

Wittgenstein and the Life of Signs

Jim Hopkins

In this paper I try to show that Wittgenstein's discussions of meaning have greater scope and unity than is commonly recognized. In particular, his claims about the use of signs can be seen to integrate consideration of a range of topics -- including essentials of the computational theory of mind, functionalism, psychosemantics, and radical interpretation -- whose mutual relevance is only now being understood. So despite the great attention already given to his views, his way of relating these topics in a unified account has more than historical interest.

1. The life of signs: two opposed accounts.

We can start to see this by examining some links between Wittgenstein's early (*Tractatus*) and late (*Investigations*) approaches to mental and linguistic representation, as these are indicated by a characteristic remark from the latter book.

§432. Every sign *by itself* seems dead. *What* gives it life. In use it is *alive*. Is life breathed into it there? -- Or is the *use* its life?

Speaking of the life of signs is a particular way of speaking of the nature of meaning. Signs which have life in this sense are *animated*: they are not mere marks or sounds, but represent or refer to things, and so can be true or false. In this remark Wittgenstein opposes two possible accounts of this animation. The first is that life is 'breathed into' the sign, and so comes from outside the use. The second, by contrast, is that the use somehow constitutes the life.

This opposition holds, among other things, as between Wittgenstein's accounts of meaning in the *Tractatus* and the *Investigations*. The idea that life is 'breathed into' signs is the idea that their meaning is provided by mental activities or processes of thought; and the *Tractatus* contains an account of these activities and how they support linguistic meaning. The idea that the use of signs should be seen as constituting their life, by contrast, is that of the *Investigations* itself.

In *The Blue and Brown Books* Wittgenstein describes the same opposition as follows:

It seems that there are *certain definite* mental processes bound up with the working of language, processes through which alone language can function. I mean the processes of understanding and meaning. The signs of our language seem dead without these mental processes; and it might seem that the only function of the signs is to induce such processes, and that these are the things we ought really to be interested in....But if we had to name anything which is the life of the sign, we should have to say that it was its *use*. (3 - 4)

Here Wittgenstein characteristically omits to say that the mistaken idea with which he is contrasting his own view is one which he adopted in earlier work. But he makes this clear in the *Investigations*, saying that his task is to attain

...greater clarity about the concepts of understanding, meaning, and thinking. For then it will also become clear what can lead us (and did lead me) to think that if anyone utters a sentence and *means* or *understands* it, he is operating a calculus according to definite rules. (§81)

2. Fregean sense and the life of signs.

Wittgenstein also takes it that considering what gives life to signs is useful for understanding other approaches to meaning. For example he says:

Frege ridiculed the formalist account of mathematics by saying that the formalists confused the unimportant thing, the sign, with the important thing, the meaning. Surely, one wishes to say, mathematics does not treat of dashes on a bit of paper. Frege's idea could be expressed thus: the propositions of mathematics, if they were just complexes of dashes, would be dead and utterly uninteresting, whereas they obviously have a kind of life. And the same, of course, could be said of any proposition: Without a sense, without the thought, a proposition would be an utterly dead and trivial thing. And further it seems clear that no adding of inorganic signs can make the proposition live. (4) Thus the Fregean notion of sense (or the mental act of grasping the sense) can also be taken as the source of the life of signs. When we grasp the sense of a word, or the thought expressed by a sentence, this mental act apparently determines what the word refers to, or the circumstances in which the sentence is true. Once we think of a sign in this way it is no longer a mere mark on paper, but something meaningful, as can be seen in the way we subsequently use it. To see how this idea figures in the *Tractatus*, let us first consider Wittgenstein's early conception of meaning more generally.

3. The *Tractatus* account of representation: element-combination pictures.

As is well known, in the *Tractatus* Wittgenstein regarded linguistic signs (words and sentences) as particular instances of a more general kind of representation, which he called *pictures*. These included mental representations, since in thinking, as he holds 'we make to ourselves pictures of facts' (2.1ff). As other examples he cites alphabetic notation considered as describing sounds; hieroglyphics; musical notation; maps, blueprints, and models; and also such information-bearing mechanical structures as the groove on a gramophone record, the perforated roll in a player piano, and the system of holes in the cards of a programmable loom.

What unites this variety of representations, according to Wittgenstein, is that each can be taken to consist of elements which stand for things in the domain they represent, and these elements are combined (related, ordered, configured) in the representation in a way which represents, or stands for, how things are in the domain represented. Thus in the case of language the representing elements are words and the representing combinations are sentences. The words stand for things and the way the words are combined determines how the sentence represents these things, that is, how things are if it is true.

Thus we can take the sentence 'Napoleon loves Josephine' as consisting of the elements 'Napoleon', 'loves', and 'Josephine' in that order, and 'Napoleon' as standing for Napoleon, 'Josephine' for Josephine', and 'loves' for the relation of loving. That these words are combined in this order, in turn, represents that these things have a certain order in the world, namely that in which Napoleon bears the relation of loving to Josephine. As Wittgenstein says, 'That the elements of picture are combined with one another in a certain way represents that the things are so combined [that is, in a definite way].' (2.15); or, more compactly, that the signs 'Napoleon', 'loves', and 'Josephine' stand in a certain order says *that Napoleon loves Josephine*. (cf 3.1432) We mark this in terms of the concept of truth by saying that 'Napoleon loves Josephine' is true just if Napoleon loves Josephine. Hence, as Wittgenstein says, 'The proposition *shows how things stand*, if it is true, and it *says that they so stand*.' (4.022)

4. Sentences (propositional signs) as *Tractatus* pictures.

Wittgenstein attaches particular importance both to the generality of this conception of representation and to the way it explains our capacity to understand and express novel sentences. To understand the working of sentences rightly, he stresses, we should understand them in terms of the abstract notion of an element-element: combination-combination representation which they embody, and so compare them with representations which have more concrete and accessible versions of this structure. So he introduces his

account of sentences as representations both via the general idea, and by a series of less abstract examples.

4.01 A proposition is a picture of reality.

A proposition is a model of reality as we imagine it.

4.011 At first sight a proposition -- one set out on the printed page for example -- does not seem to be a picture of the reality with which it is concerned. But no more does musical notation at first sight seem to be a picture of music, nor our phonetic notation (the alphabet) to be a picture of our speech. And yet these sign-languages prove to be pictures, even in the ordinary sense, of what they represent.

4.012 It is obvious that a proposition of the form 'aRb' strikes us as a picture. In this case the sign is obviously a likeness of what is signified.

4.013 And if we penetrate to the essence of this pictorial character we see that it is *not* impaired by *apparent irregularities* (such as the use of the flat and sharp signs in musical notation. For even these irregularities depict what they are intended to express; only they do it in a different way...

Again, when he comes to explain the connection of this with the understanding and expression of novel thoughts, he does so by means of a more concrete comparison.

4.016 In order to understand the essential nature of a proposition, we should consider hieroglyphic script, which depicts the facts that it describes.

And the alphabetic script developed out of this without losing what was essential to depiction.

Seeing the combinatorial working of these relatively clear examples of depiction enables us to grasp how we can understand novel sentences on the basis of familiarity with their components and modes of combination:

4.02 We can see this from the fact that we understand the sense of a propositional sign without its having been explained to us.

4.021 A proposition is a picture of reality: for if I understand a proposition, I know the situation it represents. And I understand the proposition without having had its sense explained to me. A proposition *shows* its sense. A proposition shows how things stand if it is true. And it *says that* they do so stand. This applies also to the idea that we can use sentences to express thoughts not expressed before, since we can combine familiar words together in novel ways:

4.027 It belongs to the essence of a proposition that it should be able to communicate a *new* sense to us.

4.03 A proposition must use old expressions to communicate a new sense. A proposition communicates a situation to us, and so it must be essentially connected with the situation.

And the connection is precisely that it is a logical picture.

A proposition says something only in so far as it is a picture.

The clearest examples of such representations are those in which spatial position is used to represent spatial position, 'As when in the law-court in Paris a motor-car accident is represented by means of dolls, etc.' (*Notebooks*, 7e) Here, as Wittgenstein says, 'the mutual spatial position of things then expresses the sense of the proposition'. For this reason 'the essential nature of the propositional sign becomes very clear when we imagine it made up of spatial objects' (3.1413). By contrast the role played by temporal order in speech, or linear order in print, is far less easy to survey.

To characterize the full working of this order we would have to specify (i) the elements of the linguistic system (words); (ii) the rules for combining these elements into representing combinations (the syntax of sentences); (iii) the relations of reference which obtain between the elements and the worldly things for

which they stand (the rules of reference which hold for words); and (iv) the way the rules of combination specified in (ii) make use of the relations of reference specified in (iii) to determine the truth-conditions of sentences via their syntax (the referentially based truth-conditional semantics of sentences). We can fairly easily imagine rules of the relevant kind for simple sentences like 'Napoleon loves Josephine', and much work on the syntax and semantics of natural language since the *Tractatus* can be seen as an attempt to understand the working of language along these lines. Such work, however, indicates that the full task -- particularly that involved in (iv) -- is forbiddingly complex. As Wittgenstein says, 'the tacit conventions on which the understanding of everyday language depends are enormously complicated' (4.002). What we learn when we learn language is by no means easy for us to survey.

5. Thoughts (mental representations) as *Tractatus* pictures.

As noted, Wittgenstein also applied this general conception of representation to the mind, by taking thoughts as representations in the minds (or brains) of persons. Thus as he wrote to Russell, in reply to questions about the *Tractatus*:

"...But a *Gedanke* [thought] is a *Tatsache* [fact]: what are its constituents and components, and what are their relation to 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. Again the kind of relation to the constituents of the thought and of the pictured fact is irrelevant. It 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. (*Notebooks* 1914-16, Blackwell 1961, 156-7)

On this view the capacity to think is explicated via the capacity to form a certain kind of mental representation. A person who entertains the thought that P forms in his mind a *Tractatus* representation that P, so that thinking involves the same kind of element-element configuration-configuration structure as we find in language. The actual elements of mental representation, and how they relate to one another and to things and situations in reality, are matters for empirical investigation: 'It would be a matter of psychology to find out.' This potential task for psychology, however, seems on the face of it even more difficult than that of spelling out the representational character of language as characterized in (i) - (iv) above. For in the case of language we start with a basic idea of what the representing elements and their combinations and relations to things are, whereas in the case of the mind (or brain) all this has yet to be discerned.

6. Thoughts and propositional attitudes more generally.

Just as person who thinks that P forms a mental representation that P, so also does a person who believes that P, expects that P, wishes that P, etc. This apparently applies to the whole range of states of mind -- the 'propositional attitudes' -- which we describe in the same *that P* way. Wittgenstein registers this assimilation in a discussion of the picture theory in *Philosophical Remarks*, saying:

I only use the terms the expectation, thought, wish, etc., that *p* will be the case, for process 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 call only 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.) (p 70)

The notion of multiplicity which Wittgenstein refers to here is that which holds between a *Tractatus* representation and what it represents; these, he says, 'must both possess the same logical (mathematical)

multiplicity.' (*Tractatus* 4.04). In holding that expectation, wishing, thinking, etc., *that P* have the multiplicity expressed by 'P' Wittgenstein is holding that they involve mental representations of the kind he describes in the letter to Russell above; that is, representations which have the same capacities to represent as the words and sentences of natural language. In the *Tractatus* itself he expresses the connection between these states, and its relation to his theory of representation, even more briefly. The relevant discussion occurs at 5.541ff:

...with certain forms of proposition in psychology, such as 'A believes that p is the case' and 'A has the thought p', etc.....considered superficially, it looks as if the proposition p stood in some kind of relation to an object A...It is clear, however, that 'A believes that p', 'A has the thought p', and 'A says p' are of the form ' "p" says p': and this does not involve a correlation of a fact with an object, but rather the correlation of facts by means of the correlation of their objects...This shows too that there is no such thing as the soul -- the subject, etc. -- as it is conceived in the superficial psychology of the present day. Indeed a composite soul would no longer be a soul.

The 'etc.' which Wittgenstein here appends to 'A believes that p' and 'A has the thought that p' indicates the application of what he says to the further states we describe in the same *that P* way. These, he says, are to be understood in terms of the form 'p' says p, which involves 'the correlation of facts by means of the correlations of their objects'. This means that these states are to be understood as involving a representation of the kind we have been considering, that is, a complex representation 'p' which says that p; and these representations themselves are facts (composed of objects), which are correlated with the facts they represent via correlations among the objects which compose them. As thinking subjects, according to the *Tractatus*, we are not to be understood as simple souls; rather we are complexes of representations, which are in turn complexes of psychological elements correlated with the world we think about.

Wittgenstein makes this psychological aspect of the picture theory clearer in his discussions of it in later works. Thus in *Philosophical Remarks* he says that 'What is essential to intention is the picture: the picture of what is intended' (p 63), and he treats expectation in the same way. 'Our expectation anticipates the event. In this sense, it makes a model of the event...' so that 'intention reaches up to the paradigm, and contains a general rule.' (p 70) Or again in *Philosophical Grammar*, reflecting on the ideas of the *Tractatus*: 'If I try to describe the process of intention, I feel first and foremost that it can do what it is supposed to only by containing an extremely faithful picture of what it intends.' (p 148) The essential feature of these states of mind -- that which gives them the multiplicity which is reflected in their linguistic description, and which distinguishes them from such primitive processes as salivation -- is the way they involve mental representations. These representations in turn constitute a system of the same general kind as spoken language, not less powerful, and apparently more basic; for evidently the use of language itself depends upon prior representation in thought.

7. How mental representation informs linguistic representation.

This enables Wittgenstein to explain how mental representation animates linguistic representation, that is, how the activity of thought 'breathes life' into signs in use. The basic idea, is simply, e.g., that 'when you utter the word 'Napoleon', you designate that man and no other.' (*Philosophical Grammar*, §61). Thus when a speaker utters 'Napoleon loves Josephine' she designates (or means) Napoleon by 'Napoleon', the relation of loving by 'loves', and Josephine by 'Josephine'; and by uttering the words in that order, she also means that the thing designated by the first bears the relation designated by the second to the thing designated by the third -- that is, that Napoleon loves Josephine.

Wittgenstein describes this by saying that such a speaker maps the sentence 'Napoleon loves Josephine' to the situation in which Napoleon loves Josephine, that is, uses the sentence as a model, or projection, of that situation. As he says:

3.1 In a proposition a thought finds an expression that can be perceived through the senses.

3.11 We use the perceptible sign of a proposition (spoken or written, etc.) as a projection of a possible situation. The method of projection is to think the sense of the proposition.

3.12 I call the sign with which we express a thought a propositional sign. -- And a proposition is a propositional sign in its projective relation to the world.

As Wittgenstein says elsewhere, what a sentence represents is its sense (2.221), and this is a possible situation (4.03, above). So thinking the sense of the proposition is simply thinking of the sentence as representing the situation which one means it to, and thereby linking it with the situation in which it is true. In doing this, as Wittgenstein says, anyone who 'utters a sentence and *means* or *understands* it' is 'operating a calculus according to definite rules.' (*Investigations* §81). A speaker is able to this in virtue of having learned a natural language, and hence having learned what speakers of this language use its words to stand for, and also the representational significance of uttering these words in one order as opposed to others.

We know this representational significance in the sense that we know what a person would mean by the various correctly ordered utterances (grammatical sentences) of a language we understand. But of course, as observed above, we cannot specify the rules of grammatical combination for the various sentences of such a language, nor again the rules of projection which relate these to the various worldly situations in which those sentences are true. As Wittgenstein says

4.002 Man possesses the ability to construct languages capable of expressing every sense, without having any idea how each word has meaning or what its meaning is -- just as people speak without knowing how the individual sounds are produced.

Everyday language is a part of the human organism and no less complicated than it.

It is not humanly possible to gather immediately from language what the logic of language is...the tacit conventions on which the understanding of everyday language depends are enormously complicated.

The complex rules of combination (syntax) and projection (semantics) of natural language are not *known by us*, but rather *embodied in us*, like the rules of phonetics, as 'part of the human organism'. The same presumably holds for the rules of combination and projection of thought itself. This embodiment in turn makes it possible for us to grasp the sense of a sentence, or the thought expressed by it, in the way Frege described. For the embodiment includes rules relating words to things and sentences to their truth-conditions, and so implementing the mental activities which animate signs.

8. First-person or subjective semantics.

On this account words and sentences are *given* the meanings they have in use, by us and our activities of thought: by 'the mean-ing [we] put into the sentence' as Wittgenstein describes the idea in *Philosophical Investigations* §186. This is, therefore, in part a first-person account of meaning. My words owe their meanings to my meaning things by them, and so what I mean by my words is something about which I have first-person authority. Hence, as David Kaplan says in discussing this idea as it arises in Frege and Russell, this can be regarded as a *subjectivist* conception of semantics. As David Kaplan says,

...In using language we merely encode what is already thinkable...When we speak, we assign meanings to our words; the words themselves do not have meanings. These assignments are, in theory, unconstrained (except by whatever limitations our epistemic situation places on what we can apprehend). In practice, it

may be prudent to try to co-ordinate with the meanings others have assigned, but this is only a practical matter...Although...meanings may be regarded as objective, in the sense that the same possible meanings are accessible to more than one person, the *assignment* of meanings is subjective, and thus the *semantics* is subjective. Since each individual user must *assign* meanings rather than receiving them with the words, each user's semantics is autonomous. What the language community does is make available to each of its members is a syntax, an empty syntax to which each user must add his own semantics.

On this view words and sentences attain common, public meanings (insofar as they do) just because each person assigns the same meanings to them as others do; and so our success in communication depends upon our being able to understand the meanings others assign, and to assign likewise, or at least in ways they can understand. In speaking English we mean the property *red* by 'red', but we could assign this word another meaning, and are restricted in our assignments mainly by the need to understand and be understood by others. Likewise in learning language one learns the words and sentences of the language, and also learns what meanings the people in one's linguistic community assign to them, as Augustine describes the infant doing in the quotation with which Wittgenstein begins the *Investigations*. So if we speak of a linguistic rule -- for example the rule that 'Napoleon' stands for Napoleon, or 'loves' for the relation of loving -- we are also speaking of a *convention or form of agreement among persons* in their assignments of meaning. (Thus Wittgenstein speaks of 'tacit conventions' in 4.002 above.)

This account seems latent in many claims about meaning advanced by philosophers and others; and it seems particularly plausible for words such as 'pain'. It seems that each of us is introspectively presented with the subjective feeling of pain, and that to understand the word 'pain' is to assign this feeling to the word, thereby establishing the relation of reference which holds between word and feeling. In this we think of the laying down of this rule, or the establishing of this practice, as effected by a mental act conducted in the first-person perspective of our own case. And this, of course, marks a clear point of contact between the *Tractatus* account of meaning and the *Investigations* treatment of private language. For what is at issue in the latter discussion is precisely the viability of subjectivist semantics for words for sensations, a case in which the idea seems so plausible that we are inclined to accept it without even noticing that we are doing so.

9. Thought, action, and rules of projection.

We have seen that both mental and linguistic representation involve representing elements which stand for things in the world and are related to one another in combinations or configurations which represent how things are. Someone who uses a sentence to say something does so in accord with the rules of syntax and semantics for a particular natural language. The rules of combination and projection for mental representation, however, have even wider scope, for they govern all thought, and hence all intentional action. Thus a person who acts intentionally starts out with a mental representation of what is intended, and this is linked to the intended situation by the rules of projection which connect this representation to the world. So in acting to bring about the intended situation, an agent acts in accord with these rules of projection; and if successful, the agent actually brings about the situation, and so follows out the whole pattern covered by the rule.

Wittgenstein discusses this aspect of his account in a particularly pregnant series of remarks, in which he indicates the full account of representation and the mind which underpins his work.

4.14 A gramophone record, the musical thought, the written notes, and the sound-waves, all stand to one another in the same internal relation of depicting that holds between language and the world.

They are all constructed according to a common logical pattern.

(Like the two youths in the fairy-tale, their two horses, and their lilies. They are all in a certain sense one.)

4.0141 There is a general rule by means of which the musician can obtain the symphony from the score, and which makes it possible to derive the symphony from the groove on the gramophone record, and, using the first rule, to derive the score again. That is what constitutes the inner similarity between these things which seem to be constructed in such entirely different ways. And that rule is the law of projection which projects the symphony into the language of musical notation. It is the rule for translating this language into the language of gramophone records.

4.015 The possibility of all imagery, of all our pictorial modes of expression, is contained in the logic of depiction.

Here Wittgenstein takes a series of representations of the kind he has been discussing -- the musical thought, the written notes, and the groove in a gramophone record -- and indicates how he takes them to be related to one another by the concept of a rule. Since he has just previously compared musical notation with natural language, he clearly means what he says about the one to apply to the other as well. He is concerned with 'rules of projection', or semantic rules, and since he is considering the musical *thought* (Gedanke), he is considering the semantics of mind as well as language.

We can spell out the example as follows: the composer is able to have a musical thought -- to represent music in her mind -- in virtue of the way the elements of her musical thought are connected with the elements of music itself, that is, with the patterns of sound (and hence with patterns of sound-waves in the air) which are the objects of a musical thought. Just as the elements of linguistic thought are not words but are connected with the world in the way words are, so the elements of musical thought are not written notes but are connected with sounds in the way notes are. (Cf. the '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' cited above; for the musical thought the analogue would be psychic constituents 'which correspond to the notes of musical notation.')

So if, for example, the series of sounds is ABCD, the composer will be able to think of it via a combination of thought-elements, say abcd, where 'a' stands for A, 'b' stands for B, 'c' stands for 'c', 'd' stands for 'D' and 'abcd' (in that combination) for the series ABCD. The capacity to think, and hence to form 'rules of projection' of this sort, is natural to the mind; so these connections will have been set up as the composer heard music and came to be capable of remembering and thinking of it, say in childhood. Later, when she came to learn musical notation, she will have come to represent the written notes and rules of combination as well. Then she will have used her representations of the sounds and of the written notes to correlate the two, in accord with the conventions of her musical community.

Thus for the written notes 'A', 'B', 'C', etc., the composer will have formed additional thought-elements A, B, C, D, etc., and used these to learn that 'A' stands for A, 'B' for B, 'AB' in that order for the sequence AB, and so on. These are rules of projection linking language and the world which are agreed among people, and set up by the prior rules of projection which link these people's thought to both language and the world. They are analogous to the rules in accord with which a person comes to use words in the course of learning a natural language. (In this, like the infant described by Augustine in the quotation in *Investigations* §1, the composer will have also had to make use of thought-elements standing for the relation of one thing standing for another, etc.) Having learned this notation, the composer can use the series of written notes 'ABCD' to give her musical thought abcd 'an expression which can be perceived through the senses' as Wittgenstein says at 3.11. And in this she uses the written notes as 'the projection of a possible situation', namely that in which there occurs the particular sequence of sounds ABCD which

she has composed in her mind.

Both the musical thought and the written notes are thus linked by 'rules of projection' to this possible sequence of sounds, and these are also the rules which the musician must follow out in order to play the written notes correctly, that is, 'to obtain the symphony from the score'. The musician, that is, must first form in his mind a representation of the sounds which constitute the composer's musical thought, which he will do when he reads the notes and follows the same rules of projection as the composer does, and so assigns the same sounds to written notes. Then he must use this mental representation to guide his action, by forming an intention to play those notes, and fulfilling this intention. In this he follows out the rule of projection which links his thought-elements -- those in the mental representation which is part of his intention -- to a sequence of sounds, by producing that sequence of sounds itself. This, again, is parallel to what someone does in following a spoken or written instruction, rule, or order, and is an instance of the way our minds and actions are linked by a law of projection in intentional action.

But now we come to the gramophone. Even to a reader familiar with more recent thinking about the mind it may come as a shock that Wittgenstein should speak here in the same breath of human language and 'the language of gramophone records', and that he should compare the process of derivation involved in the human activities of writing or playing music to the mechanistic inscription or derivation of sound via the groove in a gramophone. He writes as if he is suggesting that we should regard the rules of combination and projection which constitute human thought and language as embodied in physical structures, and human mental representation as achieving its effects, including intentional action, by processes of physical causality.

10. Rules of projection and the mind as an information-processing mechanism.

There is a particularly straightforward interpretation of this passage, which I think should be borne in mind. This is that, as it seems, Wittgenstein is here suggesting that human thought is realized mechanically; and that he is doing so because this accords with his view. For as we have seen, he holds that we think, act, and speak in accord with complex rules of combination and projection, but 'without having any idea' as to what these rules are. The rules are 'part of the human organism, and not less complicated than it.' But then since we have no idea of these rules -- so that they are *not* 'in our minds' -- how should they part of our organism except as realized in our bodies?

Although he considers the matter elsewhere, Wittgenstein does not seem here to be asking us to think of such rules as tacitly known or somehow otherwise cognized. (The 'tacit conventions' of which he speaks as 'enormously complex' seem not to include the rules of combination and projection of thought itself, for these are not conventions.) Since cognition is supposed to be effected in accord with these particular rules this would risk regress, and we would in any case still want an account of how what is so cognized is embodied in the organism. So on this view, as it seems, the physical body (presumably the brain) *must* contain structures which effect the task of regulating thought and action in accord with these rules, and hence physical realizations or surrogates of the rules themselves. Since the task is not discharged in the conscious mind, there is no suitable locus which is 'part of the organism' besides the body, and hence the brain.

This would mean that according to the *Tractatus* the non-conscious aspects of human thinking are realized by an embodied and mechanized physical symbol system, which can be construed as a set of mental models, or again as a language of thought. (Since for Wittgenstein a language is just a particularly abstract kind of element-element: combination-combination model, the distinction between mental model and internal language is at best one of degree.) Since the symbols in question bear information about how

the world is or is intended to be, such a physical symbol system is also an information-processing device; and Wittgenstein's description seems intended to convey the idea that the working of the mind (or brain), like that of a gramophone, is to be understood in terms of the storage, transformation, and transmission of information. How far Wittgenstein actually espoused this view is uncertain, but as we shall see he gave it interesting consideration.

11. Mind-models, psychology, and the brain.

This kind of view has more recently been formulated and discussed in terms of the idea that the mind/brain can be compared to a digital computer. The *Tractatus* was composed before the modern notion of a computer was formulated or related to the mind. Still Wittgenstein seems to have considered something in the nature of precursors of these ideas in terms of his *Tractatus* account of representation, and in connection with the topic of deriving intentional action in accord with a rule. In particular, Wittgenstein described the mechanization of the mind in terms of what he called a mind model, and argued that we are inclined to think in terms of such models, and that this leads us to conclusions about the brain.

Thus in the *Blue and Brown Books* (cf 117-8) he stresses that we are liable to hold that our abilities to do things like add or multiply 'must correspond to a state of the brain, even though on the other hand we know almost nothing about such psycho-physical correspondences.' He traces this to our practice of calling

the ability to solve a mathematical problem, the ability to enjoy a piece of music, etc., certain states of mind; we don't mean by this expression 'conscious mental phenomena.' Rather a state of mind in this sense is the state of an hypothetical mechanism, a mind model meant to explain the conscious mental phenomena. (Such things as unconscious or subconscious states are features of the mind *model*.)...We regard these phenomena as manifestations of this mechanism, and their possibility is the particular construction of the mechanism itself. (117-8)

He gives an example showing the way we conceive this mind-model, when we think of it as relevant to determining whether someone is following a rule. Thus as he says in the case of someone being given orders to make sequences of movements in accord with signs:

...When we considered the question whether B was guided by the signs, we were all the time inclined to say some such thing as that we could only decide this question with certainty if we could look into the actual mechanism connecting seeing the signs with acting according to them. For we have a definite picture of what in a mechanism we should call certain parts being guided by others. In fact, the mechanism which immediately suggests itself...[is]...of the type of a pianola. Here in the working of the pianola we have a clear case of certain actions, those of the hammers of the piano, being guided by the hammers of the pianola roll. We could use the expression "The pianola is *reading off* the record made by the perforations on the roll", and we might call patterns of such perforations *complex signs* or *sentences*, opposing their function in a pianola to the function which similar devices have in mechanisms of a different type...You see here the connection between the idea of being guided and the idea of being able to read any new combination of signs; for we should say that the pianola *can* read *any* pattern of perforations, or a particular kind, it is not built for one particular tune or set of tunes...(118)

The information-bearing structure which figures in this example -- the perforated roll of the pianola -- is clearly a representation of the element-element: configuration-configuration kind specified in the *Tractatus*, and is comparable to the groove on the gramophone record which Wittgenstein considers there. (In this instance Wittgenstein adopts a characterization in terms of what we would now call a language of

thought, by explicitly comparing the pattern of perforations to complex signs or sentences, and stressing the combinatory flexibility this ensures; but he could as well have said that the perforations constitute a *Tractatus* picture or model of the notes to be played.) Such a mind-model serves to bridge psychological and mechanical functioning. The model has the multiplicity to relate to (the physical realizations of) internal and external signs, but is nonetheless clearly mechanical, and so capable of being realized by the brain. So in using psychological models of this kind we think of the brain as the locus in which the complex activities of rule-following which give life to signs is finally implemented.

12. The 'real criteria' for mental processes.

Wittgenstein links this conception with the idea that the essential aspects of rule-following (the 'real criteria' of rule-governed cognition and action) are to be found by studying the working of such an internal model. We can see why this should be so. The model is supposed to specify how a particular kind of causal and functional role -- that informing and governing the behaviour which we recognize as manifesting human thought and language -- could be discharged. Since the discharge of this role realizes actual human thinking and use of language, it distinguishes genuine cases of these from others which might seem similar but are not real examples. Hence this conception engenders the idea of a causal function which, as Wittgenstein says, is 'the function, that makes thought what it is.' So it is natural to look to a fuller specification of this function in trying to describe the essential features of thought and language; and this, indeed, is what functionalists were later to do, when the formulation of computational concepts was more advanced and better known.

As noted above, the conviction that cases of rule-following are distinguished in this way goes with the idea that there must be corresponding differences in the brain, in which the causal roles specified by such a mind-model are ultimately discharged. Hence, as Wittgenstein says,

Of course we are inclined to say that what happened in the accomplished reader and in the beginner...could not have been the same. The difference lying, if not in their conscious states, then in their brains. We here imagine two mechanisms, the internal working of which we can see, and this internal working is the real criterion for a person's reading or not reading. But in fact no such mechanisms are known to us....(120)

Parts of this discussion are repeated in the *Investigations* consideration of reading-machines (cf, e.g., §156-8) but with the reference to the pianola (and hence to the *Tractatus* conception of representation) truncated. Still there can be no doubt that in both discussions Wittgenstein is concerned with the same topic -- the derivation of action in accord with rules -- as he was considering in the passages quoted from the *Tractatus* above. (Indeed the *Investigations* discussion concerns one of the *Tractatus* examples, that of alphabetic notion considered as a representation of sound.)

13. Mind-models and the life of signs.

Wittgenstein makes the same points in other places, and in ways which are more explicitly connected with our topic. Thus in *Philosophical Grammar* he says

Knowing, being able to do something, a capacity is what we would call a *state*...If knowledge is called a [non-conscious] state it must be in the sense in which we speak of a state of the body or of a physical model. So it must be in a physiological sense or in the sense used in a psychology that talks about unconscious states of a mind-model. Certainly no one would object to that; but in that case one still has to be clear that we have moved from the grammatical realm of 'conscious states' into a different grammatical realm. (p 48)

So if we are to speak of a state of non-conscious or tacit knowledge of the rules of language, we should

construe ourselves as speaking of physiological or psychological mechanisms. And Wittgenstein makes the possible role of such tacit or unconscious knowledge particularly explicit when he says: We may say "Thinking is operating with symbols"...It is correct to say "Thinking is a mental process" only if we also call seeing a written sentence or hearing a spoken one a mental process. In the sense, that is, in which pain is called a mental state. In that case the expression "mental process" is intended to distinguish 'experience' from 'physical processes. -- On the other hand, of course, the expression 'mental process' suggests that we are concerned with imperfectly understood processes in an inaccessible sphere. Psychology too talks of 'unconscious thought' and here "thought" means a process in a mind-model ('Model' in the sense in which one speaks of a mechanical model of electrical processes.) By contrast when Frege speaks of the thought a sentence expresses the word "thought" is more or less equivalent to the expression "sense of a sentence".

It might be said: in every case what is meant by "thought" is the *living* element in the sentence, without which it is dead, a mere succession of sounds or series of written shape. (p 106)

Here Wittgenstein begins with the idea that thinking is operating with symbols, and then goes on to consider three possible conceptions of what gives sentences their life: conscious thinking; unconscious thinking, which he takes as a process specified by the kind of mind-model represented above by the pianola; and Frege's conception of a thought. Again, second of these is clearly related to the *Tractatus* account, for such a mind-model would specify the functional and causal role of the rules which regulate thought, and hence the processes which constitute it, and thereby give life to signs.

14. Mind-models and the basic unanswered question of the *Tractatus*.

Wittgenstein's interest in mind-models also seems connected with a fundamental philosophical omission in the *Tractatus* account of representation. This account, as we have seen, is built around the idea that we make ourselves pictures of facts, and hence that the elements of thought are *referentially correlated with* (that is, they *stand for*, or *refer to*) things in the world. This notion of mental reference is basic to all that the *Tractatus* says about mental representation; and it is presupposed, and given no explication, in the course of the book. But if mental representation involves a referential mapping of the elements of thought to things in the world, we can certainly ask how such a mapping is constituted. What makes it the case that a thought, or an element of thought, refers to one object rather than another, and how we are to establish that this is so? This question is not addressed in the *Tractatus*; and it is one upon which Wittgenstein came increasingly to focus, saying in *Philosophical Grammar* that it 'contains the whole question of representation.' (§62, discussed below).

More recently this has become the central question of what Fodor calls psychosemantics, the study of the semantic character of thought. As Fodor says, '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.' And although Wittgenstein proposed no answer to this question in the *Tractatus*, he seems then to have thought of it as admitting empirical study. For as we have seen, when Russell asked both what the elements of thought were, and how these elements related to what the thought was about, Wittgenstein's reply was that both these matters required psychological investigation. As he says, the relation to the constituents of the thought of the pictured fact -- that is, the psychosemantic relations, including mental reference, with which we are concerned -- 'would be a matter of psychology to find out'.

Wittgenstein thus seems to have supposed that once the abstract structure of his account of representations as element-element: combination-combination pictures was in place, the question of the actual nature of

various particular representations could proceed by straightforward empirical investigation. Just as a linguist seeks to discover the actual sounds used as words in a given language, and how they are related to things, so a psychologist might (someday) seek to discover the actual representing elements used in thought, and how they are related to things. So in considering what he calls mind-models, Wittgenstein was focusing upon the kind of model which might explicate the notion of mental reference which he had presupposed in the *Tractatus*.

In the cases upon which Wittgenstein concentrates -- those in which a person follows a written or spoken rule, a musician plays from a score, etc. -- there is a series of derivations from input to output which accord with rules of projection. As described above, these run from written representation (the musician sees the written notes) to mental representation (the musician understands the notes to be played) , and from mental representation through action to reality (the notes are played). In such cases the human organism actually produces the concrete internal representation which corresponds to the external representation upon which the agents acts, and the concrete external situation which both the external and internal representation depict. So it seems that studying this process of acting in accord with a rule would enable us to trace the links between representation and reality -- to follow out the rules of projection as they are actually realized -- and thereby to find both what the elements of thought are, and how they relate to the things for which they stand.

This has been a starting point for many approaches to psychosemantics, either on the input or the output side. Still -- and as recent work has come to stress -- there is a clear reason to regard it as unsatisfactory. For on accounts of this kind mental representation actually shows its semantic character only in instances of thought and action which are *correct*. Thus a musician whose inner representation refers to a certain note -- and so is correlated by a 'rule of projection' with the note -- will in fact play this particular note only if he *fulfils* his intention to play it, that is, if he plays (and so follows out the rule of projection linking his thought-element to that note) *correctly*. So in seeking to understand representation by studying its causal production, we must be able to determine when the causal processes in question are unfolding correctly. If not then the connections which we observe between inner and outer items may not be those of mental reference which we wish to study; but if so we are already presupposing a grasp of the notion of correct accord between representation and represented which we wish to investigate.

15. The turning point: psychosemantics and the notion of correctness.

The idea of mental reference left unexplained in the *Tractatus* thus presupposes that of correctness in rule-governed thought or action, and so cannot be explicated apart from an account of these notions as well. And we can see Wittgenstein coming to stress this and related points, as he considers how far internal representation can be understood in terms of causal or functional role. Thus take the following from *Philosophical Grammar*:

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. It contains a roll, rollers, etc., on which the piece of music is written in some kind of notation (the position of holes, pegs, and so on). It's as if these written signs gave orders which are carried out by the keys and hammer. And so shouldn't we say that the sense of the sign is in 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 sign 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?

You might then say that the sense of the signs is not their effect, but their purpose. But consider too, that we're tempted to think that this purpose is only a part of the larger purpose served by the pianola. -- This purpose, say, is to entertain people. But it's clear that when we spoke of "the sense of the signs" we didn't mean any part of that purpose. We were thinking rather of the purpose of the signs *within* 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. "I am saying 'go away' because I want you to leave me alone", "I am saying 'perhaps' because I am not quite sure." (p 70)

As stressed above, Wittgenstein has no objection to regarding a human being as a representation-driven mechanism of this kind. The idea that the mind/brain works in such a way is an empirical hypothesis, but one which we are liable to espouse prematurely, as 'a form of account which is very appealing to us' (*Investigations* §159). His point in the above passage -- as in others in *Philosophical Grammar* in which he pursues the same line of thought -- is not that this kind of account is false, but rather that it does not after all explain what it is to act in accord with a rule. In particular, accounts of this kind do not explain what it is for such action to be correct ('the signs on the role produce hisses and bangs instead of notes...what the criterion of obedience here?') and hence for an internal sign to be *rightly* correlated with one thing or another in the world. An account in terms of mind-models, or the causal role of internal signs, fails to address the central question of correctness.

Wittgenstein continues this theme in more detail in a passage in *Philosophical Grammar* which brings together a number of the ideas we have been discussing, and illustrates this point of contrast between the approaches of the *Tractatus* and the *Investigations* fairly clearly.

62. "That's *him*" (this picture represents him) -- that contains the whole problem of representation...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 I am seeing? Well, it isn't myself, but a kind of picture. But why do I call it a 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 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. Mislead 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?"

In the proposition "I believe that p is the case" we feel that the essential thing, the real process of belief, isn't expressed but only hinted at; we feel it must be possible to replace this hint by a description of the mechanism of belief, a description in which the series of words "p" would occur in the description as the cards occur in the description of the loom. This description, we feel, would at last be the full expression of the thought.

Let's compare belief with the utterance of a sentence; there too very complicated processes take place in the larynx, the speech muscles, the nerves, etc. These are *accompaniments* of the spoken sentence. And the sentence itself remains the only thing that interests us -- not as part of a mechanism, but as part of a calculus.

"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 ask: "What is a thought? What kind of thing must something be to perform the function of thought? This question is like: "What is a sewing machine, how does it work? -- And the answer which would be like ours would be "Look at the stitch it is meant to sew; you can see from that what is essential in the machine, everything else is optional.

So what is the function, that makes thought what it is? --

If it is its *effect*, then we are not interested in it.

We are not in the realm of causal explanations, and every such explanation sounds trivial for our purposes...(p 103-5)

We can summarize part of Wittgenstein's thinking in these remarks as follows. §62 Makes clear that there is a basic question, which 'contains the whole problem of representation': namely, how a thought manages to represent someone (or something). This question can be raised even when the thought is a picture constituting an intention in my own mind and represents myself. In this we see one of the main ideas of the *Tractatus*, as explicated above: the idea that an intention is to be understood in terms of a mental representation of the intended action. But here the connection between the elements of the representation and what it represents (e.g. between my image of myself and me) is no longer taken for granted, but rather seen as constituting the problem which requires to be addressed. So the problem is now precisely that of understanding the relation which was presupposed but not explicated in the *Tractatus*, that is, the relation of reference which holds between thought and its objects. ('What connection is there between your thought and Napoleon?...How does thought manage to represent?')

To make the problem vivid Wittgenstein here takes a case in which the relation of thought and its referent can seem mysterious because the referent is remote in time and space, and indeed no longer alive. ('Can you mean in Europe someone who's in America? Even if he no longer exists?') And he now makes clear that, in his view, this problem requires to be approached by considering the use of language. The relevant use in turn has two aspects: the use as it occurs the linguistic ascription and expression of thought ('the word "Napoleon" occurs in the expression of my thought'), and the use as it occurs in other contexts, including non-psychological ones ('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 Wittgenstein stresses that the connection between the use of language and what it represents is not to be understood as made by an act or process of meaning something by a word. We do of course say that we meant such-and-such a person by a word, and this is true; but this way of speaking misleads us into thinking that there is a process of meaning someone or something by a word which goes on "inside us".

§63 Continues this argument by explaining how we are misled when we think that meaning something by a word is a process "inside us". We think that this -- and the related *Tractatus* process of thinking the sense of a sentence (thinking a proposition) -- is something we could explicate via a mind-model, or an account of symbolic causal mechanisms which realize thought. In these mechanisms there would be internal surrogates of the words by which we mean things, working as part of a mechanism which both represents the situations we think about and can be linked to them causally. Then, among other things, just as we can understand the process by which the punched cards which drive a Jacquard loom both represent the pattern to be woven and also cause the weaving of it, so we could understand the process by which a plan in my mind both represents what I am to do and causes my actions in doing it. In this we should see realized the links between the sentence which describes my action, my mental representation of the action, and the action itself; and so we should see 'the kind of relation [which holds between] the constituents of the thought and of the pictured fact'. This, however, is an error. The study of causal and functional role cannot yield the understanding which is required, partly because, as remarked earlier, the relation of correctness cannot be explicated in these terms.

We can see that in these passages Wittgenstein is replacing the *Tractatus* view that the representational power of thought breathes life into signs by one in which the use of language is required to explain the representational power of thought. He is thus also rejecting a conception according to which thought is a mental state which has the full semantic content (reference and truth-conditions) ascribed to and by sentences in use, but independently of language and prior to it. This rejection is carried out in more detail in *Philosophical Investigations*.

In the *Investigations* Wittgenstein locates this suspect notion of thought in two places. First, he finds it in the kind of account of the learning of language offered by Augustine ('as if the child...already had a language, only not this one...as if the child could *think*, only not yet speak.' §32). Secondly he finds it in the Fregean notion of an act (grasping) of thought, which fixes the reference and truth-conditions of sentences in use. This he describes in terms of 'the tendency to assume a pure intermediary between the propositional *signs* and the facts' (§94). These two suspect conceptions, as we can now see, are linked and combined in the mechanistic explication of thought associated with the *Tractatus*. For according to this explication a naturally given and mechanically functioning system of internal symbols (a language of thought) serves both to realize thought prior to and independently of the acquisition of spoken language, and also to implement mental acts of understanding and meaning which assign reference and truth-conditions to sentences as they are used, as is done by the grasping of sense. (These same connections also appear in the work of Fodor, as well as elsewhere in linguistics and cognitive science.) This is why Wittgenstein takes these various conceptions of thought to stand or fall together, and why he rejects the priority of thought together with claims about its mechanization in the successive remarks we have just considered.

16. Reformulation of the basic question and the role of consciousness.

These remarks -- which we have only partly discussed -- indicate a complex transformation in Wittgenstein's views. He is now examining the basic question of the referential character of thought, which he failed to address in the *Tractatus*. In doing this, however, he is also altering the terms in which he poses the question. For he is relinquishing the view that thought is realized in an inner system of elements and combinations analogous to language, such as he described in his letter to Russell, and such as might be compared to the punched cards which drive a Jacquard loom. Since thought may not consist of such elements, the question of understanding its referential nature is not best put in these terms.

(Wittgenstein is continuing to investigate mental reference, while dropping the assumption that there is a language of thought.)

We shall consider Wittgenstein's reformulation of this question in the *Investigations* shortly. For now let us note two things about the question. First, it is very general: it applies to anything which can be thought about, including abstract entities such as the property of being red as well as concrete particulars such as Napoleon. Secondly, the question seems particularly related to conscious awareness, to which Wittgenstein often alludes in discussing it. This is because we think of such awareness as awareness *of* things, so that consciousness seems a locus of mental reference which is particularly clear and unproblematic.

In view of this it is worth noting that we can arrange items to which this question applies in a rough order by their relation to consciousness. Thus we have:

(i) Items which are parts or aspects of consciousness itself: e.g. pains and other sensations, and their phenomenological properties.

(ii) Items which are explicitly represented in consciousness, such as things perceived in the immediate environment, or (to take Wittgenstein's example above) myself and the things I intend to do, as represented in my intentions.

(iii) Items which are not explicitly represented in consciousness, but which are somehow mentally meant or referred to along with others which are. Thus, to take an example we shall consider shortly, if I tell someone to add two repeatedly to a given number, I may explicitly think of some instances of what I mean him to do; but my request also encompasses -- and in this sense also refers to, or is about -- further instances which I do not think of.

(iv) Items such as Napoleon, which may be thought about but which cannot directly impinge upon consciousness because they are remote in space or time, or no longer exist.

While it is clear that the general question of the relation of thought to its objects applies to all these kinds of items, we can see that differences among them are important for Wittgenstein's discussion. Thus in the remark considered above he uses the fact that Napoleon no longer exists to make it seem problematic that we should simply be able to mean or refer to him in thought. More background might be required to enable us to regard it as problematic that we should simply be able to mean, or mentally refer to, a content of consciousness such as a pain. These differences also play a role in Wittgenstein's argument.

17. The question of representation in the *Investigations*: rule following and the constitution of correctness.

In the *Tractatus* Wittgenstein assumed that we know how our thoughts relate to the world, and could use this to explain how our sentences do. In *Philosophical Grammar*, as we have seen, he pressed his enquiry further, and asked how thought manages to represent or refer to things. In the *Investigations*, as we shall now see, he pursues these questions still further, and in a way which is continuous with *Tractatus* 4.0141 discussed above: that is, by relating thought, language, and the things to which they refer, via the notion of a rule.

Wittgenstein discusses rules in the *Investigations* together with the now familiar opposition between acts of meaning and use as providing the life of signs. In §138 he relates these as follows:

§138. But can't the meaning of a word that I understand fit the sense of a sentence that I understand? Or the meaning of one word it the meaning of another? -- Of course, if the meaning is the *use* we make of the word, it makes no sense to speak of such 'fitting.' But we *understand* the meaning of a word when we hear or say it; we grasp it in a flash, and what we grasp in this way is surely something different from the 'use'

which is extended in time.

Here Wittgenstein's 'but we *understand* the meaning of a word when we hear or say it; we grasp it in a flash' introduces the acts of understanding and meaning which he takes us to hypostasize as giving life to signs. And although he is dealing with what he regards as a philosophical error, it is also important that the sentence he uses to introduce the error is true. Just as we really do mean or understand things by our words, so we really do sometimes grasp the meaning of a word in a flash. What happens in such a case, as Wittgenstein supposes, is just that in a flash of understanding we acquire the capacity to use the word correctly; and since this includes the practical ability to relate the word to the things to which it refers, this is also part of the capacity we acquire. But according to Wittgenstein we are liable to misconstrue this, by assuming (i) that the grasping is of something distinct from the use -- such as the concept, meaning, or Fregean sense (thought) introduced in the first sentences of this remark -- and (ii) that this distinct thing, or our grasping of it, determines the reference of the word, and hence how we use it.

18. Intermediaries and correctness: the explanatory priority of correct use.

The first of these misconstructions is expressed here in the claim that 'what we grasp in this way is surely something different from the 'use' which is extended in time.' To accept this is to take the verb 'to grasp' here as having an object -- e.g. a sense -- distinct from the use of the sign in question, and hence, in Wittgenstein's terms, 'to assume a pure intermediary between the propositional *signs* and the facts' (§94). This, as Wittgenstein expects, will go with the idea that this intermediary somehow determines reference and truth-conditions, and hence use. So this idea -- which is part of many standard expositions of the notion of sense -- is part of the topic of the remarks which follow. We have seen that in his work prior to the *Investigations* Wittgenstein focused on explicating the notion of correctness. Accordingly, we find that in these remarks he is concerned with two questions: whether the supposed grasping of distinct intermediaries can explain correctness in the use of signs; and if not, how such correctness is to be explained. Thus in the case of sense, if what is required to explain use is a *correct* grasping of sense (or a grasping of the *correct* sense), we must ask how such correctness is determined.

In the next remarks Wittgenstein carries this argument forward, and illustrates its connection with the *Tractatus*, by considering the idea that what is grasped in a flash is a picture. Thus he takes the example of the word 'cube' and a mental picture of a cube (§139), or such a picture together with a 'method of projection' (§140). (We can see that this links with the question of mental reference discussed earlier, for the question as to what relates my mental picture of a cube to actual cubes is comparable to that as to what relates my image of me to myself or my thought of Napoleon to Napoleon.) Wittgenstein treats these examples by observing that the correctness of the grasping of the picture (or sense), or the method of projection which accompanies it, is answerable to that of the use made of the sign, and not vice-versa. We can see that this dialectic would apply to any intermediary, and Wittgenstein at once extends it to the case of someone learning to write series of numerals or numbers. Here, for example, what is grasped in a flash may be an algebraic formula from which the series can be derived. But again what counts as a correct grasping, or a correct application of the formula, is determined by the use made of it: 'The application is still a criterion of understanding' (§146) Thus the Fregean idea that the grasp of sense (meaning, etc.) is explanatorily prior to that of correct use (or practice) is shown to be mistaken. A comparable argument extends to the state of understanding more generally (§147), and the 'state of a mental apparatus (perhaps of the brain)' which might be supposed to underlie it. (§149) This last argument repeats points considered above, for the 'mental apparatus' here is the 'mind-model' of the previous discussions, and this supposed mechanism is again understood on the model of the pianola (§157) and

realized by 'what goes on in the brain and nervous system' (§158). So according to these arguments nothing in our understanding of meaning is explanatorily prior to that of correct use (or practice); for we determine the correctness of the mental acts, states, or mechanisms which we associate with meaning by reference to this.

19. Determining the correctness of use.

This establishes the basic role of correct use, but leaves us with the question as to what constitutes such correctness and how we recognize it. This is the question Wittgenstein starts to address in the next phase of his argument. In considering the pupil learning the number-series, he has so far taken it for granted that the way we are trying to get him to the series is correct. ('...he continues the series correctly, that is, as we do it.' (§145)). He now extends his enquiry, to ask how this correctness -- which includes the correctness of our own use -- is itself determined. And since he is raising this question about the notion which is most basic for the explanation of correctness generally, he is in effect asking what determines the correctness of all the other notions which he has considered. Hence, as we shall see, he is in effect asking what makes it the case that we are able to think, judge, or use words in accords with rules or norms at all.

Wittgenstein does this, as it will be important to note, by linking the notion of correctness with that of *interpretation*. He considers a pupil being taught the use of a simple arithmetic rule, such as that for adding two. This pupil is supposed, among other things, to learn to continue the series 2, 4, 6, 8,and so on indefinitely. The question then is, in what circumstances are we to interpret the pupil as having learned to follow the rule *correctly*? Wittgenstein notes that a learner might continue the series correctly (as we should say) up to 1000, but then go on to write 1004, 1008.... We should regard this as incorrect, but it would not necessarily show lack of an understanding (some understanding) on the pupil's part. It might be that going on in this different way was natural to the pupil, and we might find an interpretation which explained this, and according to which it was indeed the correct thing for him to do.

§185...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 at 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 naturally to this person to understand our order and our explanations as we should understand the order "Add 2 up to 1000, 4 up to 2000, 6 up to 3000 and so on."

Such a case would present similarities with one in which a person naturally reacted to the gesture of pointing with the hand by looking in the direction of the line from finger-tip to wrist, not from wrist to finger-tip.

This illustrates, as Wittgenstein says later, that "there are criteria for his 'thinking he understands', attaching some meaning to the word, but not the right one." (§269) So we are immediately faced with a deeper question. How do *we* know that *we* are supposed to follow the rule for adding two in the particular way that we do -- how we know that our practice, as opposed to that of the person we treat as deviant, is actually *correct*? This is the topic which Wittgenstein makes explicit in the next remark, by raising what is now the fundamental question of correctness: "To carry it out correctly! How is it decided what is the right step to take at any particular stage?"

§186 What you are saying then, comes to this: a new insight -- intuition -- is needed at every step to carry out the order '+ n' correctly. -- To carry it out correctly! How is it decided what is the right step to take at any stage? -- "The right step is the one that corresponds with the order -- as it was *meant*" -- So when you gave the order +2 you meant that he was to write 1002 after 1000 -- and did you also mean that he should

write 1868 after 1866, and 100036 after 100034, and so on -- an infinite number of such propositions? "No: what I meant was, that he should write the next but one number after *every* number that he wrote; and from these all those propositions follow in turn." But that is just what is in question: what, at any stage, does follow from that sentence. Or again, what, at any stage, are we to call "being in accord" with that sentence (and with the *mean*-ing you then put into that sentences -- whatever that may have consisted in.) It would almost be more correct to say, not that an intuition was needed at every stage, but that a new decision was needed at every stage.

And this question, as Wittgenstein makes clear, seems exceedingly difficult to answer. Clearly no answer can be based on the acts or states of understanding or grasping a sense, or the mental or physiological mechanisms assumed to underpin them, which have been considered in the argument so far; for the status of these things as correct depends upon the use to which they are related, and it is the correctness of this which is now in question. Wittgenstein has his interlocutor turn again to the idea that this question is answered by reference to acts of meaning ("The right step is the one that accords with the order -- as it was *meant*"), but this fails to extend to the indefinitely large number of cases covered by the rule ('and did you also mean that he should write 1868 after 1866, and 100036 after 100034, and so on -- an infinite number of such propositions?--') Nor is it to the point to say that this further reach of the rule is determined by what *follows from* the rule, for this is equally in question. The question as to what accords with a rule is that of what accords with a sentence (or sense or meaning) generally, and this includes deductive accord as well. (And the question concerns the constitution of correctness in applying rules generally, and so applies to the rule for *modus ponens* as much as to that for adding 2.)

20. Correct use and sentential accord generally.

We thus seem faced with a deep and general problem, which Wittgenstein again puts in terms of the notion of interpretation as follows:

§198 "But how can a rule show 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 with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning...

This same question also extends to intention, desire, belief, and so forth, that is, to the states of mind which we describe in the *that P* way we have been considering. For it is clear, for example, that a question as to which action really accords with the mathematical rule expressed by the sentence 'Add 2' is also a question as to which action really accords with the intention or desire described by the same sentence, that is, the intention or desire to add 2. Wittgenstein makes this explicit in §197, when he turns from the case of rules to that of intentions, and hence of mental states in general.

...we say that there isn't any doubt that we understand the word, and on the other hand its meaning is in its use. There is no doubt that I now want to play chess, but chess is the game it is in virtue of its rules (and so on). Don't I know, then, which game I want to play until I *have* played it? or are all the rules contained in my act of intending? Is it experience that tells me that this sort of game is the usual consequence of such an act of intending? so is it impossible for me to be certain what I am intending to do? And if that is nonsense -- what kind of super-strong connexion exists between the act of intending and the thing intended?

We can now see that the original question relates to the link between 'the act of intending and the thing intended', and so to all items or states which have *sentential content* -- content *that P*, which is assigned by the use of a sentence 'P' -- and any of the actions or items in the world which are supposed (or meant)

to *accord with* these bearers of content. Thus consider any sentence 'P' which can be used to specify something a person can do. ('P' might be 'Take the square root', 'Turn left', 'Find something *this* colour', 'Create a diversion' or whatever.) For each such 'P' we have the same questions: What makes it the case that it is correct to act in accord with 'P' in one way rather than another; and how do we know that this is so? And again we have corresponding questions about intention, belief, desire, and other states of mind. What makes it the case, and how do we know, that one action rather than another fulfils the desire or intention that P, or renders the belief that someone has done this true?

These questions relate equally to thought (propositional attitudes) and language, and concern both the constitution of the norms we take to govern these phenomena, and our knowledge of these norms. So, in Wittgenstein's terms, they are also questions about 'the hardness of the logical must' -- questions as to what constitutes this 'super-strong connection', and how we can know about it. So as he puts the question for propositional attitudes more generally

437. A wish seems already to know what will or would satisfy it; a proposition, a thought, what makes it true -- even when that thing is not is not there at all. Whence this determining of what is not yet there? This despotic demand? ("The hardness of the logical must.")

We take thoughts and motives, words and sentences, and worldly things and situations to be related in a way which we describe in terms of *normative accord*, that is, accord which can be regarded in one way or another as correct or incorrect. As illustrated, we speak of such accord in many different ways: in terms of the *truth* of sentences or thoughts, the *fulfilment* of intentions, the *satisfaction* of desires or wishes, the *realization* of hopes and fears, the fact that a step *follows from* a premise, or *according to* a logical or mathematical rule, and so on. And as §197 and §437 make clear, since this includes the accord of states of mind and sentences with objects and situations in the world, it also encompasses the relation of thought to its objects which Wittgenstein considered in the remarks from *Philosophical Grammar* quoted above. This question, in turn, is central to our sense of ourselves as agents who can knowingly think, speak, and act. Knowing the contents of our thoughts, intentions, or sentences is knowing the actions, objects, or states of affairs which are supposed to accord with them in this sort of way; and these are matters which we take ourselves not only to know about, but to know more about than others characteristically do or can know -- matters in the sphere of our first-person authority about our own minds and meanings. Yet although such knowledge seems absolutely fundamental to us, we also seem quite unable to give any account of it, or to describe any justification for it. We acknowledge the normative requirements of thought and language spontaneously and without reflection, and we take them for granted in what we say, think, and do. Trying to answer Wittgenstein's explicit questions, however, we can seem quite unable to elucidate either the basis of these requirements or the knowledge of them which comes so readily to us. Hence as Saul Kripke puts the point in his celebrated exposition, if we take Wittgenstein's remarks as sceptical 'It seems that the entire idea of meaning vanishes into thin air.' In this case scepticism would represent '*all* language, *all* concept formation, to be impossible, even unintelligible', so that 'assertions that anyone ever means anything are meaningless.' Such scepticism is thus 'insane and intolerable'

21. Another version of the question.

Wittgenstein's argument here can be seen to turn upon the way we articulate mental states by describing them by embedded sentences. We can represent this by observing that each of us speaks and understands an idiolect, which we take to be that of a natural language. In understanding the sentences of our idiolects we know the conditions in which they are true. So for example each speaker of an idiolect of English knows

The sentence 'Freud worked in Vienna' is true (in my idiolect) just if Freud worked in Vienna.
 The sentence 'Wittgenstein lived in Vienna' is true (in my idiolect) just if Wittgenstein lived in Vienna.
 The sentence 'The moon is blue' is true (in my idiolect) just if the moon is blue.

and so on.

Each of us thus knows, or can readily come to know, an indefinitely large number of truths relating the sentences of his or her idiolect to objects and situations in the world. We can schematize this by saying that each knows indefinitely many instances of the form:

T: 'P' is true (in my idiolect) just if P

Where occurrences of 'P' might be replaced by appropriate sentences of the idiolect.

In commonsense psychological thinking we re-employ these sentences in the familiar and significant way Wittgenstein stresses. In understanding people we use a vocabulary of words like 'desires', 'believes', 'hopes', 'fears', etc., each of which admits of complementation by a further sentence. So we speak of the desire, belief, hope, fear, etc., *that P*, where 'P' can be replaced by a sentence suitable for specifying the object, event, or situation towards which the motive is directed. This assignment of sentences to motives enables us to use a finite stock of psychological concepts as the basis for a potential infinity of ascriptions of mental states, and allows the truth-conditions of sentences to do duty as conditions of satisfaction, fulfilment, realization, etc., which we ascribe to motives themselves. Thus we have, e.g.

A belief described by 'P' (a belief that P) is true just if 'P' is true, that is, just if P.

An intention described by 'P' (an intention that P) is fulfilled just if 'P' is true, that is, just if P.

A desire described by 'P' (a desire that P) is satisfied just if 'P' is true, that is, just if P.

A fear described by 'P' (a fear that P) is realised just if 'P' is true, that is, just if P.

A hope described by 'P' (a hope that P) is fulfilled just if 'P' is true that is, just if P

etc.

In light of this we can see the various notions of normative accord which we apply to motives -- fulfilment, satisfaction, realization, etc. -- as analogues of the notion of truth as applied to sentences. But then, clearly, our use of words and sentences in describing thoughts and others states of mind -- both in our own case, and that of others -- presupposes and depends upon our use of the same words and sentences to describe the world, as schematized in **T**.

For each of us, however, **T** represents a massive (potentially infinitary) claim to empirical knowledge about our own uses of sentences; and this is something about it seems we could be wrong (and in respect of which persons with certain neural deficits perhaps actually are incompetent). I take it that I use the word 'Freud' to refer to a certain man, and the word 'Vienna' to refer to a certain city, and so hold the sentence 'Freud worked in Vienna' to be true just if that man (Freud) worked in that city (Vienna). But what guarantees that I am right about this? What makes it the case, and how does each of us know, that the words and sentences in his or her idiolect -- and hence the desires, beliefs, thought, feelings, etc., which we articulate by these sentences -- *really* relate to the world in the way we *think* they do?

This is another version of Wittgenstein's question. (Wittgenstein concentrates on what he takes to be the basic case, in which the question is that of *acting* in accord with a rule, and hence in accord with sentence of one's idiolect; but this is encompassed in the broader formulation.) As before, the question relates not only to the sentences by means of which we describe the world, but also to the whole range of states of mind whose descriptions embed those sentences, and the referential relations carried by the words in them. So again the question extends through the range of our first-person authority. Hence once we pose

the question, it becomes clear that there is a sense in which we cannot answer it. If I am asked how I know that I hold the sentence 'Freud worked in Vienna' as true in my idiolect just if Freud worked in Vienna, there is nothing to the point which I can reply besides insisting that I do. For clearly anything else I might think or say is equally in question.

22. Wittgenstein's solution: an approach via radical interpretation, focusing on 'regular connection' between verbal and non-verbal actions.

We have now seen how Wittgenstein's thinking from early to late is linked by a concern with mental and linguistic representation. In the *Tractatus* he took mental reference for granted, and used this to explain representation in language. In his middle period he began to enquire into the presuppositions of this earlier account, and sought to understand mental reference in connection with the notion of correctness. So in the *Investigations* he addresses these matters together, through a question which encompasses both thought and language. We take both our sentences ('Add 2') and our thoughts (the intention to add 2) to relate to our actions, and to the world more generally, in ways we describe in terms of normative accord. What makes such descriptions true, and how do we come to know them?

We saw that Wittgenstein raised this question by reference to interpretation -- by citing the possibility of an interpretation which represents an intuitively mistaken way of following a rule as correct in some different or unexpected sense, and by stressing that interpretations alone do not determine meaning. Also I think he solves the problem by reference to interpretation; this emerges in the following remarks, which are, even for him, unusually difficult to understand.

§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. For there is no regular connection between what they say, the sounds they make, and their actions; but still the sounds are not superfluous, for if we gag one of their 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'.

In these remarks Wittgenstein at first explicitly states the question of accord which he has been raising, using the examples of rules and orders, and people who respond to these in different ways ('Which one is right?'). He then replies to his own question indirectly, by describing an hypothetical radical interpreter, similar to the figure considered later by Quine and Davidson, who has the task of making sense of both the interpretees' utterances and actions, without prior knowledge of either. And here he makes a series of claims about the finding of empirical regularities ('regular connection') between sounds and actions which make interpretation possible.

What are we to make of these claims, and how do they constitute an answer to the general questions about

correctness or normative accord which Wittgenstein has raised? Without going further into exegetical detail, I think we can take Wittgenstein here to be making a series of related points which we can partly bring out as follows. We are concerned with the interpretation of speech, and speech is a kind of behaviour which has a marked and complex (syntactic) order, in virtue of which we can understand it with particular clarity and precision. But as §207 suggests, speech is also a kind of behaviour which cannot be understood in isolation from the rest of the behavioural order of which it is a part. (To see this consider trying to work out the meaning of radio broadcasts in a foreign language on the basis of their syntax alone, in the absence of any clue which might provide information as to what the broadcasts were about.) Despite the order in people's productions of sounds or marks, if these were not a co-ordinated part of a larger pattern of action, we could not interpret such sounds or marks, or regard them as language at all. By contrast, we can understand the order in much non-linguistic behaviour without relying on speech, at least up to a point. We can generally see the purposive patterns in people's behaviour in terms of their performance of commonplace intentional actions, and their being engaged in various everyday projects: 'the usual human activities' which constitute 'the common behaviour of mankind'. But as Wittgenstein has previously stressed, unless we can link such actions with speech, we cannot, in many cases, know what people think; and in the absence of speech it would be doubtful how far we could ascribe precise thoughts or motives to people at all (Cf. §25, §32; and also §342).

So we have a general claim about interpretation and understanding. Words with no relation to deeds are unintelligible, and deeds with no relation to words are inarticulate. It follows that the kind of understanding of people which we actually attain, in which we take deeds to spring from motives with precise and determinate content, requires that we integrate our understanding of verbal and non-verbal action, and hence that we correlate and co-ordinate the two. In doing this we tie the complex order present in utterance to particular points in the framework of action and context, so as to interpret language; and this in turn enables us to interpret the rest of behaviour as informed by thought which, like that expressed in language, has fully articulate content.

23. Regular connection and intention.

Wittgenstein makes this conception of the relation of interpretation, utterance, and action clearer in further remarks. Thus in §243 he writes:

A human being can encourage himself, give himself orders, obey, blame, and punish himself; he can ask himself a question, and answer it. We could even imagine 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.)

But could we also imagine a language...the individual words of [which] are to refer to what can only be known to the person speaking; to his immediate private sensations...[?]

This again involves the figure of the explorer/interpreter, in what is clearly a variation on the theme of §206-7. As before, according to Wittgenstein, we find meaning where we find a kind of correlation between utterance and action which a radical interpreter can specify. Here the idea is applied to 'resolutions and decisions', that is, to utterances which express the interpretee's intentions in the actions with which they can be correlated. Just as in §206-7 the radical interpreter is required to find regular connections between utterances of orders and actions which are obeyings of those orders, so in §243 the radical interpreter is required to find regular connections between expressions of intention and actions which are fulfillings of those intentions. In the earlier remarks Wittgenstein claimed that such correlations

are necessary for interpretation, and here he adds that they also suffice for it. For given such correlations the explorer-interpreter can both 'predict these people's actions correctly' and 'succeed in translating their language into ours.'

This remark also makes more explicit how Wittgenstein takes the specification of this kind of regularity to bear upon mental states like intention and desire. In correlating sentences expressing 'resolutions and decisions' with actions in this way, an interpreter is relating specifications of intention to the actions which fulfil those intentions. This is particularly obvious in the case of resolutions and decisions, but it also applies to the rules and orders which Wittgenstein considers in other remarks; for in general the linguistic expression of a rule or order can also be regarded as a specification of the intention with which the person who follows the rule or executes the order thereby acts.

To frame hypotheses about such correlations, as Wittgenstein observes, the interpreter/explorer must use his own language. So we can say that in specifying an interpretive regularity of the kind with which Wittgenstein is concerned, an interpreter seeks to map a single sentence of his or her own idiolect (i) to the interpretee's verbal behaviour, construed as an utterance of a sentence expressing a rule or order, or again a decision, resolution, or the like; and (ii) to the interpretee's non-verbal behaviour, construed as an action in accord with the rule, order, decision, etc.; and so (iii) to the interpretee's intention in acting, which is taken as expressed by the sentence in (i) and therefore fulfilled by the action in (ii). So an interpretive understanding of a single regular connection of this kind represents these three elements -- the interpretee's utterance, intention, and action -- as at once empirically correlated and in normative accord.

24. Interpretation is both empirical and normative, and so provides the kind of account of correctness which is required.

Thus in §206-7 and §243 Wittgenstein sketches a conception of interpretation which is at once empirical and normative. Interpretation consists in the detection and understanding of a natural order in behaviour, which includes regular connections between utterances and other actions; so this can be considered as a form of empirical enquiry. The empirical order which interpretation detects, however, is also a normative order. In interpreting even non-verbal actions we find them to be, as Wittgenstein says, intelligible or 'logical', that is, to accord with standards of intelligibility and logic; and in discerning the regular connections between verbal and non-verbal action which are required for language we find the non-verbal actions to occur in normative accord with the verbal ones, for example as followings of rules or orders which are expressed in utterances.

Interpretation thus discerns a kind of regularity which is genuine and which can be objectively characterized; and interpretive characterizations of this order are in terms of the norms and rules about which Wittgenstein has raised his constitutive and epistemic questions. So as §206-7 implies, these questions can be answered by reference to interpretation. Correctness is constituted as a form of natural order, and this kind of order is understood in interpretation. Roughly, what makes it the case that some particular thing I do is correct or incorrect is that my behaviour manifests the kind of order which renders it interpretable, so that there are objectively ascribable but normative regularities in terms of which it can be understood; and this particular thing coincides or fails to coincide with one of these regularities. So reference to interpretation provides an answer to Wittgenstein's general questions, that is, an account as to how these normative connections are both constituted and known.

The instances of 'regular connection' of which Wittgenstein speaks in §207 thus have a triple status: they obtain; we interpret them in normative terms; and they sustain interpretation of this kind. Hence we can see them not only as causal regularities, but as regularities which correctness demands, and which support

the kind of interpretation in which correctness or the lack of it is ascribed. This status explains the apparently 'superstrong connection' or the 'despotic demand for what is not yet there' which hold between the act of intending and the thing intended, or other elements of interpretive regularity generally. And it emerges clearly in the examples with which Wittgenstein both illustrates the questions and attempts to dissolve them.

§198. "But how can a rule show 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 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 connection 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 connection; 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. The notion of 'regular use' or 'practice' here introduces the notion of interpretable regularity which we have been considering above. Wittgenstein's example is the practice of going by a sign-post, in which there is a regular connection, brought about by training, between the sign and actions which accord with it. (The sign-post is of course not an utterance of an interpreter; but still it can be regarded as a concrete instance (token) of one of an interpreter's sentences, and so as involving the kind of sign/action regularity described above.) Thus an interpreter could use observation of behaviour connected with the sign-post to work out that the sign meant, say, 'turn left'; and also a person who used such a sign could point to the sign-post itself as part of specifying a rule or giving an order, or as specifying his or her desire or intention to act accordingly, that is, to turn left.

To interpret the sign this way is perforce to hold that a person trying to act in accord with it by turning right would not be acting in accord with it, and so in that sense would be behaving incorrectly. The sign-action regularity thus covers behaviour which both has a causal explanation, and can also be assessed for correctness. The regularity of which the sign is part is also essential to this potential for correctness, since, as §207 makes explicit, a degree of regularity in persons' behaviour in relation to sign-posts (use of the sign) is required for the cogent assignment of an interpretation to the sign, and hence also to the ascription of the desire or intention which agents link with the sign; and such an interpretation also specifies the norm against which correct use of the sign is assessed. So, in Wittgenstein's terms, we begin to understand 'what this going-by-the-sign really consists in' when we see the matter both as one of causal connection and also in terms of the linked notions of practice, interpretation, and correctness.

25. The interpretive discernment of regularity can be compared to the framing of empirical hypotheses and the finding of natural laws.

The question of determining the right way to follow a signpost, or a rule in general, is thus an empirical one: it is that of charting a 'regular connection', that is, finding the right characterization of a certain regularity. In this case, however, the regularity is that of behaviour related to a norm or standard of correctness. This also yields answers to the specific questions Wittgenstein poses. Consider first the analogue of that in §206: What should we say if one person responds in one way and another in another to the sign-post and the training connected with it -- which one is right? On the exegesis so far this is

straightforward. If the best interpretive explanation we can give of the role of the sign in the lives of those who use it is that it means 'turn left', then someone who responds to the training and the sign by turning right is so far responding incorrectly. This, indeed, is comparable to the case with which the enquiry began, of the learner who responds to our training with '+2' by going on '1004, 1008'.... As in that case, misinterpreting the sign can be also compared with misinterpreting the gesture of pointing; and indeed if we take the kind of sign-post Wittgenstein actually describes in his argument [§85 'A rule stands there like a sign-post...But where is it said which way I am to follow it; whether in the direction of its finger or (e.g.) in the opposite one?'] we see that the comparison is nearly exact.

As before, someone who is acting incorrectly may yet have what deserves to be called his own understanding of the training and the rule. As Wittgenstein holds, this should also show in regular behaviour on his part, which we should be able to interpret. If we succeed in formulating the way this person understands 'turn left', then there will also be the possibility that he will fail to act in accord with the rule as he understands it. Thus someone might regularly turn right at the sign, leading us to suppose that he understood it this way; then on occasion he might encounter the sign (or in another case hear the order 'turn left') and turn left, but then correct himself, and turn right. This too we could interpret, for as Wittgenstein emphasizes, self-correction -- and other kinds of behaviour which show sensitivity to norms -- are also observable aspects of our natural history, and hence material for radical interpretation.

§ 54. Let us recall the kinds of case where we say that a game is played according to a definite rule. The rule may be an aid in teaching the game. The learner is told it and given practice in applying it. -- Or it is an instrument of the game itself. -- Or the rule is employed neither in the teaching nor in the game itself; nor is it set down in the list of rules. One learns the game by watching how others play. But we say that it is played according to such-and-such rules because an observer can read these rules off from the practice of the game -- like a natural law governing the play.-- But how does the observer distinguish in this case between player's mistakes and correct play? -- 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 recognise that someone was doing so even without knowing his language.

Although this remark is an early one, it has a clear bearing on the themes we have been considering, which are made fully explicit only later in the book. In particular, since Wittgenstein compares games and language, his focus on a game learnt solely by observing the behaviour of others anticipates his remarks about the explorer/interpreter, who learns the language of the monologue people through observation in precisely this way. We may suppose, moreover, that such an interpreter would be aided in this work by attending to the 'characteristic signs', stressed here, of people's awareness of the relation of their own behaviour to their own norms, which can be recognised 'even without knowing [their] language'. We can thus see this early remark as ending with a reference to the idea of radical interpretation which Wittgenstein takes up more explicitly via the role of the explorer in §207 and §243. Also we can see that in this early remark too Wittgenstein is considering interpretable regularities, which in this case can be 'read off from the practice of the game -- like a natural law governing the play.' Here again Wittgenstein indicates that these are at once natural regularities subject to disciplined empirical study ('like a natural law'), and also regularities which have the further status of activity in accord with rules or norms; and that this is reflected in further observable behaviour relating to them. Wittgenstein sketches his views repeatedly.

The comparison of radical interpretation to the working out of the rules of a game on the basis of observation serves also to reinforce other points in our exposition so far. In remarks plainly continuous with this, but applied explicitly to language, Wittgenstein constructs nearly the same argument for the case of rules in the home language as he will deploy in the case of radical interpretation considered in §207.

82. What do I call 'the rule by which he proceeds'? The hypothesis that satisfactorily describes his use of words, which we observe; or the rule which he looks up when he uses signs; or the one which he gives us in reply if we ask him what his rule is? -- But what if observation does not enable us to see any clear rule, and the question brings none to light?...What meaning is the expression "the rule by which he proceeds" supposed to have left to it here?

An interpreter trying to discern rules of language can construct hypotheses to account for the linguistic behaviour she observes, and can also make use of the interpretee's own expressive or self-ascriptive account. Interpretive hypotheses will naturally enjoy a maximum of support when these sources of information coincide, as they do in instances of interpretive regularity. If, however, no such hypothesis is satisfactory, then the notion of rule may be inapplicable to the data of observation; there is, as he says in §207, 'not enough regularity' in the data for us to describe it in terms of the concept of a linguistic rule.

26. Interpretation ratifies the first-person perspective on mind and meaning.

This clearly also applies to the alternative formulation of Wittgenstein's question above, namely, what makes it the case, and how am I to know, that the sentences of my idiolect relate to my own activities and to the rest of the world as I take them to? It is clear that I cannot answer this by *saying* how my sentences relate to these things, for this 'still hangs in the air together with what it interprets, and cannot give it any support' (§198). In this context my interpretations of my own rules are no more than 'the substitution of one expression of the rule for another', and do not determine what they mean. Rather this is determined by the *use* I make of my sentences, which manifests '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). This is my linguistic practice, which another can interpret as in regular connection with my non-verbal actions, and thereby understand in terms of my fulfilling intentions, following rules, and other forms of normative accord. And although I do not (and could not) interpret myself in this way, this also constitutes my first-person expressions of my own meanings and motives as correct and authoritative. So Wittgenstein's point is not that we face a problem in justifying our assignments of meaning, or the way we follow rules. It is only that we are bound to *think* that there is such a problem (and indeed an insoluble one) so long as we do not acknowledge the asymmetric role of interpretation, and hence suppose that what can be cited in justification of the ways we think and act must somehow be employed or available to each of us in the Cartesian perspective of our own case. Hence according to Wittgenstein the first-person perspective -- that in we are authoritative about mind and meaning, and in which we find no doubt or indeterminacy -- does not contain within itself the materials required for its own justification. Like Descartes, we tend to assume that where clarity and certainty are, there ground and justification must also be, and this is an error. The beliefs which characterize the first-person perspective are constituted as knowledge by a possible relation to others, which consists in our being such as to be interpretable by them.

27. Expression, interpretable regularity, and further propositional attitudes.

The *Investigations* thus replaces the first-person perspective on mind and meaning of the *Tractatus* with an account grounded in the interpretive apprehension of the natural regularities which, in his terms,

constitute our 'form of life' (§241). These regularities, however, also relate to the first-person perspective, for they encompass utterances which express intentions and other sententially described motives upon which non-verbal actions are based. Thus according to Wittgenstein in the verbal expression of pain 'words are connected with the primitive, the natural, expressions of the sensation and used in their place' (§244); and something similar holds for wishing, expecting, and the like. As he says

441. By nature and by a particular training, a particular education, we are disposed to give spontaneous expression to wishes in certain circumstances. (A *wish* is, of course, not such a 'circumstance'.) In this game the question whether I know what I wish before my wish is fulfilled cannot arise at all. And the fact that some event stops my wishing does not mean that it fulfils it. Perhaps I should not have been satisfied if my wish had been satisfied....Supposed it were asked "Do I know what I long for before I get it?" If I have learned to talk, then I do know.

We can see how this relates to such remarks as §206-7 and §243 on interpretation. The natural capacity for self-expression cited here ('By nature...we are disposed') is one which yields a kind of regular connection between utterance and other action, and so the kind of regularity which makes interpretation possible. (The expression of wishes or expectations, for example, might be part of the behaviour which enabled an interpreter observing someone speaking in monologue to work out what he meant, as in §243.) That I spontaneously express my wishes, and that 'some event stops my wishing does not mean that it fulfils it' are aspects of the same 'regular connection': one according to which a wish is satisfied just by an event described by the same sentence as that in which the wish is expressed, and therefore just by an event in normative accord with the wish itself. Hence the interpretable regularities which fix the semantic character of sentences also serve to fix those of the states of mind, such as wishing, which we use sentences to articulate and express.

This holds for all the propositional attitudes which have this kind of first-person ascriptive expression; so all can be regarded as bound to their objects in the same linguistically mediated way. Thus Wittgenstein puts the point in a way which admits of generalization:

444. One may have the feeling that in the sentence "I expect he is coming" one is using the words "he is coming" in a different sense from the one they have in the assertion "he is coming." But if it were so how could I say that my expectation had been fulfilled? If I wanted to explain the words "he" and "is coming", say by means of ostensive definitions, the same definitions of these words would go for both sentences. But it might now be asked: what's it like for him to come. -- The door opens, someone walks in, and so on. --What's it like for me to expect him to come? -- I walk up and down the room, look at the clock now and then, and so on. -- But the one set of events has not the smallest similarity to the other! So how can one use the same words in describing them? -- But perhaps I say as I walk up and down: "I expect he'll come in" -- Now there is a similarity somewhere. But of what kind?!

445. It is in language that an expectation and its fulfilment make contact.

28. Thought, practice, and the fixing of reference.

This should, I think, be seen as a claim about the representational or semantic character of states of mind which have a sentential (*that P*) articulation generally. The referents of such states of mind, and the conditions in which they are satisfied, fulfilled, etc., are those of the sentences by which we articulate them, and hence are determined by the practice of using those sentences. As noted above, this reverses the Tractatus conception of thought as independent of language and explanatory prior to it. So this remark is

continuous with *Philosophical Grammar* §63-4, in which Wittgenstein put forward a linguistic explication of the reference of thoughts while rejecting mechanistic ones.

The claim in *Philosophical Grammar* was that it was not via an internal symbol system, but rather in linguistic practice, that a thought about Napoleon made contact with Napoleon himself. And like the argument in §441-5, this turned on the way the same words are used both in the expression of a thought and in the practice of describing and referring to its objects. (Cf. '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.') My uses of "Napoleon" in expressing my thoughts link these thoughts to the referent of that term, as I use it in other contexts; and my uses also enjoy interpretable connections with those of others, which ultimately reach back to Napoleon himself. So what enables me to speak of Napoleon also enables me to think of him. The interpretable order of which my linguistic expressions of thought are a part actually encompasses the object of my thought, despite its remoteness in time and space from my activity of thinking; and this enables us to see how thought and object are related.

We can see how this is carried into the fuller argument of the *Investigations*. In §437 Wittgenstein had posed a question: 'A wish seems already to know what will or would satisfy it; a proposition, a thought, what makes it true...Whence this determining of what is not yet there?' §206-7 have enabled us to see that we fix the conditions in which sentences are true via the regular connections manifest in interpretable practice, and §441-4 that these connections include sentential expressions of motives (expectations, wishes, etc.) which specify the conditions in which these motives are fulfilled, satisfied, etc., and hence the things which they are about. So the same regularities as secure the references of sentences also and thereby secure those of thoughts which sentences express. §445 thus answers the question of §437, by saying that the 'determining of what is not yet there', by which 'an expectation and its fulfilment make contact' is effected by the use of language.

29. Practice and consciousness.

The idea that the intentionality of thought is to be explained via linguistic practice also contradicts the idea that mental reference is simply given to us in consciousness itself. Thus in §205 -- just before indicating the role of interpretable regularity in §206-7 -- Wittgenstein has his interlocutor urge that the existence of a custom or technique (interpretable practice) is not necessary for an intention to play chess. The idea is apparently that one can simply *intend* or *mean* to play chess, as one can simply *intend* to designate or *mean* Napoleon: the relation of intention or thought to its objects is prior to that of linguistic practice.

Wittgenstein indicates the inadequacy of this by asking how the rules which define chess are 'present in the mind' of someone intending to play. This is a reminder of the point, stressed also in §186, that what someone intends, means, expects, etc., may go beyond anything which could plausibly be said to be a content of that person's conscious mind at the time. (This is clear also in the case of such an intention as that to add 2, or again my expectation that I will do so correctly in any instance if asked; and something similar holds for the variety of instances of some descriptive word, for these too are characteristically not 'present to the mind' in use.) Such consciousness-transcending mental reference, Wittgenstein implies, requires to be explained by reference to linguistic practice. Thus as Wittgenstein answers his own question about the reference of intention in §197: 'Where is the connection effected between the sense of the expression "Let's play a game of chess" and all the rules of the game? -- Well, in the list of the rules of the game, in the teaching of it, in the day-to-day practice of playing.'

Here Wittgenstein again returns to the idea of the sense of a sentence, that is, the thought expressed by it. So this again is a remark about how such a sense or thought -- that expressed by "Let's play a game of chess" -- relates to the things it is about. The answer again is in terms of use or practice. The verbal expression of a desire or intention to play chess is part of an interpretable practice; and other parts of that same practice include the spoken and written expressions of the rules of chess, the many co-ordinate actions which constitute the teaching and learning of chess, the playing of actual games, and so forth. In understanding such a practice we relate speakers to things in their immediate environment, and also to things beyond it [(ii) - (iv) in section 16 above]. For the parts of such a practice admit any degree of spatial and temporal distance from one another, as the case of adding 2 already illustrates. Interpretation links the disparate parts of such a practice as elements of interpretable regularities, and so binds them together in an intelligible (and causally connected) whole. What *cannot* be present to consciousness *can* be a part of such a practice. So intention and other sententially described states -- including thoughts -- refer beyond consciousness, via the practices in terms of which they are expressed and described. Finally, as Wittgenstein argues in §243ff, what holds for mental reference which goes beyond consciousness in this way also holds for mental reference within consciousness itself [(i) in section 16 above]. Just interpretable practice seems unnecessary for us to mean chess by 'chess', so it seems unnecessary for us to mean pain by 'pain'. It seems that we can simply associate words with sensation, and could do so even if there were nothing which could enable another person to judge what sensations we had, and so to interpret these words (§256). But it is a corollary of the arguments above that someone could not lay down the meaning of a word in this purely first-person way.

This establishing of meaning is supposed to bring it about that the speaker can use the word correctly, that is, as in accord with a rule. We have seen, however, that what makes it the case that someone follows such a rule correctly -- as opposed to merely seeming to herself to do so -- is that her performance can be assessed as part of an interpretable practice. So if this supposed inner ostensive definition really were beyond the scope of interpretation, it would also be beyond the scope of correctness. In the case of such an imagined, purely first-person conception of sensation, there would be no standard of correctness (§258). It is natural to think that even so one would still *believe* that one was associating word and sensation correctly (§260). But this is to presume that in the imaginary situation one *thinks* or forms *beliefs* about the sensation correctly, and this is equally unwarranted. For thought and belief are also not purely first-person matters, but subject to external assessment. So the intentionality of consciousness -- consciousness as permeated by sententially articulable thought -- requires to be understood in terms of language and practice as well.

30. Summary and conclusion.

As presented here, Wittgenstein's argument has both a conceptual and an empirical aspect. The conceptual claim is that we should regard sentential content -- and hence the assignment of both propositional attitudes to persons and propositions to their sentences -- as having a particular place in our 'form of life', that is, as constituted via our activities of making sense of human behaviour. We can cast light on this, in turn, by considering the activity of a radical interpreter, who is set the task of understanding the speech and actions of people unfamiliar to him. This enables us to see that our natural but disciplined practice of ascribing sentential content requires the discernment of regularities which hold between linguistic and non-linguistic behaviour (practice). Given these regularities, we can characterize both utterances and non-verbal activities as the product of motives which have the same kind of sententially ascribed contents as the utterances themselves; and this is the kind of content which we ascribe to our own thoughts. In the

absence of such regularities, however, this kind of content could not be ascribed in an empirically disciplined way; and so far as we understand such content as constituted via the explanatory practice in which it is assigned, this is to say that such content would not exist.

This conception, we should note, is consistent both with the claim that thinking and intending are (or are realized by) events in the brain, and that such events are causes of -- and in this sense explanatorily prior to -- behaviour and practice. Ascribing a causal priority to events in the brain should be no barrier to recognizing that our ability to conceive these events in the particular sentential ways we do depends upon our capacity to relate them to behaviour and the environment, and via the discernment of regularity, in the way Wittgenstein indicates. On Wittgenstein's account the assignment of content is like a kind of measurement, which presupposes that the objects to be measured (the forms of verbal and non-verbal behaviour to be assigned motive- and content-specifying sentences) can be related in certain ways. In particular, the sentential measurement of intentionality can be compared with that of weight (§142). Just as the assignment of weights to objects by balancing them against standards would lose its point if these balancings did not correlate with the weight-related behaviour of objects generally, so assigning contents (and hence motives) to behaviour via sentences would lose its point if the utterance of sentences failed to correlate with motivated behaviour generally. Without the order required for the imposition of a metric, assignments of measurement lapse; and this applies to the assignment of sentential content to states or events in the brain.

What Wittgenstein offers as against the *Tractatus* (Augustine/Frege/Fodor) idea of thought is thus a series of proposals about the nature of sentential content itself. The order which sustains the ascription of such content is manifest in linguistic practice, which is why such practice (use) constitutes the life of signs. Since assignments of content to thought presuppose this order, we cannot take the content of thought to be independent of that of language. So the idea that our thoughts have determinate and meaning-constituting content prior to language can be seen as another form of the Cartesian illusion characterized above.

Attending to the contents of our thoughts as these seem immediately given to us in consciousness, we fail to apprehend the full range of conditions involved in their attribution. We take as simply given (or given by some sort of invisible underlying mechanism) what is in fact constituted via our understanding of a behavioural order in which an essentially salient role is occupied by language and practice.

We have seen that in arriving at this conception Wittgenstein put aside the idea of an inner code. As re-enacted in the *Investigations*, this rejection was prompted not only by the realization that such a code could not be semantically self-sufficient, but also by empirical ignorance. (§158...'Now ask yourself: what do you really *know* about these things?') So it is worth observing that an advocate of such a code could accept, in the spirit of the *Investigations*, that the code could be assigned sentential content only via the kind of order shown in linguistic practice, while also holding, in the spirit of Augustine or Fodor, that it nonetheless has such content before language is acquired. The idea would be that the elements of the code might already (innately) be such as to have a determinate explanatory relation to behaviour, including future speech, even before such behaviour actually develops. The structure of the code -- to take one of Wittgenstein's own examples -- might be compared to that inside a seed, which is already such as to specify the nature of the plant (linguistic and non-linguistic behaviour) which will grow from it, and so could be assigned content on the basis of this order-determining power. Wittgenstein was inclined to reject such a picture, and indicates an alternative in which thought acquires sentential focus only together with the behaviour to which it relates. Such a view, which could not be justified by philosophical argument alone, seems also to be gaining support in cognitive science.