and Understanding.* Chicago: University of Chicago Press.

Baker, Gordon. 2004. *Wittgenstein's Method, Neglected Aspects*, (Ed.) K. Morris. Malden: Blackwell.

Bar-On, Dorit. 2004. *Speaking My Mind.* New York: Oxford University Press.

Bar-On, Dorit. 2009. First-Person Authority: Dualism, Constitutivism and Neo-Expressivism. *Erkenntnis*, pp. 53-71.

Berker, Selim. 2008. Luminosity Regained. *Philosopher's Imprint*, pp. 1-22.

Blackburn, Simon. 1984. *Spreading the Word.* New York: Oxford University Press.

Blackburn, Simon. 1998. Wittgenstein, Rorty and Minimalism. *Mind*, pp. 157-181.

Block, Ned. 1994. Consciousness. In *A Companion to the Philosophy of Mind*, (Ed.) S. Guttenplan, pp. 210-219. Cambridge: Blackwell.

Block, N. 1995. On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences*, pp. 227-47.

Boghossian, Paul. 1989. Content and Self-Knowledge. *Philosophical Topics*, pp. 5-26.

Boghossian, Paul. 1990. The Status of Content. *The Philosophical Review*, pp. 157-184.

Bonjour, Laurence. 1978. Can Empirical Knowledge Have a Foundation? *American Philosophical Quarterly*, pp. 1-13.

Bonjour, Laurence. 2001. Toward a Defense of Empirical Foundationalism. In *Resurrecting Old-Fashioned Foundationalism*, pp. 21-40. Lanham: Rowman & Littlefield.

Boyle, Matthew. 2010. Bar-On on Self-Knowledge and Expression. *Acta Analytica*, pp. 9-20.

Burge, Tyler. 1979. Individualism and the Mental. *Midwest Studies in Philosophy*, pp. 73-121.

Burge, Tyler. 1988. Individualism and Self-Knowledge. *The Journal of Philosophy*, pp. 649-663.

Candlish, S., and George Wrisley. (2008). Private Language. http://plato.stanford.edu/entries/private-language/ (accessed June 20, 2010)

Chisolm, Roderick. 1982. *The Foundations of Knowing*. Minneapolis, MN: University of Minnesota Press.

Chrisman, Matthew. 2009. Expressivism, Truth, and (Self-)Knowledge. *Philosophers' Imprint*, pp. 1-26.

Conant, James. 2002. The Method of the Tractatus. In *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, (Ed.) E. Reck, pp. 374-462. New York: Oxford University Press.

Crary, Alice. 2000. Introduction. In *The New Wittgenstein*, (Eds.) A. Crary, and R. Read, pp. 1-18. New York: Routledge.

Davidson, Donald. 1987. Knowing One's Own Mind. *The Proceedings and Addresses of the American Philosophical Association*, pp. 441-458.

Dennett, Daniel. 2002. Seeing is Beleiving--Or is it? In *Vision and Mind*, (Eds.) *A.* Noë, and E. Thompson, pp. 481-496. Cambridge: The MIT Press.

Dennett, Daniel. 2007. Philosophy as Naive Anthropology. In *Neuroscience & Philosophy*, pp. 73-96. New York: Columbia University Press.

Dummett, Micheal. 1993. What does the Appeal to Use Do for the Theory of Meaning? In *The Seas of Language*, pp. 106-116. New York: Oxford University Press.

Evans, Gareth. 1982. *The Varieties of Reference.* Oxford: Oxford University Press.

Finkelstein, David. 2001. Wittgenstein's "Plan for the Treatment of Psychological Concepts". In *Wittgenstein in America*, (Eds.) T. McCarthy, and S. C. Stidd, pp. 215-36. Oxford: Clarendon Press.

Finkelstein, David. 2003. *Expression and the Inner.* Cambridge: Harvard University Press.

Fogelin, Robert. 1987. *Wittgenstein.* London: Routledge.

Fogelin, Robert. 2009. *Taking Wittgenstein at his Word.* Princeton: Princeton University Press.

Fumerton, Richard. 2003. Introspection and Internalism. In *New Essays on Semantic Externalism and Self-Knowledge*, (Ed.) S. Nuccatelli, pp. 257-276. Cambridge: The MIT Press.

Fumerton, Richard. 2006. *Epistemology.* Malden: Blackwell Publishing.

Fumerton, Richard. 2008. The Problem of the Criterion. In *Oxford Handbook on Skepticism*, (Ed.) J. Grecco, pp. 34-52. New York: Oxford University Press.

Gasking, David, and Alan Jackson. 1967. Wittgenstein as Teacher. In *Ludwig Wittgenstein: the Man and his Philosophy*, (Ed.) K. Fann, pp. 49-55. New York: Dell Publishing Co.

Gertler, Brie. 2001. Introspecting Phenomenal States. *International Phenomenological Society*, pp. 305-328.

Gertler, Brie. 2008. Self Knowledge.
http://plato.stanford.edu/archives/win2008/entries/self-knowledge
(accessed June 30, 2010)

Glock, Hans. 1996. Necessity and Normativity. In *The Cambridge Companion to Wittgenstein*, (Eds.) D. Stern, and H. Sluga, pp. 86-100. New York: Cambridge University Press.

Grahek, Nikola. 2007. *Feeling Pain and Being in Pain.* Cambridge: The MIT Press.

Hacker, Peter. 1975. *Insight and Illusion* (1st Edition). New York: Oxford University Press.

Hacker, Peter. 1986. *Insight and Illusion* (Revised Edition). New York: Oxford University Press.

Hacker, Peter. 1990. *Wittgenstein Meaning and Mind.* Malden: Blackwell Publishers Inc.

Hacker, Peter. 2000. Was he trying to whistle it? In *The New Wittgenstein*, (Eds.) A Crary, and R. Read, (pp. 353-388). New York: Routledge.

Hacker, P. M. S. (2003). Wittgenstein, Carnap and the New American Wittgensteinians. *Philosophical Quarterly* , 1-23.

Hacker, P. M. S. (2005). Of Knowledge and of Knowing that Someone is in Pain. In A.
Pichler, & S. Saatela (Eds.), *Wittgenstein: The Philosopher and his Works* (pp. 244-270). Bergen: The Wittgenstein Archives and the University of Bergen.

Hacker, Peter. 2007. Gordon Baker's Late Interpretation of Wittgenstein. In *Wittgenstein and his Interpreters*, (Eds.) G. Kahana, E. Kanterian, and O. Kuusela, pp. 88-122. Malden: Blackwell.

Horwich, Paul. 1998. *Meaning.* New York: Oxford University Press.

Horwich, Paul. 2006. A World Without Isms. In *Truth and Realism*, (Eds.) M. Lynch, and P. Greenough, pp. 188-203. New York: Oxford University Press.

Jackson, Frank. 1967. A Note on Incorrigibility and Authority. *Australasian Journal of Philosophy*, pp. 358-363.

Jackson, Frank. 1973. Is there a good argument against the incorrigibility thesis? *Australasian Journal of Philosophy*, pp. 51-62.

Jacobsen, Rockney. 1996. Wittgenstein on Self-Knowledge and Self-Expression. *The Philosophical Quarterly*, pp. 12-30.

Kahane, Guy, Oskari Kuusela, and Edward Kanterian. (2007). Introduction. In *Wittgenstein and his Interpreters*, (Eds.) G. Kahane, O. Kuusela, & E. Kanterian, pp. 1-36. Malden: Blackwell.

Kuusela, Oskari. 2005. From Metaphysics and Philosophical Theses to Grammar: Wittgenstein's Turn. *Philosophical Investigations*, pp. 95-133.

Kuusela, Oskari. 2006. Do the Concepts of Grammar and Use in Wittgenstein Articulate a Theory of Language or Meaning? *Philosophical Investigations*, pp. 309-341.

Malcolm, Norman. 1966. Wittgenstein's Philosophical Investigations. In *Wittgentsein The Philosophical Investigations*, (Ed.) G. Pitcher, pp. 65-103. Garden City: Anchor Books.

Malcolm, Norman. 1978. Wittgenstein's Conception of First-Person Psychological Sentences as "Expressions". *Philosophic Exchange*, pp. 59-72.

Malcolm, Norman. 2001. *Ludwig Wittgenstein: A Memoir* (2nd ed.). New York: Oxford University Press.

Mays, Walter. 1967. Recollections of Wittgenstein. In *Ludwig Wittgenstein: the Man and his Philosophy*, (Ed.) K. Fann, pp. 79-88. New York: Dell Publishing Company.

McDowell, John. 1996. *Mind and World.* Cambridge: Harvard University Press.

Miller, Alexander. 2001. On Wright's Argument against Deflationism. *The Philosophical Quarterly*, pp. 527-531.

Mulhall, Stephen. 2007. *Wittgenstein's Private Language.* New York: Oxford University Press.

Peacocke, Christopher. 1998. Conscious Attitudes, Attention, and Self-Knowledge. In *Knowing Our Own Minds*, (Eds.) C. Wright, B. Smith, and C. MacDonald, pp. 63-98. New York: Oxford University Press.

Putnam, Hilary. 1980. Brains and Behavior. In *Readings in Philosophy of Psychology*, (Ed.) N. Block, (Vol. I, pp. 24-36). Cambridge: Harvard University Press.

Putnam, Hilary. 1981. *Reason, Truth and History.* New York: Cambridge University Press.

Quine, Willard. 1960. *Word and Object.* Cambridge: MIT Press.

Quine, Willard. 1970. *Philosophy of Logic* (2nd edition.). Cambridge: Harvard University Press.

Quine, Willard. 1976. Carnap and Logical Truth. In *Ways of Paradox*, pp. 107-132. Cambridge: Harvard University Press.

Quine, Willard. 1990. *The Roots of Reference.* La Salle: Open Court Publishing Company.

Rorty, Richard. 1970. Incorrigibility as the Mark of the Mental. *The Journal of Philosophy*, pp. 399-424.

Rorty, Richard. 1965. Mind-Body Identity, Privacy, and Categories. *The Review of Metaphysics*, pp. 24-54.

Russell, Bertrand. 1921. *The Analysis of Mind.* New York: Humanities Press.

Ryle, Gilbert. 1949. *The Concept of Mind* (2002 edition.). Chicago: New University of Chicago Press.

Searle, John. 1969. *Speech Acts.* New York: Cambridge University Press.

Sellars, Willfrid. 1969. Language as Thought and as Communication. *Philosophy and Phenomenological Research*, pp. 506-27.

Sellars, Willfird. 1997. *Empiricism and the Philosophy of Mind.* Cambridge: Harvard University Press.

Shoemaker, Sydney. 1994a. Self-Knowledge and "Inner Sense": Lecture I: The Object Perception Model. *Philosophy and Phenomenological Research*, pp. 249-69.

Shoemaker, Sydney. 1994b. Self-Knowledge and "Inner Sense": Lecture II: The Broad Perceptual Model. *International Phenomenological Society*, pp. 271-290.

Shoemaker, Sydney. 1994c. Self-Reference and Self-Awareness. In *Self-Knowledge*, (Ed.) B. Gertler, pp. 80-93. New York: Oxford University Press.

Stern, David. 2004. *Wittgenstein's Philosophical Investigations.* New York: Cambridge University Press.

Stevenson, Charles. 1995. The Emotive Meaning of Ethical Terms. In *20th Century Ethical Theory*, (Ed.) S. Cahn, pp. 116-128. Upper Saddle River: Prentice-Hall.

Stoljar, Daniel. 2007. *The Deflationary Theory of Truth.*
http://plato.stanford.edu/entries/truth-deflationary/
(accessed August 10, 2010)

Strawson, Peter. (1966). Review of Wittgenstein's Philosophical Investigations. In *Wittgenstein: The Philosophical Investigations*, (Ed.) G. Pitcher, pp. 22-64. Garden City: Anchor Books.

Tomberlin, James. 1968. The Expression Theory of Avowals. *Philosophy and Phenomenological Research*, pp. 91-96.

Williamson, Timothy. 2000. *Knowledge and its Limits.* New York: Oxford University Press.

Wittgenstein, Ludwig. 1969. *The Blue and Brown Books.* New York: Barnes and Noble.

Wittgenstein, Ludwig. 1970. *Zettel.* (Eds.) G. Anscombe, and G. Von Wright. Berkeley: University of California Press.

Wittgenstein, Ludwig. 1974. *Philosophische Grammatik*, (Ed.) R. Rhees, (Trans.) A. Kenny). Oxford: Blackwell.

Wittgenstein, Ludwig. 1976. *Wittenstein's Lectures on the Foundations of Mathematics.* Hassocks: The Harvester Press Ltd.

Wittgenstein, Ludwig. 1979. *Wittgenstein's Lectures Cambridge, 1932-1935*, (Ed.) A Ambrose. Totowa: Rowman and Littlefield.

Wittgenstein, Ludwig. 1980a. *Remarks on the Philosophy of Psychology* (Vol. II). (Eds.) G. Von Wright, and H. Nyman, (Trans.) C. Luckhardt, and M Aue. Malden: Blackwell.

Wittgenstein, Ludwig. 1980b. *Remarks on the Philosophy of Psychology* (Vol. I). (Eds.) G. Anscombe, and G. Von Wright, (Trans.) G. E. Anscombe. Malden: Blackwell.

Wittgenstein, Ludwig. 1993. *Philosophical Occasions.* (Eds.) J. Klagge, and A. Nordmann. Indianapolis: Hackett Publishing Company.

Wittgenstein, Ludwig. (2005). *The Big Typescript.* Malden: Blackwell.

Wittgenstein, Ludwig. 2009a. *Philosophical Investigations.* (Trans.) G. Anscombe, P. Hacker, and J. Schulte. Malden: Blackwell.

Wittgenstein, Ludwig. 2009b. Philosophy of Psychology - A Fragment. In L. Wittgenstein, *Philosophical Investigations*, (Trans.) G. Anscombe, P. Hacker, and J. Schulte, pp. 182-244. Malden: Blackwell.

Wright, Crispin. 1992. *Truth and Objectivity.* Cambridge: Harvard University Press.

Wright, Crispin. 1998. Self-Knowledge: The Wittgensteinian Legacy. In *Knowing Our Own Minds*, (Eds.) C. Wright, B. Smith, and C. Macdonald, (pp. 13-46). Oxford: Clarendon Press.

Wright, Crispin. 2003a. *Saving the Differences: essays on themes from Truth and Objectivity*. Cambridge: Harvard University Press.

Wright, Crispin. 2003b. Comrades against Quietism. In C. Wright, *Saving the Differences: essays on themes from Truth and Objectivity*, pp. 125-150. Cambridge: Harvard University Press.

Wright, Crispin. 2003c. What Could Anti-Realism about Ordinary Pyschology Possibly Be? In C. Wright, *Saving the Differences: essays on themes from Truth and Objectivity*, pp. 407-42. Cambridge: Harvard University Press.