# CHAPTER 6

# Rules, Privacy, and Physicalism

JIM HOPKINS

The main themes of Wittgenstein's later philosophy are particularly concentrated in his remarks on sensation and private language. Among these are the role of rules in thought and language; the role of "pictures" in generating philosophical problems; the aspects of our "form of life" which render our practices possible; and his far-reaching critique of what might be called the Cartesian picture of the mind. As I hope to indicate, Wittgenstein's work on these topics can be seen as consonant with more recent philosophical thinking about the mind, and also as providing ways to extend it.

I hope also to support the idea that there is another aspect to Wittgenstein's work, linked with the ideal of self-knowledge which has been central to philosophy from its beginnings. It might be said that in Wittgenstein's late work, "Know thyself" becomes "Understand your own philosophical errors and confusions." So he writes in such a way as to evoke and intensify his reader's allegiance to sources of error and confusion, at the same time as criticizing them. As long as the reader retains these allegiances, he will think that it is Wittgenstein who is mistaken, confused, or putting in question what is obviously true. This also is particularly salient in the remarks on sensation (cf. the "But you will surely admit that there is a difference…" or the "And yet you again and again reach the conclusion…" of PI §304); and I hope in passing to suggest something of its role in the remarks on rules as well.

Wittgenstein's repeated evoking and dispelling of confusion and accusation in this way is part of a pattern—together with direct statements, such as the "And your scruples are misunderstandings" of PI §120, or the "Now, ask yourself: what do you *know* about these things" of PI §158—in which he establishes a particular relation of authority in regard to his reader. Negotiating this relation, in turn, is an important part of learning—or failing to learn—from Wittgenstein's work. (Thus some readers lose patience and turn away; others submit by trying to assume—and so inevitably

failing to earn—what they take to be a similar position themselves; and so on.) This is a matter of independent interest, which we will not pursue here.

## 1. THOUGHT AND CONTENT

To compare Wittgenstein's ideas with some now current it will be useful to begin with the *Tractatus*. One achievement of this book was to concretize the account of thought which Wittgenstein had encountered in the "great works of Frege." In 1897 Frege had contrasted the objectivity of thoughts with the subjectivity and privacy of sensations, ideas, and other mental items. He argued that "a man never has somebody else's mental image, but only his own; and nobody even knows how far his image (say) of red agrees with somebody else's." As regards thoughts, by contrast, "it is quite otherwise…[since] one and the same thought can be grasped by many men" (Frege 1952, 79).

This antipsychologism left it unclear how thought (or Frege's act of "grasping" a thought) might actually take place. Wittgenstein made good this lack, by framing an account of thought as realized by internal representations. These had as yet unknown "psychical constituents" with "the same sort of relation to reality as words"; and they were mapped to possible states of affairs in complex abstract relations which "it would be a matter of psychology to find out."[1] This was part of a more general account of representation, according to which all representations were such "pictures," and hence among the facts of the world.

On this account a person who thinks (or believes, intends, desires, etc.) that P, does so by framing a thought-token which bears a referential and structural relation to the state of affairs in which P.[2] As Wittgenstein says at TLP 5.542, constructions such as "A believes that P," "A has the thought that P," and "A says P" all involve "the correlation of facts by means of the correlations of their objects." These "correlations of facts" hold between the psychical facts constituting mental representations on the one hand, and the possible states of affairs represented by them on the other. Thus the mental representation which realizes the thought that I add 2 does so because the "psychical constituents" which make it up are correlated with such things as myself

---

1. His fuller account, as expressed in a letter answering questions by Russell, goes as follows:

> "But a Gedanke [thought] is a Tatsache [fact]: what are its constituents and components, and what is their relation to those of the pictured Tatsache?" I don't know *what* the constituents of a thought are but I know *that* it must have such constituents which correspond to the words of Language.… [The] kind of relation of the constituents of the thought and of the pictured fact…would be a matter of psychology to find out…"Does a Gedanke [thought] consist of words?" No! But of psychical constituents that have the same sort of relation to reality as words. What those constituents are I don't know. (NB App. III, 130–131)

2. This partly relies on a further feature of the *Tractatus*, namely that the correlation of thoughts and sentences with *possible* states of affairs is effected via their correlation with *actual* objects, as analysis would ultimately reveal.

and the act of adding 2, so that the thought as a whole can be correlated with the situation in which I add 2. When many men grasp or express the same Fregean thought—when they all have or express the (abstract) thought that P—this is because their individual thought-tokens are related to the same state of affairs—the state of affairs in which P—in the same referential and structural ways.

We can summarize this by saying that as 5.542 suggests, the *Tractatus* account of thought provides an approach to content as a whole. We describe a range of phenomena, including intention, thought, and meaning, in a way which we can regard as assigning a (propositional or sentential) content to the items described. Thus we use a sentence "P" ("I add 2") to describe a thought as the thought that P (the thought that I add 2), and we apply this mode of description to a range of related phenomena. We speak of an utterance or sentence as meaning or saying that P (My utterance of "I add 2" says or means that I add 2), and likewise of a perception as a perception that P, of an intention, hope, fear, or desire as the intention, hope, fear, or desire that P, and so on. All these phenomena, on the *Tractatus* account, are unified and explained by their relation to the mental representation of the possible situation that P, which in one way or another constitutes the thought, perception, meaning, intention, and so on. The relevant representations and relations, moreover, are (at least in principle) empirically accessible: for they are among the facts of the world, about which psychology might hope to find out.

These content-sustaining correlations extend to "A says P," and hence to the linguistic expression of thought, because of the particular role which Wittgenstein assigns to thought in animating language. He takes it that in *meaning* or *understanding* a sentence we think of it as relating to things in a certain way, and in doing this we confer meaning upon it. As he describes the idea in the *Investigations*: "'Thought must be something unique.' When we say, and *mean*, that such-and-such is the case, we—and our meaning—do not stop anywhere short of the fact; but we mean: *this— is—so*" (§95). Thus for example we might utter the sentence "Russell befriended Wittgenstein" *meaning* by "Russell" the man Bertrand Russell, by "befriended" the (past) act of befriending, by "Wittgenstein" the man Ludwig Wittgenstein, and by the order in which we utter these words that Russell was the befriender and Wittgenstein the befriended of the two. In this we would "use the perceptual sign…(spoken or written, etc.) as a projection of a possible situation," namely that in which Russell befriended Wittgenstein, and so would mean that *this-was-so*. The same holds for *understanding* the sentence, as uttered by another. In both cases we "think the sentence-sense" ("Denken des Satz-Sinnes," TLP 3.11) and thereby determine the conditions in which the sentence, as we mean or understand it, is true.

This illustrates how the "correlations of facts by means of the correlations of their objects" mentioned in TLP 5.542 constitute not only our thoughts but also the knowledge of language which enables us to communicate and coordinate our actions. And this also entails that according to the *Tractatus* the use of a word or sentence as we observe it in linguistic practice does *not* constitute its meaning. For this use is both animated and fixed by something which is less public and more basic, namely the acts

of meaning or understanding—the "thinking the sense"—which assign to these signs the referents and truth-conditions in accord with which we then proceed to use them.[3] So as Wittgenstein says in the *Blue Book* meaning and understanding seem "processes through which alone language can function," and without which "the signs of our language seem dead" (BB, 3). Again, as Wittgenstein says in "a mythological description" in the *Investigations*: "The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space" (§219). It is clear that if such a *stamping with a particular meaning* is done by a mental act or process, then this act or process will fix what the correct use of the rule is to be.

The idea that we mentally establish the relation which makes signs meaningful in this way naturally goes with another, namely that our mental act somehow links signs with a family of abstract objects. (In the *Tractatus* the link is with actual objects and their properties, and, via these, with possible objects in their combinations in states of affairs. In other accounts the abstract objects differ, encompassing senses, possible worlds, numbers, and so on.) This is no accident: for on the one hand only such abstract objects have the greater-than-empirical range and rigidity which we associate with meaning; and on the other the only contact we can have with such objects—and hence with the meanings we might suppose them to constitute—is via the remarkable reach of thought.[4]

## 2. THE PROBLEM OF REPRESENTATION

As is well known, a conception akin to the *Tractatus* account of mental representation was later spelled out in detail by Fodor and others, as part of an attempt to understand the mind as a symbol-processing mechanism. Having attained this view Fodor realized that it left a basic question unanswered. As he said, "We must now face what has always been *the* problem for representational theories to solve: what relates internal representations to the world? . . . I take it that this problem is now the main content of the philosophy of mind."[5]

Wittgenstein had recorded a similar realization nearly half a century before— and prior to the development of the digital computer—in relation to his *Tractatus* account. Thus in *Philosophical Grammar* I, §§62–63 he wrote:

> 62 "That's *him*" (this picture represents *him*) – that contains the whole problem of representation.

3. This is closely related to the conception Kaplan describes as subjectivist semantics. See Almog, Perry, and Wettstein 1989, 600.

4. Despite Wittgenstein's emphasis on the "hardness" and "crystalline purity" or the order of possibilities as conceived in the *Tractatus*, the role of abstract objects is minimized. Thoughts are naturally connected with actual concrete objects, and nonexistent possible objects are recombinations of the components of these actual ones. This eliminates the need for abstract objects which exist independently of concrete ones.

5. Fodor 1981, 61. See also the lucid exposition in Fodor 1990.

…

I have the intention of carrying out a particular task and I make a plan. The plan in my mind is supposed to consist in my seeing myself acting thus and so. But how do I know, that it is myself that I'm seeing? Well, it isn't myself, but a kind of picture. But why do I call it the picture of *me*?

…

"I thought Napoleon was crowned in the year 1805."…What connection is there between your thought and Napoleon? – It may be, for example, that the word "Napoleon" occurs in the expression of my thought, plus the connection that word has with its bearer; e.g. that was the way he signed his name, that was how he was spoken to and so on.

"But when you utter the word 'Napoleon' you designate that man and no other" – "How then does this act of designating work, in your view? Is it instantaneous? Or does it take time?" – "But after all if someone asks you 'did you mean the very man who won the battle of Austerlitz' you will say 'yes'. So you meant that man *when you uttered the sentence*." Yes, but only in the kind of way that I then knew also that $6 \times 6 = 36$.

The answer "I meant the victor of Austerlitz" is a new step in our calculus. The past tense is deceptive, because it looks as if it was giving a description of what went on "inside me" while I was uttering the sentence.

("But I meant *him*". A strange process, this meaning! Can you mean in Europe someone who's in America? Even if he no longer exists?)

63 Misled by our grammar, we are tempted to ask "*How* does one think a proposition, *how* does one expect such and such to happen? (how does one do that?)"

"How does thought work, how does it use its expression?" – This question looks like "How does a Jacquard loom work, how does it use the cards".…

"How does thought manage to represent?" – the answer might be "Don't you really know? You certainly see it when you think." For nothing is concealed.

How does a sentence do it? Nothing is hidden.…

We are not in the realm of causal explanations, and every such explanation sounds trivial for our purposes. (Cf. BB, 117–118)

These remarks indicate both Wittgenstein's dissatisfaction with his *Tractatus* approach and how he sought to improve it by moving toward that of the *Investigations*. In PG I, §62 he takes up the *Tractatus* explication of intention. My intention is realized by a *Tractatus* picture: a mental representation of myself performing the intended action, which serves as a blueprint for my doing so. (As he says in PR §21, "What is essential to intention is the picture: the picture of what is intended.") In fulfilling my intention I follow out the rules which link the constituents of the picture—including those which represent myself—to the objects and situations for which they stand, and so bring about the situation I originally intended. (As he says at PG I, §58, "intention reaches up to the paradigm and contains a general rule.")

Moreover, as §63 makes clear, Wittgenstein also envisaged an account of the mind as a symbol-processing mechanism. The processes by which a representation

which realizes an intention actually produces the intended action might be compared to those by which the punched cards which direct a Jacquard loom actually govern the weaving of the pattern they also represent. This provides a simple kind of model, to which Wittgenstein often returned, for the way mental representation as conceived in the *Tractatus* might admit of psychological investigation.

Yet all this, as Wittgenstein urges in §62, still leaves us facing "the whole problem of representation." For in approaching thought in this way we presuppose, but do not explain, the most basic connection: the content-fixing link between the components of thought and things in the world. Without an understanding of this, we also have no account of the acts of thought (here that of meaning Napoleon by "Napoleon") by which we assign meaning to words and sentences; and such an act in any case seems mysterious and problematic.

As §63 stresses, we are tempted to suppose that we might learn about these things by investigating the physical symbolic realization of thought, or its causal role. This, however, is not so. For as Wittgenstein had earlier remarked, making use of another symbol-processing mechanism:

> But one might say something like this. The sentences that we utter have a particular purpose, they are to produce certain effects. They are parts of a mechanism, perhaps a psychological mechanism, and the words of the sentences are also parts of the mechanism (levers, cogwheels, and so on). The example that seems to illustrate what we're thinking of here is an automatic music player, a pianola. . . . And so shouldn't we say that the sense of the sign is its effect? – But suppose the pianola is in bad condition and the signs on the roll produce hisses and bangs instead of the notes. – Perhaps you will say that the sense of the signs is their effect on a mechanism in good condition, and correspondingly that the sense of an order is its effect on an obedient man. But what is regarded as a criterion of obedience here? (PG I, §33)

Here Wittgenstein is concerned with a particular variant of the *Tractatus* account of sense as a relation between a sentence and a possible situation: one in which this relation is fixed by causal or functional role. On such an account the sense of a sentence used as an order will be understood in terms of its effect on the behaviour of another, and the sense of an inner representation in terms of its effect on the behaviour of the agent whose representation it is. (The two are presumably connected via the idea that the order achieves its sense-constituting effect on behaviour by causing an inner representation which realizes an intention to act in accord with the order, which in turn causes the obedience itself. So in this case both the external and the internal symbols would relate to the same situation, and so would have the same sense.) Hence, as he adds in the next paragraph,

> when we spoke of "the sense of the signs" . . . [we] were thinking rather of the purpose of the signs *within* the mechanism of the pianola. – And so you can say that the purpose of an order is its sense, only so far as the purpose can be expressed by a rule of language. (PG I, §33)

Such a causal account cannot solve the "problem of representation," since a sign can have sense but fail to have the *right* effects. The effects which we might take to be sense-constituting are those which obtain just when the agent acts *in accord with* the order, or again (to relate the example of the pianola here to §§62–63 above) when the inner representation supposed to constitute an intention causes an action in which the intention is *fulfilled*. So it is this relation of accord (obedience, fulfilment, etc.), which requires investigation: and this relation, as it appears, is expressed by a linguistic rule (e.g., a "grammatical" rule such as he describes at PI §458: "If an order runs 'Do such-and-such' then executing the order is called 'doing such-and-such.' ").

This, however, also indicates the kind of solution to these problems he now entertains. He takes it that the role he has previously given to thought as linking words and things (the meaning Napoleon by "Napoleon" in PG I, §62 above) requires to be replaced by that of the practice of language itself. (What thought could not reach out to by itself, as one might say, can nonetheless be part of the social practice of using language, which is actually extended in both space and time, and further than any single individual can reach.) Likewise he seeks to understand the relation of accord which holds between intention and action, order and obedience, and so on, in terms of a rule for the use of language. As regards the explication of content, as we might say, it seems that where in the *Tractatus* thought about language was, there in the *Investigations* the practice of using languages is to be.

Accordingly we see that the acts of meaning imagined in the *Tractatus* can also be located in the account of language with which Wittgenstein begins PI §1. Augustine describes how he observed his elders refer to objects and grasped "that the thing was called by the sound they uttered when they *meant* to point it out." In this we might take Augustine to be understanding that the elders *meant* particular things by their words, as one might understand that a speaker means Napoleon by "Napoleon." Learning the public use of words thus appears here as learning that others mean particular things by them, so that one can go on to mean the same things by them oneself.

Wittgenstein does not, of course, deny that we intend or mean things by our words. Rather he argues that such intentions presuppose the ability to engage in the practice of language (see, e.g., PI §§190, 205), so that we misconstrue them if we take them either to impart meaning to such practice or to explain our ability to engage in it. So in this first remark he also gives an example of practice—that of the shopkeeper and his use of words—and afterwards describes its basic explanatory role.

> "But how does he know where and how he is to look up the word 'red' and what he is to do with the word 'five'?"——Well, I assume that he *acts* as I have described. Explanations come to an end somewhere.—But what is the meaning of the word "five"?—No such thing was in question here, only how the word "five" is used. (PI §1)

Here he indicates both the basic role of practice (and in relation to reference and the postulation of abstract objects) and also briefly poses the questions he will raise

and answer about it when he comes to consider acting in accord with a rule. And shortly after introducing the notion of practice, he starts explicitly to question the idea of mental processes which might be supposed to give meaning to signs in it, and so would fix both meaning and correctness in use. Thus in PI §19 he asks "how do you do this: how do you *mean that* while you *say* 'Slab!'?" and in §22 he cautions (using words closer to the *Tractatus* and *Philosophical Grammar*) that "'*meaning*' (thinking) the sentence" is not a kind of act. He continues by introducing an example of *interpretation* in §32, and in §34, stressing that "neither the expression 'to intend the definition in such-and-such a way' nor the expression 'to interpret the definition in such-and-such a way' stands for a process which accompanies the giving and hearing." As we will see, this will prove a key notion in his argument.

## 3. CONTENT AND PRACTICE

For present purposes we can describe this change as follows. The relation of thought to its objects can be regarded as a form of mental reference, and that of accord (as between intention and action, etc.) as a psychological analogue of truth. (As the thought that P—the thought described by "P"—is *true* just if P, so the intention that P—the intention described by "P"—is fulfilled just if P; likewise the order that P is obeyed just if P, the hope, fear, or expectation that P realized just if P, etc.) So as the truth of a sentence secures the reference of its terms, this sort of accord secures the kind of mental reference with which Wittgenstein was concerned. (Since the reference to myself in my intention is ensured by the fulfilment of the intention—the intention will not be fulfilled unless *I* act accordingly—this notion of reference can be explicated together with the notion of accord to which it is related.) But further, concentrating on this notion of accord, and in the particular case of following a rule, provides an alternative approach to thought, and one which renders speculation about mental representation and its relation to objects irrelevant.

In Kantian terms we can regard thinking or judging as applying a concept in accord with a rule, and hence as a particular case of following a rule. We use the concept C to think of things *as C's*, as we do when we think, concerning some item, *this is C*. The ability to think of things in such a way is shown in the ability to describe them in the same way, as we might by saying "This is C." In this we follow a rule of language: a rule for correct description, where description of this kind reflects a correct use of a concept, or the correct formation of a thought. The same holds in analogous cases, for example where someone responds to "Add 2" by adding 2, or regards an intention described by "Add 2" as fulfilled if he has done so. In investigating such instances of following a rule, therefore, we investigate the correctness of thought together with the practice of using language in accord with rules. And in this, as Wittgenstein says, nothing is hidden. We already assess our linguistic activities for correctness, so it is our public practice of assessment that we require to understand.

Thus the approach of the *Tractatus*, in which thought and content are explicated by inner representation, yields to that of the *Investigations*, in which they are explicated by the following of rules in linguistic practice. So Wittgenstein's question of PI §198,

"But how can a rule shew me what I have to do at *this* point? Whatever I do is, on some interpretation, in accord with the rule"[6]

addresses, among other things, the correctness of my own thoughts and my first-person account of them, including their relation to objects; and similarly his question of §206,

what if one person reacts in one way and another in another to the order and the training? Which one is right?

addresses the correctness of thought in general, including my own, as raised in §198 via the notion of interpretation.

Wittgenstein also makes the connection between concepts and rules explicit in his discussion. He often treats connected questions in succeeding remarks, and after raising the question of rules and correctness in §206 he indicates his answer there and in more detail in §207. Then in §208 he makes clear that these remarks both relate to concepts, since to master a concept is to acquire the ability to act in accord with a rule.

208. Then am I defining "order" and "rule" by means of "regularity"?—How do I explain the meaning of "regular", "uniform", "same" to anyone?—I shall explain these words to someone who, say, only speaks French by means of the corresponding French words. But if a person has not yet got the *concepts*, I shall teach him how to use the words by means of *examples* and by *practice*.—And when I do this I do not communicate less to him than I know myself.

This also appears in details of his argument. For example in his discussion of the private use of a word "S" he has his interlocutor (who advocates the notion) exclaim:

260. "Well, I *believe* that this is the sensation S again."

And he replies

Perhaps you *believe* that you believe it!

---

6. Here as elsewhere I have replaced Anscombe's "shew" by "show."

Here the reply turns on the idea that the use of a word which was private in the sense under consideration would not accord with a rule, and so would not express a concept. In the absence of the concept S, there is no real thinking of an item as S, and so no real belief that I have the sensation S. One might believe that one had such a belief; but this would be mistaken.

## 4. EXAMINING CORRECTNESS

In accord with this exegesis we can see Wittgenstein's argument in the *Investigations* as having, among others, two connected aspects. First, he seeks to establish that linguistic meaning is constituted in linguistic practice, so that the ability to engage in such practice is basic to such states or acts as understanding or grasping meaning as well. Second, he seeks to explicate what such practice involves, and how correctness is determined within it. In both aspects of this argument, as we have noted, he opposes the basic role of practice to that of mental acts or processes, such as the "thinking the sense" of the *Tractatus*, which might seem to provide a link to an alternative source of meaning, and so would relegate practice to a secondary role.

As we have noted, such meaning-constituting processes would have two connected features. They would animate practice, giving otherwise dead signs the life we find in use (Wittgenstein mentions this succinctly at PI §403); and they would also determine what the correct practice in using signs was to be. In the *Tractatus*, as we saw, this idea took the form of the mental determining of reference and truth-conditions; in the *Investigations* it takes the form of the mental determining of correct use more generally. So as against this idea Wittgenstein argues in a number of cases that there can be no such mental determination of use.

Thus in PI §138 he asks how a Fregean grasping of a sense can determine a use which is extended in time. In considering this he urges that any mental act—such as my picturing a cube in connection with the word "cube"—admits of further interpretation as regards its content, and so could not "force a particular use on me" (§140). Rather, it seems, the correctness of my use, as otherwise ascertained, seems to determine whether my initial understanding itself has been correct. A comparable argument applies to the state of understanding itself, conceived as the source of correct use (§146); to the event of coming to understand (§147); and to the "state of a mental apparatus (perhaps of the brain)" which might be supposed to underlie this (§149). (This last argument repeats points considered in *Philosophical Grammar* above, for the "mental apparatus" is modelled on the pianola at PI §157.) This is how it would be if the meaning of signs was actually constituted by their use, rather than "breathed into it" (PI §432) from such an external source.

We can follow this contrast through to the remarks in which Wittgenstein raises the basic questions of correctness via the notion of interpretation. He initially approaches this via the example of a pupil being taught to continue a series by

adding 2. In his first use of the example he assumes that the way we are trying to teach the pupil to go on is correct: "he continues the series correctly, that is, as we do it" (§145). In the next phase of his argument he brings this into question, by pointing out that it is possible that a pupil who *seems to understand* the teaching and examples as we do might actually not do so.

> Now we get the pupil to continue a series (say + 2) beyond 1000—and he writes 1000, 1004, 1008, 1012.
>
> We say to him: "Look what you've done!"—He doesn't understand. We say: "You were meant to add *two*: look how you began the series!"—He answers: "Yes, isn't it right? I thought that was how I was *meant* to do it."———Or suppose he pointed to the series and said: "But I went on in the same way."—It would now be no use to say: "But can't you see . . . . ?"—and repeat the old examples and explanations.—In such a case we might say, perhaps: It comes natural to this person to understand our order with our explanations as *we* should understand the order: "Add 2 up to 1000, 4 up to 2000, 6 up to 3000 and so on." (§185)

In this the teacher does not at first interpret the pupil's responses correctly, nor does the pupil rightly understand the examples and the teaching, although it seems to each that this is so. So the question arises as to how we know that we understand what we take to be our own rules correctly, as opposed to merely seeming to ourselves to do so. And since these include the rules which constitute our concepts, the question at once becomes radical. So as Wittgenstein says in his next remark:

> To carry [the order "+2"] out correctly! How is it decided what is the right step to take at any particular stage?—"The right step is the one that accords with the order—as it was *meant*." . . . But that is just what is in question: what, at any stage, does follow from that sentence [from "Add 2" meant as an order]. Or, again, what, at any stage we are to call "being in accord" with that sentence (and with the *mean*-ing you then put into the sentence—whatever that may have consisted in). (§186)

Here the question carries to linguistically expressed thought in general: to the "being in accord" with a sentence which we find between rule and instance, premises and conclusion, intention and action, order and execution, and all the rest. The question is both epistemic and constitutive, and relates to Wittgenstein's basic notion of practice as well as to the mental acts he criticizes as seeming to animate or determine it. How do I know that I am following a particular rule—using a particular concept—correctly, as opposed to merely seeming to myself to do so? And how is such correctness constituted? (Cf. the "what this going-by-the-sign really consists in" of §198.)

In order to connect these questions with more familiar work in philosophy, we might say that they concern the opposition between *appearance* and *reality*, as applied both to meaning and to thought. We are clear that it *seems* or *appears* to us

that our sentences have meaning and our thoughts and experiences have content. But do we really think and mean as we take ourselves to? If meaning or content is real, what makes it so, and how do we distinguish this reality from mere appearance or seeming? These questions are deep, and as Wittgenstein raises them they seem particularly hard to answer. For of course—since these are exactly the claims we have been asked to justify—we cannot simply answer that we *know* this reality, or that we have first-person authority with respect to it, in our intuitive knowledge of what we think and mean.

Deprived of this kind of answer, we feel at a loss; for what we seem required to justify includes the traditional and supposedly indubitable starting points of philosophical enquiry (Descartes's *cogito*, the empiricists' immediate knowledge of impressions and ideas, Kant's "I think"), and hence the foundations apparently required for any philosophical investigation, including that of thought and meaning themselves. And as Saul Kripke observed in his pioneering exposition, if we take our inability to answer Wittgenstein's questions as grounds for scepticism, "It seems that the entire idea of meaning vanishes into thin air." We seem left with the mere appearance of meaning, and so threatened with taking "*all* language, *all* concept formation, to be impossible, even unintelligible," so that "assertions that anyone ever means anything are meaningless." Such scepticism, as Kripke says, would be "insane and intolerable."[7] So once aware of these questions, how should we address them?

## 5. ESTABLISHING CORRECTNESS

Despite the depth and apparent difficulty of Wittgenstein's questions, his answers to them are relatively straightforward. The difficulty we encounter in his text arises not so much from his positive account, as from the way he uses his remarks to expose points of potential confusion. To bring this out let us first take a basic but relatively simple part of his response to the questions, and then consider the rhetoric by which he treats them in more detail.

From early in the *Investigations* Wittgenstein notes that we can approach questions as to what rule a person is following by a kind of explicit interpretive observation, in which we discern and describe regularities in a person's behaviour. (Cf. the "am I defining 'order' and 'rule' by means of 'regularity'?" quoted from §208 in the previous section.) As we shall see, he discusses this most fully in §§206–207, where he uses the observation that "the common behaviour of mankind is the system of reference by means of which we interpret an unknown language" (§206)

---

7. Kripke 1982. The quotations are from pages 22, 62, 77, and 60 respectively. I trust it will be clear from what follows that my own account differs from Kripke's; but his recognition of the way meaning threatens to vanish under the impact of Wittgenstein's arguments, as considered in the discussion of PI §198 below, marks him as one of the few who have really reckoned with their force.

to explain what constitutes following a rule correctly. Prior to this, however, he has already introduced the idea in some detail.

Thus in §32 he considers how someone observing others in a strange land will have to guess at the meaning of the ostensive definitions the people try to give him there, and will "guess sometimes right, sometimes wrong." In §54 he says that we can say that a game is played in accord with definite rules when an observer "can read these rules off from the practice of the game—like a natural law governing the play." And he extends this to language by adding that such an observer can also distinguish between mistakes and correct play, since "There are characteristic signs of it in the players' behaviour. Think of the behaviour characteristic of someone correcting a slip of the tongue. It would be possible to recognize that someone was doing so even without knowing his language."[8] Again in §82 he asks, "What do I call 'the rule by which he proceeds'?" and immediately offers as a preliminary answer "The hypothesis that satisfactorily describes his use of words, which we observe . . . "[9] Later he extends this to going on with a number series by framing a "hypothesis" about the other's use in algebraic terms (§151); to considering which way of going on from examples is the correct one (§§206–207); and to interpreting the speech and other actions of people foreign to us, even if they talk only in monologue (§243). In these cases, as other remarks make clear, considering such interpretive observation is a way of making explicit the *criteria* we use in ascribing understanding, intending, and related states. (See for example §269.)

As Wittgenstein also considers (§§32, 82, 185, 207), we can go wrong in trying to understand another in this way. First, we may simply be mistaken as to what rule another is following—we may in effect frame the wrong hypothesis as to the regularity which the other's behaviour actually displays. Second, we may be mistaken in supposing that someone is following a rule at all: despite initial appearances her behaviour in using a term may be disordered or random. In the first kind of case we may correct our hypothesis, and in the second we may find that no hypotheses proves satisfactory. (And of course, as Wittgenstein notes in §163, the presence of regularity, and hence the satisfactoriness of any such hypothesis, may itself be a matter of degree. For while there are strict rules of language, as the example of the number series makes clear, these are not always obeyed, or rightly characterized, by those who obey them; and other cases, as he notes in §83, we may proceed without rules, or make them up or change them as we go along.)

The argument about correctness which begins at §185 also turns on such a mistaken hypothesis. The reader observing the pupil's responses of "2," "4," "6," and so on at first construes them as (hypothesizes that they are) instances of obedience

---

8. This example as well as that of the reading-machines (PI §157) indicates that Wittgenstein is including in his consideration rules which we learn to follow without ever learning to formulate, and hence without articulate beliefs as to what rules we are following.

9. As he says in *The Big Typescript*, 199e: "I have indicated two ways of ascertaining the rule according to which he acts. One of them, the hypothetical, consisted in observing his actions, and then the rule was of the same kind as a proposition of science."

to the order to add 2. In this case, taking the example in terms of what Wittgenstein says in §82, this would seem to be "the hypothesis that satisfactorily describes his [the pupil's] use of words, which we observe"; and here also "Add 2" would be "the [sign] which he [the pupil] gives us in reply if we ask him what his rule is." This hypothetical construal, however, proves to be unsatisfactory, as emerges when the pupil continues past 1,000. Still, as Wittgenstein says in §82, the teacher *might* be able "to see [a] clear rule," which he could describe by a further hypothesis ("Add 2 up to 1,000, 4 up to 2,000, etc."); and this might prove correct, in the sense that it would be borne out by further observation. This shows both that someone might take teaching and examples differently than we do, and also that just as different hypotheses represent different events as in accord with natural laws, so different hypotheses about the rule someone is following represent different behaviours as in accord with these rules. And again, these are both aspects of regarding the understanding of what rule a person is following as a kind of discerning of order in behaviour ("like a natural law governing the play"). We do not know in advance how much order we will discern, or what this order will be; and we may fail to discern it correctly.

But of course in our own case we do not observe ourselves or interpret our own words in this way. We apply the criteria which such observation makes explicit in the case of others, but not ourselves (§377). As subjects and agents we simply use our words and concepts: we spontaneously think and act in accord with them, and take ourselves to do so correctly. So without interpretation we already know what we think and mean, and this first-person authority extends to the contents of our mental states and acts more generally. But should this come seriously into question for us—should I have to ask myself, as prompted by §§185–186, which of two competing formulations rightly describes the rule I am following, or again what this rule requires of me at some particular point—then interpretation in my own case would provide no help. For since any interpretation I might offer is part of my own practice, it "still hangs in the air along with what it interprets, and cannot give it any support" (§198). This also applies to any justification I might attempt for my judgments as to what I intend or mean, or for the way I follow a rule, or for holding that I am doing so correctly. If what is in question is my capacity to think correctly, and if my attempts at justification are exercises of the same capacity as is in question, then such attempts are circular. So here, as Wittgenstein says "my spade is turned," and I am inclined to say "This is simply what I do" (§217).

We can thus see this part of Wittgenstein's argument as focusing on an asymmetry regarding interpretation as between the first- and third-person cases. In the third-person case interpretation (as a form of investigation which proceeds by the tacit specification and confirmation of interpretive hypotheses) enables us to describe what rule a person is following, and hence what he must do to follow it. In the first-person case, in which my interpretation does not serve as such an hypothesis about my own practice, but rather continues and amplifies that practice, my interpretation "hangs in the air with what it interprets." Insofar as we fail to

distinguish these very different cases we may take Wittgenstein as arguing that as regards rule-following *all* interpretive understanding hangs in the air, and thus as advocating the kind of scepticism Kripke describes. But if we resist this conflation, then we can also see that the difference in perspective highlighted by his remarks also provides straightforward answers to the questions he raises.

For roughly, what makes it the case that my understanding of my own concepts (rules, etc.) in my own case is correct, as opposed to merely seeming so to me, is that I admit of understanding by others (via my participation in "the common behaviour of mankind" which, as Wittgenstein says in §206, is "the system of reference by means of which we interpret an unknown language"). This entails that there are regularities in my verbal and nonverbal behaviour which others might interpret, so that in this they might understand me as thinking and acting in accord with the same concepts (rules) as I take myself to, and as doing so correctly. So the fully objective correctness which I presuppose but cannot justify in my own case, can nonetheless be justified on my behalf by the interpretive understanding of others.

Such justification does not require that others should actually understand me or share my concepts, but only that this should be possible. For insofar as it is possible that another should succeed in understanding me as correctly thinking as I take myself to do, then in fact I do think that way, and correctly. (So, as Wittgenstein notes at §213, this approach does not require that we ever actually interpret one another in such an explicit and hypothetical way.) The reality as opposed to the appearance of meaning—and the justification of our claims about this reality, including those as to the meanings of our sentences and the contents of our thoughts—is constituted by a discernable order in our behaviour, which we can make articulate by interpretation.

## 6. PERSPECTIVE AND PUZZLEMENT

If this is right then a main reason for finding Wittgenstein's discussion puzzling is a failure to distinguish the role of interpretation in one's own case from that in the case of others. Wittgenstein in effect sets the stage for this, by introducing two quite distinct conceptions of interpretation, which we can describe as first and third person. The third-person conception is that discussed in section 4 above, and concerns the interpretive understanding of one person by *another*. In interpreting another in this sense, I specify what that *other* means, and so speak in the third person. By contrast with this third-person conception of interpretation there is also a first-person one. For I may also specify what I *myself* mean, as when I say "By 'Moses' I mean such-and-such a man." As noted in section 3 above, Wittgenstein relates this to the idea of meaning as a mental state or process which gives meaning to a word (or stamps a rule with a particular meaning), and thereby determines what its correct use will be. He argues against this by stressing that what "comes before the mind" in meaning or understanding does not "force a use on us" since we

can always interpret or specify it in more than one way. This again is the first-person conception, that of *interpretation in one's own case.*

Although these conceptions are clearly distinct, Wittgenstein anticipates that his reader will be inclined to conflate them, via the idea that in understanding either my own words or those of another I give them meaning by the same mental processes. So for example in learning to use a rule from *another* I become able to interpret the rule—to stamp it with meaning *in my own case*—in the same way as the other does, and hence in the way I might read from his behaviour in understanding him.[10] Thus at successive stages in his argument Wittgenstein represents both conceptions of interpretation in the same way, that is, by means of an algebraic formula. Thus in PI §146 he implies that in thinking of understanding as "the *source* of the correct use" we are thinking of it as analogous to the grasp of such a formula, and remarks:

> But this is where we were before. The point is, we can think of more than *one* application of an algebraic formula; and every type of application can in turn be formulated algebraically; but naturally this does not get us any further.—The application is still a criterion of understanding.

This turns on the first-person conception: if I think of my own state of understanding as like my grasp of an algebraic formula, I can see that this state, like the formula itself before my mind, might be interpreted in more than one way, and so fails to determine what use on my part will actually be correct. (And rather than my understanding determining what use is correct, the correctness of my use determines whether I have actually understood.) But then in §151 Wittgenstein considers an individual B who tries to understand what rule *another* individual A is following, and then says that he knows how to go on.

> What happened here? Various things may have happened; for example, while A was slowly putting one number after another, B was occupied with trying various algebraic formulae on the numbers which had been written down. After A had written the number 19 B tried the formula $a_n = n^2 + n - 1$; and the next number confirmed his hypothesis.

This is the third-person conception. The aim, as in §82, is to find "the rule by which he proceeds," and so "the hypothesis that satisfactorily describes his use of words, which we observe." As Wittgenstein here notes, we can frame such hypotheses and confirm them like any others. True, the predictions of such hypotheses are potentially infinite, and competing hypotheses might yield the same predictions over any finite range. But this applies to hypotheses in general, and presents no special difficulty in the case of meaning—unless we are inclined to suppose that the fact that we

---

10. Thus compare §213 with the "Now, *how* was it possible for the rule to have been given an interpretation during instruction, an interpretation which reaches as far as to any arbitrary step?" from RFM VI, §38.

mean such-and-such things, or obey such-and-such rules, cannot be a mere matter of fact, and so answerable to observation and hypothesis in this kind of way.

So when at §213 Wittgenstein says that "this initial segment of a series obviously admitted of various interpretations (e.g. by means of algebraic expressions)" his use of "interpretation" is ambiguous. Indeed he characteristically deploys both conceptions in the same remark (§146 begins with the third person and moves to the first, while §151 can be seen as starting with the first and moving to the third), leaving the reader to distinguish their application. This again is not straightforward, for much of his argument holds for both. As §213 illustrates, since both can be exemplified by the use of an algebraic series, both are subject to the claim that "we can think of more than one application of an algebraic formula, and every application can in turn be formulated algebraically." This in turn entails that on both conceptions, as Wittgenstein's interlocutor complains at §198, "Whatever I do is, on some interpretation, in accord with the rule." Again, it entails that on both, as Wittgenstein says at §201, "if everything can be made out to accord with the rule, then it can also be made out to conflict with it." Despite these similarities, however, the two conceptions have different roles and fates in Wittgenstein's argument. The first-person conception is shown to fail to sustain meaning on its own, and to be significant only as it works in coordination with the other; the third-person conception is elaborated as part of an account of correctness in practice, in which the first-person conception also plays a role, as the preliminary sketch in section 4 indicates.

In that sketch we noted that Wittgenstein's account required his reader to distinguish the role of interpretation in the first- and third-person perspectives. We now see that to do this in the context of Wittgenstein's argument the reader must also distinguish the first- and third-person conceptions of interpretation which he introduces side by side. Accordingly his interlocutor struggles particularly intensely with this at §198:

> "But how can a rule shew me what I have to do at *this* point? Whatever I do is, on some interpretation, in accord with the rule."—That is not what we ought to say, but rather: any interpretation still hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning.
>
> "Then can whatever I do be brought into accord with the rule?"—Let me ask this: what has the expression of a rule—say a sign-post—got to do with my actions? What sort of connexion is there here?—Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it.
>
> But that is only to give a causal connexion; to tell how it has come about that we now go by the sign-post; not what this going-by-the-sign really consists in. On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom.

Here the interlocutor's "Whatever I do is, on some interpretation, in accord with the rule" apparently conflates the two conceptions via their role in §185. The

example of the pupil in that remark clearly raises the initial question "How can a rule show me what I have to do at this point?" For the pupil has *apparently* been using the rule "Add 2" correctly up to "1,004"; and as *he* has understood the rule this was, at that point, the right thing to say. The pupil has, however, learned that there is another interpretation of the rule—on which, at that same point, he should instead have written "1,002"—and this was apparently the interpretation the teacher had in mind all the time. Also, as the interlocutor is aware, when we consider a formula such as "Add 2" as "before the mind" of either the teacher or the pupil, we can see that there is more than one way to apply it, as is exemplified here.

So on the first-person conception the argument so far seems to imply that the responses "1,002" and "1,004" can equally be justified, and that this can be generalized, since "Whatever I do is, on some interpretation, in accord with the rule." But also the interlocutor has seen that when the pupil produced a response to "Add 2" which *apparently* conflicts with the rule, the teacher might still provide an interpretation ("Add 2 up to 1,000, 4 up to 2,000, etc.") on which this response would be correct. So on the third-person conception too it appears that "Whatever I do can be brought into accord with the rule," which is how the interlocutor puts his question the second time. Thus the interlocutor's successive expressions of puzzlement seem themselves to mark the conflation, by the way the first (with "show me") alludes to the first-person case, whereas the second (with "brought into accord") fits interpretation by another as well.[11]

The interlocutor's questions thus show that he has been participating fully in Wittgenstein's dialectic so far. His fault lies only in not yet distinguishing clearly between the two conceptions of interpretation in play, and so in not yet seeing—as Wittgenstein will bring out in the remarks that follow—that they yield different answers to his final question. But at this stage, and on both conceptions, his questions can be answered as Wittgenstein initially replies. On both, interpretation in my own case "still hangs in the air with what it interprets, and cannot give it any support"; and on both "Interpretations do not by themselves determine meaning." (Third-person interpretations may *ascribe* meaning in describing the use of words, but as in the case of laws of nature it is the real regularity under description, and not the describing of it, which determines the meaning so ascribed.)

It is thus appropriate that Wittgenstein's interlocutor—or the reader who has followed the argument with equal tenacity—should now struggle with the way the constraints of meaning seem in danger of vanishing. (Kripke apparently did so, when his way of seeing Wittgenstein's argument as putting these constraints in question "struck [him] with the force of a revelation" (1982, 1).) And the problem highlighted

---

11. The same tacit shift in perspective occurs more clearly *Remarks on the Foundations of Mathematics*: "How can the word 'Slab' indicate what I have to do, when after all I can bring any action into accord with any interpretation? How can I follow a rule, when after all whatever I do can be interpreted as following it?" (VI, §38). Here the "I can bring any action…" involves interpretation by myself, as opposed to "whatever I do *can be interpreted…*" which suggests interpretation by another.

by Kripke is not lessened but made more acute by the "answer" presented in §201, namely that "if everything can be made out to accord with the rule, then it can also be made out to conflict with it. So there would be neither accord nor conflict here." For if there were "neither accord nor conflict" with a rule then meaning really would have vanished. To have meaning we require accord and conflict, and these must not be just apparent but also real. Hence Wittgenstein's reminder in the next remark that "to *think* one is obeying a rule is not to obey a rule" (§202). He notes that we cannot make this distinction "privately"; but how the distinction is made *non-*"privately," and indeed what privacy is supposed to be, has so far not been explained.

As subsequent remarks indicate, the interlocutor has yet to see that although on both conceptions of interpretation there is a sense in which whatever I do can be brought into accord with a rule, on neither is this actually relevant to my linguistic practice. It is not relevant as regards the first person, because in practice "there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call 'obeying the rule' and 'going against it' in actual cases" (§201). And it is not relevant as regards the third person, for another who is seeking to understand the rule I am following in practice will perforce frame hypotheses which *satisfactorily* describe my utterances and other actions. This requires formulations which specify both accord and conflict, so as to admit confirmation and disconfirmation by what actually I say and do. (Thus the teacher in §185 was able to see his error, and consider another hypothesis, precisely because his initial supposition was disconfirmed at "1,004.") By considering such hypotheses, particularly in the context of an explorer who seeks to interpret a language strange to him, we can see how our understanding of one another can be regarded as having foundations which are not abstract but rather concrete and empirical.

Also we can see that these foundations enable us to extend such understanding even to others who are not members of our community and never speak to us. For as Wittgenstein says at §243, we can also imagine observing and interpreting

> human beings who spoke only in monologue; who accompanied their activities by talking to themselves.—An explorer who watched them and listened to their talk might succeed in translating their language into ours. (This would enable him to predict these people's actions correctly, for he also hears them making resolutions and decisions.)

## 7. INTERPRETATION, ORDER, AND OBJECTIVITY

Wittgenstein makes explicit both this notion of interpretation and the kind or order he takes it to involve in PI §§206–207.

> 206. Following a rule is analogous to obeying an order. We are trained to do so; we react to an order in a particular way. But what if one person reacts in one way and another in another to the order and the training? Which one is right?

Suppose you came as an explorer into an unknown country with a language quite strange to you. In what circumstances would you say that the people there gave orders, understood them, obeyed them, rebelled against them, and so on?

The common behaviour of mankind is the system of reference by means of which we interpret an unknown language.

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

Are we to say that these people have a language: orders, reports, and the rest?

There is not enough regularity for us to call it "language".

The first paragraph of these remarks makes clear that Wittgenstein intends them to address the question of correctness, as raised by the pupil who reacts differently in §185. He does this by taking up the interpretation of an "articulate language." This links his consideration of interpretation with the topic of thought, for he takes human thinking to have a complexity (or "multiplicity") which the sentences of such a language specify,[12] and which we cannot ascribe in their absence (§§25, 32, 342ff., 650). As emphasized in the *Tractatus*, the combinatory syntax of human language ("part of the human organism, and no less complex than it," TLP 4.002) provides a potential infinity of distinct but interrelated sentences, capable of sustaining the assignment of distinct but interrelated contents to an unlimited range of utterances, actions, and states of mind.

Still, as PI §207 argues, even such significantly patterned linguistic tokens cannot be understood in isolation from the activity and environment of the agent who produces them. (Trying to do so would be like trying to learn a foreign language by listening to the radio while having no information as to what the broadcasts were about.) So while deeds without words are inarticulate, words without deeds would be unintelligible. If the two are "regularly connected," however, we can understand each by relation to the other. This is because—as Augustine observes in describing

---

12. Thus in *Philosophical Remarks* he says:

I only use the terms the expectation, thought, wish, etc., that *p* will be the case, for processes having the multiplicity that finds expression in *p*, and thus only if they are *articulated*. But in that case they are what I call the interpretation of signs.

I only call an *articulated* process a thought: You could therefore say 'only what has an articulated expression'.

(Salivation – no matter how precisely measured – is *not* what I call expecting.) (PR §32)

And in this we may note also his use of "interpretation of signs," as relating to the *Tractatus* idea that such processes involve a sign which receives the projection of a possible situation.

bodily movement as "the natural language of mankind"—we can understand much nonverbal action without relying on speech. This enables us to tie the complex structure of language as it appears in utterance to further points in action and context, and so to interpret sentences and words; and by this means we also understand nonverbal action as informed by thoughts which, like sentences, have fully articulate content.

We can see this clearly in the example of the monologue people, whose verbal expressions of intention ("resolutions and decisions") are regularly connected with actions which accord with them. In understanding nonverbal bodily movements as actions—and hence, as Wittgenstein says, as "intelligible" and "logical"—we relate them to intentions in ordered series. (Thus we might observe someone moving in certain ways, with the intention of arranging leaves, sticks, and other burnable things; this with the intention of making a fire; this with the intention of cooking food; this with the intention of serving a cooked meal; and so on. As Anscombe says, all these ascriptions show "an order that is there" in the movements we observe.)[13] So we can imagine an explorer-interpreter A observing an interpretee B who in monologue uses sentences $\Sigma_1, \Sigma_2, \ldots \Sigma_n$ to express intentions relating to everyday actions. In finding B's actions intelligible A relates each to a series of intentions which A articulates in A's own language, say by $S_1, S_2 \ldots S_n$. In each instance of such understanding, therefore, A is in a position to hypothesize that some of A's own sentences (and hence words) have the same use in articulating intentions as some of B's, the use of which he observes; and so also that these sentences (and words) have the same use as A's own in specifying the observable conditions in which these intentions are fulfilled (and again in specifying B's beliefs as to whether this is so). As Wittgenstein notes, this could enable A to translate B's sentences into his own, and also to predict what B will do on the basis of what B says.[14]

Such predictions would depend on three hypotheses: (1) that A rightly understands the intentions in B's nonverbal actions; (2) that B uses his sentences to express these intentions correctly (and hence with first-person authority); and (3) that A also rightly understands these sentences as B uses them. Failure in prediction should be traceable to falsity in one of these hypotheses, and success tends to confirm all three together, for the intentions and sentences concerned. In this hypothetical procedure, moreover, nothing is assumed or taken on trust. As A tests A's understanding of B's nonverbal actions against A's understanding of B's expressions

---

13. I have discussed this order and the nature of hypotheses related to it in more detail in Hopkins 1999b. I relate it to psychoanalytic interpretation in Hopkins 1999a. These and other papers of mine are available via the website of the Department of Philosophy, King's College London, www.kcl.ac.uk.

14. Wittgenstein's emphasis on the relation of verbal and nonverbal behaviour also appears in his Cambridge lectures. See for example "if that were the rule instead of the exception, if there were a race of men who always walked straight out of the room whenever they said 'I intend to play chess'—would we still say they used the phrase in the same way we do?" (LFM, 25).

of intention, A's confidence in A's own understanding is based on A's ability to predict what B does, and A's confidence in B's first-person authority is based on its coincidence with A's own and independent understanding of the utterances and actions in which B expresses it. So given initial footholds in A's natural understanding of intention and B's observable and authoritative use of sentences in expressing it, this interpretive procedure would be both thoroughly empirical and capable of precise predictive testing.

In acting in accord with their own expressions of intention the monologue people are already acting in accord with sentences, and so in the relevant sense they are already following rules (cf. the "'being in accord' with that sentence" of Wittgenstein's original question at §186). Also, as indicated in the remarks on writing numerals in series (cf. §§143–145 as well as §§185ff.) we can readily imagine the same scenario as extended to expressions of intention such as "I will add 2" and verbal actions which accord with them, such as uttering or writing "2" followed by "4" followed by "6"… For clearly the same kind of "regular connections" as hold between verbal expressions of intention and nonverbal actions also hold for verbal actions as well. And as §§206–207 make clear, the same considerations apply beyond the monologic case. In particular we can see that A's understanding of B is partly comparable with our understanding of one another, for it is especially important to us that we can predict the actions of others on the basis of what they say.

This means that we can also use the example of A and B above to illustrate how the "regular connections" spoken of in §§206–207 secure mutual understanding in our own case. We constantly coordinate our own actions with those of others by assuming that if they are sincere in their expressions of intention and able to act as they intend, then they will act in accord with the intentions they express. (Without such coordination our actions would be liable to "fall into confusion," as Wittgenstein "feels like putting it" in §207.) In this again we rely on forms of the three hypotheses above: that we can recognize the intentions on which others act, that they express these intentions with authority, and that we understand their expressions correctly. So our successes in everyday coordination (as well as failures which we can trace to inability or insincerity on the part of others) constantly but tacitly testify that these hypotheses are correct, and hence that others use the same sentences to express their intentions, and so relate these same sentences to the world, as we do. Such coordination, indeed, is an explicit feature of Wittgenstein's examples of rules as orders (or imperatives more generally); for in these cases one person acts in accord with a sentence provided by *another*, thus showing how the "regular connections" of §§206–207 effect coordinations between utterances and actions which are interpersonal and so fully social.

In such interactions—as opposed to those of the monologue people—we both respond to the words of other and speak to them in turn. In adding this to our account, we also see how Wittgenstein takes the two notions of interpretation in his discussion to work in harmony. Thus, again, as Wittgenstein says in §87, I can explain what I mean by "Moses" by saying "I take 'Moses' to mean the man, if there was such a man, who led the Israelites out of Egypt…" Another seeking to

understand me might claim that this explanation is not final, and that it could not be completed: "As though an explanation as it were hung in the air unless supported by another one." This, however, is a misapprehension. For, as Wittgenstein says, "whereas an explanation may indeed rest on another one that has been given, but none stands in need of another—unless *we* require it to prevent a misunderstanding." So where we use our capacity to specify what we mean in this way—that is, to ensure understanding on the part of *another*—our first-person specifications also do not "hang in the air." Rather, as we can now see, they add further and more elaborately articulated elements to the "regular connections" which enable us to understand one another, and so help to ensure that we do so correctly.

In addition Wittgenstein notes that by "nature" and "training" we are disposed to authoritative verbal expressions of other states of mind as well. He specifically mentions wishes, longings, and expectations (see, e.g., §§444–445 and the "if I have learned to talk, then I do know" of §441); but what he says also applies to desires, beliefs, hopes, fears, and other states with sentential content. In these cases B's expressive self-ascription should also describe further aspects of B's actions or environment, so that A could use this link in a way analogous to that above. Thus if A takes B's use of "S" as expressing an expectation, A will also hold that B will regard whether a situation is described by "S" as determining whether, in that situation, B should take his expectation as fulfilled. This is a less direct prediction of the same general kind as we see in the case of the monologue people, but one which makes clear how the use of sentences links mental states with objects and events which are remote in space and time. So as Wittgenstein says at §445, expectation and fulfilment "make contact" in language, and in another of the "regular connections" by which interpretive understanding is sustained.

This is how Wittgenstein answers the questions he raises. Our ability to establish and follow rules rests on a system of empirical regularities, which hold over linguistic behaviour, nonlinguistic behaviour, and the environment. These are common to human beings, and with training become part of "the framework on which the working of our language is based" (§240). Our coordination in them, therefore, is "not agreement in opinions but in form of life" (§241). We can make these regularities explicit by considering, as in the case of the monologue people, how a radical interpreter might characterize them, namely by taking sentences an interpretee utters as having regularly repeatable uses (as in the familiar "that P" constructions) in characterizing the interpretee's nonverbal actions, motives (or states of mind), and environment. In this the interpreter would represent the utterances as meaningful, the motives, like the sentences, as directed on the environment ("propositional"), and the actions as rational ("intelligible" and "logical") in light of them. The interpreter's success in this would be marked by the ability to predict what others will do on the basis of what they say, as we do in the spontaneous and everyday coordination of activities which is central to our social form of life. So we can see that these same regularities—which can be extended indefinitely to avoid misunderstanding in practice—enable us to regard one another as speaking a

language which we all understand, as following rules (and using concepts) in the same ways, and as expressing what we think and mean with an authority for which our mutual understanding provides the prospect of continual ratification.[15]

Although these considerations are complex, they are reflected throughout Wittgenstein's remarks, and enable us to see how densely and precisely these remarks

15. It may be worth noting that this account apparently contradicts a distinguished tradition in the interpretation of Wittgenstein. In *The Realistic Spirit* Cora Diamond describes how

> John McDowell, in speaking of the kind of philosophical illusion from which Wittgenstein in his later work tries to free us, uses the phrase "the view from sideways on" to characterize what we aim for, or think we need to aim for, in philosophy. We have, for example, the idea of ourselves as looking, from sideways on, at the human activity of following a rule, and as asking from that position whether there is or is not something *objectively* determined as what the rule requires to be done at the next application. To think of the question in that way is to try to step outside our ordinary saying what a rule requires, our ordinary criticisms of steps taken by others, our ordinary ways of judging whether someone has grasped what a rule requires. We do not want to ask and answer these ordinary questions, but to ask what in reality there is to justify the answers we give when we are unselfconsciously inside the ordinary practice. (Diamond 1991, 184–185)

> On the reading proposed here, by contrast, Wittgenstein's discussion (particularly the material in §§206–7 and §243) aims to convey just the kind of understanding the seeking of which Diamond condemns as illusory. Wittgenstein takes it that "when we are unselfconsciously in the ordinary practice" we may well *not* be aware how that practice depends on empirical regularities and can be regarded as ratified by reference to them. Wittgenstein is sometimes explicit about this, saying, for example,

> the description of the method and unit of measurement tells us something about the world in which this measurement takes place. And in this very way the technique of use of a word gives us an idea of *very* general truths about the world in which it is used, of truths in fact which are so general that they don't strike people, I'm sorry to say, and philosophers, too. ("Notes for the 'Philosophical Lecture,'" 449)

> To appreciate the role of empirical regularities in this case—which, as Wittgenstein repeatedly reminds us (PI §§142, 242) are analogous to those which sustain measurement more generally—Wittgenstein encourages us "to step outside of our ordinary saying what a rule requires," etc. For this we adopt the external perspective on our own practices which a radical interpreter would have, and for whom the regularities in which we unreflectingly participate might be a matter of implicit empirical study.

> In this we regard the interpretive understanding of language not just "from outside" but, in the case of the monologue people "from sideways on." Taking things this way, moreover, enables us to understand the sense in which in the ordinary case, despite the ways in which "whatever I do can be brought into accord with a rule," there is nonetheless "something *objectively* determined as to what the rule requires to be done at the next application"; and again how in a "private" case there could be no such thing. Again this enables us to see "what in reality there is to justify the answers we give," e.g., when we speak with first-person authority as regards how we follow a rule, or what we intend and mean more generally. This is particularly important, for in considering how our exercise of first-person authority can be justified we naturally look to our own case, and are thus liable to fail to see "what in reality there is"—namely the possibility of our own regularities being understood *by others*—which justifies us.

> McDowell 1984 takes Wittgenstein's central point to be "to attack the assimilation of understanding to interpretation" (357); but in this McDowell seems mainly to have the

are interconnected. Thus Augustine's discussion of language in §1 serves not only to introduce errors from Wittgenstein's early philosophy, but also, via the transition at §32, to bring forward a conception of the relation of verbal and nonverbal behaviour—both repeatedly illustrated in his simple language-games and taken as discernable in a fully fledged "articulate language" by a radical interpreter—upon which his later account will turn. In this Augustine's "natural language of mankind" in §1 becomes "the common behaviour of mankind" which renders interpretation possible in §§206–207, and so elucidates how meaning is founded in our "form of life" in §241. So also we understand Wittgenstein's later philosophy by imposing on its first paragraph from Augustine the same shift—roughly, from a first-person or Cartesian view of psychological and semantic concepts to one which sees them as rooted in interpretable practice—as Wittgenstein seeks to effect in the course of the *Investigations* itself. In this we deepen and justify his claim at the end of §1 that explanations of meaning come to an end—not, as philosophers have long held, in mental reference which ultimately penetrates to abstract objects, but—in how we, as members of a social species, spontaneously *act*. Thus consider the following:

> The absent-minded man who at the order "Right turn!" turns left, and then, clutching his forehead, says "Oh! right turn" and does a right turn.—What has struck him? An interpretation? (PI §506)

Brief as this is, it is connected with almost everything we have discussed so far. The clutching of the forehead is comparable to the behaviour Wittgenstein mentions in §54 as characteristic of correcting a slip, which one could recognize "even without knowing his language." It occurs as part of a simple "regular connection" between utterance and action, such as §§206–207 enables us to see as essential to the understanding of language (and hence of sententially expressible thought), and which in §243 we find in "people who spoke only in monologue," as the soldier speaks here. This is a connection which *another* might interpret to find "the rule that he is following" by framing "the hypothesis which satisfactorily describes his use of words, which we observe." Such interpretation would yield that "Right turn" can be taken as linked in this man's behaviour with turning right, as by "a natural law governing the play," but making allowance for error (or other forms of noncompliance or nonfulfilment) on his part. This would make it possible to predict how he will act *in accord with* these words, as uttered by himself or another, as we see in this case. So

first-person conception in mind. Although Wittgenstein both raises the question and explicitly addresses it in §§206–207, McDowell does not consider how correctness in following a rule in practice is supposed to be established, nor why it is lacking in a private language. In particular he makes no mention of Wittgenstein's linking this topic to interpretation involving "the common behaviour of mankind" which makes it possible "to interpret an unknown language" via "regular connection" between utterance and action. Wittgenstein's own explicit use of the notion of interpretation here is an "assimilation of understanding to interpretation"—but in the third- as opposed to the first-person case—which McDowell does not to take into account.

the hypothesis could be strongly confirmed, and (with the soldier's help if necessary) put beyond any doubt which might actually arise. In acting, however, the soldier is not *himself* interpreting the order. Rather, as in answer to Wittgenstein's questions of §198, he is exhibiting a grasp of it (as imparted by training) "which is *not* an *interpretation*, but which is exhibited in what we call 'obeying the rule' and 'going against it' in actual cases" as in §201. (And his grasp, as in that remark, is exhibited *both* in his obedience and in his failure to obey.)[16]

## 8. SUBJECTIVITY, PRIVACY, AND APPEARANCE

This makes clear how in Wittgenstein's account the practice of language renders human thought both social and objective. In learning language we come to participate in such a system of regularities, and so to speak and think in coordination with one another, and in ways which are subject to mutual interpretive assessment. So our participation in these regularities renders meaning and understanding objective, and thereby also demarcates them from their appearances. As Wittgenstein says:

> there are certain criteria in a man's behaviour for the fact that he does not understand a word: that it means nothing to him, that he can do nothing with it. And criteria for his 'thinking he understands', attaching some meaning to the word, but not the right one. And, lastly, criteria for his understanding the word right. In the second case one might speak of a subjective understanding. And sounds which no one else understands but which I '*appear to understand*' might be called a "private language". (PI §269)

This remark—like the "it is not possible to obey a rule 'privately'" of §202—serves to connect the topic of rules with that of private language. In the terminology Wittgenstein introduces here, the deviant pupil of §185 has a *subjective* understanding of "+2," that is, something rightly regarded as *an* understanding, but one which he does not share with others. He attaches "some meaning" to

---

16. This account also enables us to approach a problem which Wittgenstein does not specifically consider, but which arises on any empirical approach to meaning. Our interpretive understanding of language and of others more generally is liable to empirical indeterminacy, as illustrated by Wittgenstein's use of algebraic formulae, and later as particularly stressed by Quine. The embedding of an interpretee's sentences in the ascription of all propositional attitudes, as emphasized on Wittgenstein's account, both spreads this indeterminacy across all sententially described states and events, and by this same means minimizes its practical importance, by making the test of each ascription at the same time a test of the interpreter's and interpretee's understanding of the sentence as well as the state of mind involved. Since our understanding of language is as it were repeated and tested for us in our understanding of each and every sententially described state of mind, we can regard it as both at the core of our understanding of ourselves and others, and also, by its role as the armature of interpretation, as being particularly strongly confirmed.

the phrase—his practice shows a genuine regularity—but "not the right one." For the right regularity for a student of *our* language is not just any regularity, but that particular regularity which coordinates the student's practice fully with our own. This would render the case one of objective (as opposed to subjective) obedience to the teacher's order, which is what both teacher and pupil intend.

Such a real but subjective understanding also contrasts with the mere *appearance* of understanding, which is all there could be if *no* hypothesis could satisfactorily describe a putative rule-follower's use of words.[17] This, by definition, is the case for a private language. So for a language to be *private* in Wittgenstein's sense is just for it to be *uninterpretable* as explicated by §§206–207 and §243. Another person "cannot understand this language" since its words "are to refer to what can only be known to the person speaking; to his immediate private sensations." In this case neither "the common behaviour of mankind" of §§206–207, nor anything else, can provide a "system of reference" for these words by which anyone besides their user might understand them.

Such a language could perforce be used only in monologue; and this is why Wittgenstein introduces it by explicit contrast to interpretable monologue in §243. Here, as we have seen, there are regularities which hold between first-person expressive uses of the words in question and other behaviour, and these provide a mode of triangulation by which an interpreter can link them with the speaker's intentions and so with the environment. In the private language, by contrast, there are to be no such links (§256), nor anything else which would render the language subject to interpretive understanding. But since, as he remarks, "to *think* one is obeying a rule is not to obey a rule"; and since this distinction turns on the possibility of the interpretive understanding of one putative rule-follower by another; there can be no such thing as following a rule "privately." Thus the conclusion as drawn in §258—that in such a case there is no "criterion of correctness" for the use of a word, so that "we cannot talk about 'right'"—already follows from the constitutive role of understanding by another as previously explicated.

## 9. INNERNESS, PRIVACY, AND OUR PICTURE OF THE MIND

Still Wittgenstein's redrawing of this conclusion has further and independent interest, both as regards his own philosophical steps and the methods by which he encourages his reader to retrace them. Wittgenstein took philosophical problems to be rooted in misunderstandings of language fostered by "pictures" of the use of words, and this was particularly so as regards the problems of mind. As he says

---

17. Here it is perhaps worth noting that when Wittgenstein says "There is not enough regularity for us to call it 'language'" in §§206–207 he is not excluding the idea that there is an order which we fail to find. We can acknowledge that order renders language interpretable without holding that we can always find such order when it is there.

425. In numberless cases we exert ourselves to find a picture and once it is found the application as it were comes about of itself. In this case we already have a picture which forces itself upon us at every turn,—but does not help us out of the difficulty, which only begins here.

427. "While I was speaking to him I did not know what was going on in his head." In saying this, one is not thinking of brain-processes, but of thought-processes. The picture should be taken seriously. We should really like to see into his head. And yet we only mean what elsewhere we should mean by saying: we should like to know what he is thinking. I want to say: we have this vivid picture—and that use, apparently contradicting the picture, which expresses the psychical.

The picture to which Wittgenstein here refers is that of the *innerness* we ascribe to mental items, when we regard them as *in* the mind, or again when we regard the mind as *in* the body. From early in his work Wittgenstein noted that our notion of mental innerness seemed to separate the contents of the mind from the public world in which we live and act. As he says in *The Big Typescript* "The expression 'that something is going on in our mind' is supposed to suggest, I believe, that it can't be situated in physical space" (174e). Accordingly we picture the mind as a kind of *nonphysical inner space*, whose contents are psychological and phenomenological as opposed to physical or neural ("not thinking of brain-processes but of thought-processes"), and detectable by a process analogous to sight (introspection: on this see also his emphasis in PI §305: "When one says 'Still, an inner process does take place here'—one wants to go on: 'After all you *see* it.' "). But then since we picture the mind as *in* the body, we also relate this nonphysical space to the space internal to the body, as in somehow wanting "to see into his head." Wittgenstein applies this notion in detail to the idea that we know what pain is only from our own case.

295. "I know.... only from my *own* case".... even if [this] gives no information, still it is a picture....

When we look into ourselves as we do philosophy, we often get to see just such a picture. A full-blown pictorial representation of our grammar. Not facts; but as it were illustrated turns of speech.

296. "Yes, but there is *something* there all the same accompanying my cry of pain. And it is on account of that that I utter it. And this something is what is important—and frightful."—Only whom are we informing of this? And on what occasion?

297. Of course, if water boils in a pot, steam comes out of the pot and also pictured steam comes out of the pictured pot. But what if one insisted on saying that there must also be something boiling in the picture of the pot?

In §297 the picture of the mind as an inner space in which we encounter the sensation of pain is compared to a picture of a pot in which water is boiling, and our idea that pain goes on in this *pictured* space to the idea that there is something boiling—not just in the pot, in which the boiling actually takes place, but—in the

*picture* of the pot. On this picture, as Frege says, "nobody even knows how far his image (say) of red agrees with somebody else's." But then, as Wittgenstein observes, the privacy which this picture represents sensation as having provides a *reductio* of the use of the picture itself.

> 293. If I say of myself that it is only from my own case that I know what the word "pain" means—must I not say the same of other people too? And how can I generalize the *one* case so irresponsibly?
>
> Now someone tells me that *he* knows what pain is only from his own case!—Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle.—Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing.—But suppose the word "beetle" had a use in these people's language?—If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*: For the box might even be empty.—No, one can 'divide through' by the thing in the box; it cancels out, whatever it is.
>
> That is to say: if we construe the grammar of the expression of sensation on the model of 'object and designation' the object drops out of consideration as irrelevant.

Here the inner space is pictured as a box rather than a pot, but the basic metaphor[18] is the same. In applying it we take ourselves to recognize the contents of the inner space independently of anything else, and so hold that "it would be quite possible for everyone to have something different in his box" and again that "the box might even be empty." If we take such things to be possible, we require grounds for judging which of these possibilities actually obtain. Since we can have no access to the containers of others, we can have no such grounds, and the contents of the containers are private. But then items which were private in this way could not be the referents of words in a public language. For such words, as noted in §229, are used in accord with public criteria, which serve to determine whether or not they are used correctly. So if a word is used in a public language—as "pain" or other words for sensation actually are—it cannot have a referent whose nature of existence is independent of these criteria and so private. Items which can vary or fail to exist while the conditions of truth or correct use for a sentence remain constant cannot be construed as referents of terms in that sentence. Hence, as Wittgenstein says in his "Notes for the 'Philosophical Lecture'": "In fact the private object is one about which neither he

18. The "pictures" in which Wittgenstein takes philosophical problems to be rooted all seem instances of conceptual metaphor, as described by Lakoff and others. So (and remarkably) something close to Wittgenstein's conception of philosophy also informs Lakoff and Johnson's *Philosophy in the Flesh*. Although the programme of this book seems to have been conceived in ignorance of Wittgenstein, it nonetheless casts light on some of his most basic claims. As noted below, I discuss the connection between Wittgenstein and conceptual metaphor in Hopkins 2000a.

who has it nor he who hasn't got it can say anything to others or to himself."[19] Accordingly Wittgenstein summarizes his argument at §304 as follows:

> "But you will surely admit that there is a difference between pain-behaviour accompanied by pain and pain-behaviour without any pain?"—Admit it? What greater difference could there be?—"And yet you again and again reach the conclusion that the sensation itself is a *nothing*."—Not at all. It is not a *something*, but not a *nothing* either! The conclusion was only that a nothing would serve just as well as a something about which nothing could be said. We have only rejected the grammar which tries 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 anything else you please.

These remarks also have the precise links with others noted above. The interlocutor's "again and again" relate respectively to the first- and third-person perspectives. The first refers back to the discussion (§§258–259) establishing that the user of a word for a private object cannot speak about it in monologue (in a private language), to which the interlocutor responded in §260 by saying "Then did the man...make a note of *nothing whatever*?" The second refers to §293 quoted above, in which "the box might even be empty," so that, as far as interpretation by another is concerned, "pain" cannot be construed as referring to such an item in any case. (The interlocutor's response in §296 was that "there is *something* there all the same.") Wittgenstein's allusions thus make clear that his argument aims to show that a private object—as opposed, from the outset (§246), to the actual sensation of pain—would be "a something about which nothing can be said." In this he rejects construing "the grammar of the expression of sensation on the model of 'object and designation,'" by rejecting "the full-blown pictorial representation of our grammar" which "forces itself on us at every turn" and so "tries to force itself on us here."

## 10. SENSATION, THOUGHT AND PHYSICALISM

So far, as we may put the point, Wittgenstein's argument has not been so much about sensations themselves as about a *picture* we use in thinking of the mental as inner. He describes his procedure by saying in PI §374 that "the best that I can propose is that we should yield to the temptation to use this picture, but then investigate how the *application* of the picture goes." His investigation shows how the use of the picture is self-refuting. The picture is supposed to depict how we think about

19. "Notes for the 'Philosophical Lecture,'" 451. See also the "What I deny is that we can construe the grammar of 'having pain' by hypostatising a private object" at p. 451 or the dialogue at pp. 452–453: "'But if he is truthful, why shouldn't we take his word for it...?' But we do!...'Then where do you disagree with us?'—When you talk about something incommunicable, private."

sensations and use language to describe them. In trying to use it, however, we find that we are representing sensations as things about which we cannot speak or think, either to ourselves or others. Since the application of this picture would have the consequence that we could not use words as we do, and as the picture is supposed to depict us as doing, we must reject the picture. So as Wittgenstein says at §305, his argument registers "our setting our faces against the picture of the 'inner process'." For his argument shows in detail (in both the first- and third-person cases) how "this picture with its ramifications stands in the way of our seeing the use of the word as it is."

In §293 and §304 Wittgenstein also goes beyond this. He specifies more precisely the way grammar and picture are linked, and also formulates his disagreement with Frege, who saw no difficulty in the idea that we frame publicly understandable thoughts about sensations which we nonetheless regard as private. As §293 makes clear, in the picture in question we represent the sensation *as an object*, and so "construe the grammar of the expression of sensation on the model of 'object and designation'." This in turn leads us to regard the sensation as private and hence, ultimately, as "a something about which nothing can be said." So as Wittgenstein remarks at §304, a main point of his argument is also that a sensation is *not* to be construed as an object. In "the grammar of the expression of sensation" the sensation itself is "not a *something*, but not a *nothing* either."[20] (If we construe the sensation itself as "a something," this will prove a something about which nothing can be said; so this is the point at which the grammar and picture force themselves on us, and also the point at which they are to be resisted.)

Wittgenstein rightly expects that his reader will find this paradoxical, and stresses that "the paradox disappears only if we make a radical break with the idea that language always functions in one way," namely, to convey thoughts, which may be about objects (somethings) as various as houses and pains. So in formulating this paradox Wittgenstein explicitly forecloses a particular response to his argument so far. This is for the reader *not* actually to reconsider the grammar which "forces itself on us here," but rather simply to follow Wittgenstein in holding that sensations are not private but public. According to Wittgenstein's argument to do this is simply to remain confused: to think of sensations in a way which renders them private, but to insist that they are public all the same. And this certainly casts no light on the sense in which, according to Wittgenstein, pain is "not a *something*" for us, nor on what it might mean for us to break with the apparently commonsense idea that we use

20. Commentators have often hesitated to take this paradoxical formulation as part of Wittgenstein's positive account of the "grammar of the expression of sensation," taking it that he is speaking here of the "grammatical fiction" (PI §307) of the "private object." But Wittgenstein makes his intention clear by demarcating the actual sensation from the grammatical fiction by the use of contradiction. The grammatical fiction is that the sensation itself is "*a something*" about which we cannot speak; the grammatical reality is that the sensation itself is "*not* a something, but not a *nothing* either," which remains a paradox on which little light has been cast.

"pain" to formulate thoughts about pains, and in the same way as we do about other items. These ideas remain genuinely challenging, and to understand them further it will be useful to link Wittgenstein's discussion more explicitly to physicalism.

As we have noted in considering Wittgenstein's tacit specifications ("looking at his beetle" in §293), to construe a sensation *as a something* is to take it as comparable to something one might see: an everyday material object or process, a patch of colour, and so forth. We can, however, provide a different but equally clear and relevant example of a something—although not that which Wittgenstein targets—by using his own analogy in §297. Here as we saw, Wittgenstein compares the inner processes he is discussing to those which go on inside a boiling pot, and the picture we frame of these processes to a picture of the pot. He thus compares the physical causes of pain to those of boiling in a pot and the expression of this activation in behaviour to that of steam from the pot.[21]

This analogy straightforwardly represents pain as a physical process which occurs within our bodies (to carry it through would be to regard pain as a kind of neural boiling).[22] So the most natural reading of the remark (allowing for

21. And of course he presents us with a series of physical causes which lead to a particular mental event, and a series of physical effects which that event causes. As is now familiar this provides premises for arguing that any such mental event should be identified with the physical event which has exactly the same physical effects, and hence exactly the same causes and effects overall. An early version of this now standard argument is in McGinn and Hopkins 1978, and is related to the claims about supervenience argued in Hopkins 1975. Although this paper was published before zombies came into fashion, the treatment of the topic there and in relation to Wittgenstein still seems to me to be correct.

22. Although Wittgenstein does not espouse physicalism, it is clear from his discussion that he took the picture he targets to isolate mental processes from neural processes with which we might link them conceptually. Thus in the *Blue Book* he considers the idea that "certain physiological processes correspond to our thoughts in such a way that if we know the correspondence we can, by observing these processes, find the thoughts" (7). Here Wittgenstein imagines this as "verified experimentally" by the use of a device (as it were an imaginary version of a brain-scanner) which makes it possible for someone to watch his own brain while he is thinking. In such a case, he says, we could regard *both* the "train of images and organic sensations" available to introspection *and* the observed working of the brain as "expressions of the thought." This, he says, illustrates the sense of the phrase "thought takes place in our heads" as "soberly understood." And he insists straightforwardly that "This phrase has sense if we give it sense," and that we could give it sense by linking it with observation and experiment in this way.

[In the thought experiment] we have given this expression its meaning by describing the experience which would justify the *hypothesis* that the thought takes place in our heads, by describing the experience which we would wish to call "observing thought in our brain". (BB, 8)

Again, he says that "we could easily imagine"

that we looked at a group of things in this room, and, while we looked, a probe was stuck into our brain and it was found that if the point of the probe reached a particular point in our brain, then a particular small part of our visual field was thereby obliterated. In this way we might co-ordinate points of our brain to points of the visual image, and *this might*

reservations expressed elsewhere)[23] would be to take it as indicating that in using the picture under investigation we impose upon a physical reality a dualistic conception of the mind. (We take that as well as the physical boiling in the pot, there is *also* a boiling in the picture of the pot, which must of course—since it goes on in no material pot—be a process of some aethereal nonphysical kind.) In this we would take pain as occurring not in the nervous system, but in the inner space by which we represent to ourselves what happens in the nervous system. We insist that the boiling with which we are concerned is not that going on in the pot, but rather another going on in our picture of the pot; and in this we lose track of the very inner processes the picture was supposed to represent.

Whether or not this is the correct reading of Wittgenstein, it enables us to see that a standard physicalistic conception provides a particularly clear example of the idea that a sensation is an inner *something*—an inner physical event or process—to which we internally apply our concept of pain. Likewise this conception provides a clear account of the way in which sensations such as pain are *private*. When we think of pain

make us say that the visual field was seated in such and such a place in our brain. (BB, 9; emphasis supplied)

In these passages Wittgenstein indicates his openness to the kind of changes in philosophical thinking which have taken place with the acceptance of physicalism. The thought experiments he describes are ones in which the kind of interpretive triangulation described in PI §§206–207 and §243 is extended to include "regular connections" with neural structures and events as well. In terms of his discussion, such an extension might justify the adoption of new and explicitly physicalistic criteria for the location and typing of mental states. And of course in considering private language we must regard such criteria as tacitly excluded by the idea that the diarist of §256 has "only the sensation," about which, therefore, no one else can know. Wittgenstein did not of course anticipate the Kripkean conception that the connection between sensation and neural event might be one of identity which was metaphysically necessary, as opposed to that of sensation to its characteristic causes and effects, which was a priori but ultimately contingent. But the acceptance of this formulation still requires to be properly integrated with Wittgenstein's own conclusions concerning necessity and rules.

It is now familiar that the first of Wittgenstein's formulations above conflates the roles of types and tokens. As Davidson was later to argue, there might be a particular (token) physiological process which not only corresponded to, but was identical with, each particular (token) process of thought; while at the same time there were no physiological types (no types described via the concepts of physiology) whose instantiations strictly corresponded with those of types of thought as sententially described. Accordingly when Wittgenstein returned to the same topic in *Zettel* he rightly stressed that the absence of a correlation of types might make it "impossible to read off thought-processes from brain-processes" (Z §608) On this see also note 23 below.

23. E.g., in *Zettel* §§608ff. Still the relevance of this is not clear, since the anomalism of the mental, with which Wittgenstein in concerned in *Zettel*, need not hold for pain. As I argue in Hopkins 1999b, the discussion in *Zettel* seems best understood as a response to the apparent contradiction between the three principles (anomalism, causation between mental and physical events, and the nomological character of causality) which Davidson discusses and reconciles in Davidson 1980. There can be no doubt that Wittgenstein would have reconsidered the *Zettel* discussion had he realized that it was type/token ambiguous; and since the apparent contradiction depends on this ambiguity, we may assume that his reconsideration would have been radical.

in our own case both our activity of thinking and the event we are thinking about are linked in the same nervous system, and so form parts of a single complex neural event. This shows how each of us really does have a unique and essentially unshareable conceptual perspective on his or her own sensations, and also why, as well as private, such events also seem to us to be *subjective*. For given the way pain and our act of conceiving pain are actually fused in our feeling pain *as pain*, the pain itself *rightly* appears to us as practically indistinguishable from our apprehension of it as such. Again, however, these are not the notions against which Wittgenstein argues. For on this account it is not only possible for us to know what others are feeling, but to learn much more about this. We can compare both the neural activities which constitute our sensations and those which constitute our thoughts, and triangulate further on these via speech and other behaviour, in the ways Wittgenstein describes in §207 and §242; and this, as is familiar, might enable us to calibrate the likeness (or unlikeness) of our sensations far more confidently and precisely than we do at present.

Taking pain as an inner something in this physical way also makes clear why the imposition of the picture Wittgenstein targets has the consequences he explores. For in using the picture of the inner realm, we disconnect our representation of the application of the concept of pain from the actual object to which this concept applies. (In imposing *this* picture of pain as a something, we eliminate from the picture the something which pain actually is. Or again, in applying the model of "object and designation" in *this* way, we eliminate the object which we actually designate.) And of course once we eliminate the real object of our concept from our representation of the use of that concept, we are left representing the concept as one which no one could apply correctly, either from within or without; and so as a concept of a something about which nothing can be said. ("The decisive movement in the conjuring trick has been made" (§308) so that the analogy "falls to pieces" in our attempt to take it seriously.)

But of course it is very hard for us to recognize that this is so. For the picture so subtly masks the real object it represents that in using it we have no reason to suppose that we are in danger of thinking that real object away. The picture represents the neurologically explicable subjectivity and privacy of sensation as a kind of inner detection of objects whose whole and ostensibly nonneural nature is shown us in a realm to which we alone have access. Without reference to the neural object, as Wittgenstein shows, this subjectivity and privacy would be unintelligible. We are left with an item we take ourselves to know internally, but which has no trace in the world outside the mind, and so cannot be made the topic of objective speech or thought. But since this thinking away does not alter the *picture*, nothing forces us to know what is wrong apart from repeated demonstration of our own failures of coherence in using it. (Our representation of the lost object is so satisfying that we take ourselves to have no need of the object itself.) So here, as Wittgenstein says at §255, the treatment of a philosophical problem is like Freud's treatment of neurosis.

In both cases the problems are rooted in representations which serve to mask the reality they represent, and which so completely substitute themselves for this

reality that we are unable to question them. What we require to release us from their grip is repeated demonstration of the irrationality to which they lead us, and by someone whose authority in this we are willing to tolerate. (And the creation of such a transference-like attitude may be one purpose of the evoking and dispelling of confusion and accusation so characteristic of the *Investigations*.) In addition this representation of the inner realm may play an important role in our cognitive functioning. For the neurological phenomena which the picture represents do actually have a double innerness (they are *internal* to the nervous system, which is *internal* to the body) which the picture succeeds in depicting (it shows them as *in* the mind, which itself is *in* the body).

We depict such innerness in a whole family of examples comparable to Wittgenstein's box and pot: they constitute the conceptual metaphor[24] of the *mind as a container*. To take a related instance, we think of anger as a *hot* fluid, somehow contained within the body. The feelings of someone who is angry may *seethe* or *simmer* and so be *agitated*. A person who is *hot under the collar* in this way may be *fuming* as the anger *rises*, or *wells up* within; and so he may have to *simmer down*, or *cool down*, so as not to *boil over*. If he doesn't manage to *let off steam*, he may *burst with anger*, or *explode with rage*. We thus represent the spectrum of feeling between calmness and uncontrollable anger in terms of the temperature of the emotion-liquid, which may be cool (no anger), agitated or hot (some degree of anger), or boiling (great anger); and the pressure caused by the emotion-heat may ultimately cause the mind/body container to burst. And in using fluids to represent inner happenings in this way we also dephysicalize them: for even if someone's anger really *wells up*, *boils over*, or *spills out*, we do not expect to find it on the carpet. This is comparable to the dephysicalization of boiling in Wittgenstein's example, if we take it as going on in the picture of the pot.

We use this metaphor both consciously and unconsciously,[25] and we can see the use that Wittgenstein targets as providing us with a way of thinking of the neural events which realize sensation. In this we represent events which are (1) by other standards perceptually and causally inscrutable and (2) hidden inside our bodies by mapping them to others which are (1) perceptually intelligible but which may be (2) hidden in containers in the external environment. Such an image modeled on the external environment portrays the events involved in sensation as both internal and intelligible. Indeed, as one might say, it seems that it is partly by imposing this picture that we are *able* to represent what occurs in our brains as

24. For this topic and its relation to philosophy see Lakoff and Johnson 1999. Although the programme of this book seems to have been conceived in ignorance of Wittgenstein, it nonetheless casts light on some of his most basic claims; for his own examples of the pictures in which philosophical problems are rooted seem all to be among the metaphors Lakoff and Johnson discuss. As noted below, I discuss the connection between Wittgenstein and conceptual metaphor in Hopkins 2000a.

25. As I discuss in Hopkins 2000a, the unconscious use of such metaphor is obvious and pervasive in the expression of phantasies about the mind.

occurring in our minds, and hence as something we can consider and reflect upon as it occurs.[26] This may be why while "setting our faces against the picture of the 'inner process'" Wittgenstein also stresses that "*Certainly* all these things happen in you" and that he does not "dispute [the picture's] *correctness*" (§§423–424), as opposed to the use of it we make in philosophy.

Thus the situation is not, as our subjectivity might lead us to suppose, that in our first-person conception of pain we are aware of the whole nature of the object we conceive. Rather in this case our conception is *partial*, and leaves the full nature of the object mainly out of account. The events we conceive internally are hidden inside us, and irrelevant for us to see or touch. So we (or our brains) do not represent them as having visible or tangible extension, or as taking place in structures with insides or outsides or surface or volume. Rather in subjectivity we omit all this, and represent only part of the neural activity connected with the location of some injury, and that location mainly just as painful.[27] So putting Wittgenstein's conclusions in these physicalistic terms, we can say that the error which leads to dualism is that of mistaking a *partial* conception of something which is complex, spatially extended, and public for a *full* conception of something which is simple, nonextended, and private.[28]

Taking matters this way enables us to return to Wittgenstein's dialectic. It is clear that the point of the interlocutor's use of "the sensation itself" in §304 is to invoke this first-person—and as we now see very partial—conception. So, as we may put the point, in thinking of "the sensation itself" as in §304 we are *not* thinking of the sensation *as a whole*, or *as the* (neural, physical, extended) *thing it actually is.* From the perspective of physicalism it is no paradox to observe that in this thinking of

26. Our main other method of effecting this seems essentially the same. In learning language we perforce learn to map the mechanisms which realize our desires, beliefs, and other motives to the sentences which describe their conditions of fulfilment, satisfaction, and the like, and thereby render them thinkable by describing them as related to the situations to which their biological function relates them causally. We seem generally to conceive the brain as the mind by coming to map aspects of neural function to aspects of the environment which we are already able to represent, and thus to render these inner aspects thinkable by a process akin to metaphor.

27. This may explain why philosophers such as Nagel and McGinn so emphasize the problem of relating consciousness to space. Nagel 1998 urges that mental concepts "don't give us the comfortable initial handle on the occupants of the familiar spatio-temporal world that pre-scientific physical substance concepts do" (339), while McGinn 1995 stresses the necessity for "'containing' the non-spatial (as we now conceive it) phenomenon of consciousness" (159). The idea that phenomena such as pain are not spatial, however, would involve a non sequitur: a move from the fact that pains (or conscious experiences more generally) are not conceived introspectively as having spatial features linked with touch and sight, to the idea that they *lack* such features, and are therefore not extended, and so nonphysical. Since conscious events are neural events, consciousness is certainly a spatial phenomenon; and this is consistent with the fact that in introspection we do not represent things as having a full complement of spatial properties

28. As noted, the ideas in this section are discussed more fully in Hopkins 2000a and Hopkins 2000b. They are applied in detail to the problem of consciousness in Hopkins 2007.

"the sensation itself" we are precisely *not* thinking of the sensation *itself*, but rather of a single *aspect* of the sensation, and that as disconnected from the rest. This is a truth (and a break with Frege) which Wittgenstein's "not a *something*" captures precisely, and which differentiates the inner from other objects of thought. If we think of a house by thinking of its nearest or most striking aspect, we take the aspect as *of the house* whose aspect it is. By contrast when we think of "the pain itself" we do not think of this as an aspect of something further (or physical), whose nature is also conveyed in its external natural expression. Rather we take it as the thing itself, whose nature is fully displayed within. This is "the decisive step in the conjuring trick," and the one Wittgenstein's "not a *something*" is intended to correct.

Likewise reference to the neural mechanisms of sensation clarifies the way their internal qualities are (not incommunicable but) actually communicated. In considering this in §310 Wittgenstein refers to natural expressions and their preconceptual empathic responses: "Imagine not merely the words 'I am in pain' but also the answer 'It's not so bad' replaced by instinctive noises and gestures." We see the relevance of this more fully when we consider how the expression of (the neural event of) pain in one person causes a corresponding event in the brain of another: an empathic neural shadow, as it were, but one which partly matches the original.[29] So again in thinking in this physicalistic way we part from the idea that language "always serves the same purpose: to convey thoughts—which may be about houses, pains, good and evil, or anything else you please" (§304). In these and other ways we can use the physicalism for which Wittgenstein cleared the way not only to understand his remarks on the mental and its innerness, but also to improve our physicalistic account of these topics themselves.

## REFERENCES

Almog, Joseph, John Perry, and Howard K. Wettstein. *Themes from Kaplan*. New York: Oxford University Press, 1989.

Davidson, Donald. "Mental Events." In *Essays on Actions and Events*. Oxford: Clarendon Press, 1980.

Diamond, Cora. *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*. Cambridge, Mass: MIT Press, 1991.

Fodor, Jerry. "Propositional Attitudes." In *Readings in the Philosophy of Psychology*, vol. 2, ed. Ned Block. London: Methuen, 1981.

Fodor, Jerry. "Psychosemantics, or: Where Do Truth Conditions Come From?" In *Mind and Cognition*, ed. William Lycan. Cambridge, Mass.: Blackwell, 1990.

Frege, Gottlob. *Translations from the Philosophical Writings of Gottlob Frege*. Ed. Peter Geach and Max Black. Oxford: Blackwell, 1952.

Hopkins, Jim. "Wittgenstein and Physicalism." *Proceedings of the Aristotelian Society* 75 (1975): 121–146.

Hopkins, Jim. "Patterns of Interpretation: Speech, Action, and Dream." In *Cultural Documents: The Interpretation of Dreams*, ed. L. Marcus. Manchester: Manchester University Press, 1999a.

29. On this see Singer 2004.

Hopkins, Jim. "Wittgenstein, Davidson, and Radical Interpretation." In *The Philosophy of Donald Davidson*, ed. Lewis Edwin Hahn. Chicago: Open Court, 1999b.

Hopkins, Jim. "Psychoanalysis, Metaphor, and the Concept of Mind." In *The Analytic Freud*, ed. Michael Levine. London: Routledge, 2000a.

Hopkins, Jim. "Evolution, Consciousness, and the Internality of Mind." In *Evolution and the Human Mind: Modularity, Language, and Meta-Cognition*, ed. Peter Carruthers and Andrew Chamberlain. Cambridge: Cambridge University Press, 2000b.

Hopkins, Jim. "The Problem of Consciousness and the Innerness of the Mind." In *Perspectives on Perception*, ed. Mary Margaret McCabe and Mark Textor. Frankfurt am Main: Ontos Verlag, 2007.

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

Lakoff, George, and Mark Johnson. *Philosophy in the Flesh*. New York: Basic Books, 1999.

McDowell, John. "Wittgenstein on Following a Rule." *Synthese* 58 (1984): 325–363.

McGinn, Colin. "Consciousness and Space." In *Conscious Experience*, ed. Thomas Metzinger. Paderborn: Schöningh / Imprint Academic, 1995.

McGinn, Colin, and Jim Hopkins. "Mental States, Natural Kinds, and Psycho-Physical Laws." *Proceedings of the Aristotelian Society, Supplementary Volumes* 52 (1978): 195–236.

Nagel, T. "Conceiving the Impossible and the Mind-Body Problem." *Philosophy* 73 (285) (1998): 337–352.

Singer, T. "Empathy for Pain Involves the Affective but Not the Sensory Components of Pain." *Science* 303 (2004): 1157–1162.

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus* [1922]. Trans. D. F. Pears and B. F. McGuiness. London: Routledge, 1961.

Wittgenstein, Ludwig. *Philosophical Investigations* [1953]. Trans. G. E. M. Anscombe. Oxford: Blackwell, 1958.

Wittgenstein, Ludwig. *Remarks on the Foundations of Mathematics* [1956]. Ed. G. H. Wright, R. Rhees, and G. E. M. Anscombe. Trans. G. E. M. Anscombe. Oxford: Blackwell, 1978.

Wittgenstein, Ludwig. *Preliminary Studies for the "Philosophical Investigations": Generally Known as the Blue and Brown Books* [1958]. Oxford: Blackwell, 1964.

Wittgenstein, Ludwig. *Notebooks 1914–1916* [1961]. Ed. G. H. Wright and G. E. M. Anscombe. Trans. G. E. M. Anscombe. Oxford: Blackwell, 1979.

Wittgenstein, Ludwig. *Zettel*. Ed. G. E. M Anscombe and G. H. von Wright. Trans. G. E. M. Anscombe. Oxford: Blackwell, 1967.

Wittgenstein, Ludwig. *Philosophical Grammar*. Ed. Rush Rhees. Trans. Anthony Kenny. Oxford: Blackwell, 1974.

Wittgenstein, Ludwig. *Philosophical Remarks*. Ed. Rush Rhees. Trans. Raymond Hargreaves and Roger White. Oxford: Blackwell, 1975.

Wittgenstein, Ludwig. *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge, 1939: From the Notes of R. G. Bosanquet, Norman Malcolm, Rush Rhees and Yorick Smythies*. Ed. Cora Diamond. Hassocks: Harvester Press, 1976.

Wittgenstein, Ludwig. "Notes for the 'Philosophical Lecture.'" In *Philosophical Occasions*, ed. James Klagge and Alfred Nordmann. Indianapolis: Hackett, 1993.

Wittgenstein, Ludwig. *The Big Typescript*. Ed. and trans. C. Grant Luckhardt and Maximilian A. E. Aue. Oxford: Blackwell, 2005.