## Wittgenstein, Davidson, and Radical Interpretation

In this paper I try to relate a number Wittgenstein's remarks in *Philosophical Investigations*, including those on following a rule, to work by Quine and Davidson, particularly the latter. The comparison turns on the notion of radical interpretation, and bears on a number of topics, including meaning, the content of psychological states, first-person authority, the cogency of interpretation, and the problem of consciousness. Since space is limited, and Quine and Davidson have written so as to be readily understood, I shall at first presuppose acquaintance with their views, and concentrate on spelling out those of Wittgenstein and describing points of connection. In later sections, however, it will be possible to relate Wittgenstein and Davidson in more detail, and also to survey matters in a less exegetical way.

I

Let us begin by considering *Philosophical Investigations* §§205-207, in which Wittgenstein explicitly introduces the notion of interpretation, and indicates its relevance to questions raised in his other remarks on mind and language.

> 205. "But it is just the remarkable thing about *intention*, about the mental process, that the existence of a custom, or technique, is not necessary to it. That, for example, it is imaginable that two people should play games in a world in which otherwise no games existed; and even that they should begin a game of chess -- and then be interrupted."
>
> But isn't chess defined by its rules? And how are the rules present in the mind of the person who is intending to play chess?
>
> 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'.[1]

The significance of these remarks partly depends on their place in Wittgenstein's argument. §205 is part of his discussion of following a rule, and also of the relation between consciousness and the intentionality, or object-directedness, of mental states. As Wittgenstein has observed, when we intend to play a game or follow a rule, we may not be consciously aware of the rules which define the game, or the series of steps we have to take in following a rule. So we have the question asked in §205, as to how, or in what sense, such things can be 'present in the mind' of a person who intends.

The same topics are continued in §206, which refers back to Wittgenstein's earlier and quite fundamental questions as to how the understanding of a rule, or thought or action in accord with a rule, is possible at all. The issues raised in those remarks concern both what constitutes the correctness of an activitiy of following a rule, and how we can know when such correctness obtains (cf the 'How can a rule show me what I have to do at *this* point...' of §198). In §206 Wittgenstein presents a consideration about interpretation -- 'the common behaviour of mankind is the system of reference by which we interpret an unknown language' -- which he evidently takes as relevant to answering these questions. In §207 he takes this idea a step further. Here he claims that if we are to understand people's language, there must be a 'regular connection' between 'the usual human activities' -- which are presumably part of 'the common behaviour of mankind' referred to in §206 -- and the utterances (makings of sounds) of people whose language we hope to understand.

This last part of the argument of §206 and §207 seems relatively clear. Everyday human actions provide a system of reference by which we interpret a language, in the sense that interpreting a language requires us to *correlate* and thereby to *co-ordinate* utterances of that language with such actions. Thus to take the most straightforeward case, to interpret a language as containing orders we must be able to correlate utterances which we take as orders with actions related to those orders -- with behaviour which we can understand as acts of compliance, defiance, and so forth. Such correlations, which show episodes in verbal and non-verbal behaviour to be related both to one another and to the context in which they occur, partly constitute the circumstances in which we would 'say that the people there gave orders, understood them, obeyed them, rebelled against them, and so on', to which Wittgenstein alludes in §206.

It seems that if we consider §206-7 in their fuller context we can see Wittgenstein as making a series of related points, which can partly be brought out as follows. Speech seems a kind of action which we understand particularly clearly, and by means of which we can obtain understanding of an agent's further actions and motives which is clear and precise, and which draws upon that agent's first-person authority. But as §207 suggests, speech is also a kind of behaviour which cannot be understood in isolation from the the rest of the behavioural order of which it is a part. If people's productions of sounds or marks 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. So if, e.g., we were presented with the highly patterned bursts of sound constituting speech, but really lacked any further information as to how the production of these bursts was interwoven with the situation and other actions of the person making the sounds, then we would not be able to make sense of them. (One can get some idea of this by trying to imagine learning a foreign language simply by listening to the radio, but without having any independent idea of the events with which the broadcasts were concerned.) The sounds of speech, however clear or clearly structured, are meaningless until systematically related to worldly objects and events; but grasp of their intrinsic nature or structure alone would not enable us to relate them to things in a precise and empirically disciplined way.

By contrast, as §207 also suggests, we can grasp 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' -- as we can in the case of one another. 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 is doubtful how far we can ascribe precise thoughts or motives to people at all (Cf. §25, §32; and also §342).

We thus 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 actions 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 structure of utterance to particular points in the framework of action and context, and thereby 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.

We can perhaps make Wittgenstein's claim in these remarks clearer by comparison with something which is more familiar and immediately acceptable. In our everyday practice of interpreting utterances we do not simply assign meanings to them; rather we characteristially take them as expressions of desire, belief, intention, and other motives. (Thus we take regularly take assertions as expressing beliefs, questions as expressing desires to know something, requests or orders as expressing desires that something be done, and so forth.) This enables us to intepret the motives which we take to be expressed in this way with precision, and also to relate such interpretation to the speaker's ability to express such motives with authority. Clearly, however, we could not take utterances as such expressions of motives with any degree of accuracy and certainty, unless we also had independent means of determining what the agent's operative desires, beliefs, or intentions really were. Evidently the means we use are the interpretation of further actions. We are able to regard utterances as accurate or authoritative expressions of motives because doing so enables us to interpret other actions, and with cogency, as stemming from those same motives, or others closely related to them. In understanding persons in this way, therefore, we in effect correlate their utterances with other actions, as effects of common causes (motives). Schematically, insofar as we take an utterance of 'S' as an expression of a desire, intention, or belief that P, and then confirm this by independently interpreting further actions as flowing from that same motive, we thereby correlate utterance and action. Such correlations, as we shall see, are at least closely akin to those Wittgenstein introduces in the remarks above.

Wittgenstein's emphasis thus falls upon the correlative interpretation of signs and actions. This kind of correlation, it is worth noting, figures from the beginning of the *Investigations*. In his opening remark Wittgenstein quotes Augustine on the way he came to understand (interpret) human language, and contrasts Augustine's account with an example of his own:

> Now think of the following use of language: I send someone shopping. I give him a slip marked 'five red apples'. He takes the slip to the shopkeeper, who opens the drawer marked 'apples'; then he looks up the word 'red' in a table and finds a colour sample opposite it; then he says the series of cardinal numbers -- I assume he knows them by heart -- up to the word 'five' and for each number he takes an apple of the same colour as the sample out of the drawer. ---- It is in this and similar ways tht one operates with words.----'But how does he know where and how he is to look up the word 'red' and what he is to do with the word "five"?' ---- Well, I assume he *acts* as I have described. Explanations come to an end somewhere...

In his first example Wittgenstein concentrates upon connections between signs and actions, and urges that these provide a point of termination in the explanation of meaning. Since these connections partly consititue the practice of using signs, the signs can be understood by reference to them. The very simple languages described in such early remarks as §2, §8, §19 and §21 provide further clear examples: they consist of regular connections between utterance and action, which enable us to interpret the former by reference to the latter. In §2, for example, utterances of 'slab' and 'block' are regularly followed by actions which are appropriate to them; and such understanding as we have of this and such other simple examples rests on our interpretation of such regularities.

Thus from the beginning of the *Investigations* Wittgenstein stresses the role of instances of regular connections between signs and actions, which he apparently regards as fundamental to interpretation and meaning. Of course actual human languages are far more complex than Wittgenstein's early and deliberately simplified examples; and they possess the feature Wittgenstein so emphasized in the *Tractatus*, that their sentences can be understood from the words composing them and the way these are combined. So in treating rules and orders in 'an articulate language' in §206-7, Wittgenstein is turning to the actual case, and indicating that comparable regularities must hold here as well.[2] (The germ of this

development is indicated in §1, in the way the shopkeeper generates his use of 'five red apples' from distinct practices related to 'five', 'red', and 'apples' respectively.)

Remarks §206-7 also refer ahead, for example to those with which Wittgenstein begins his discussion of 'private' language.

> 240. Disputes do not break out (among mathematicians, say) over the question whether a rule has been obeyed or not. People don't come to blows over it, for example. That is part of the framework on which the working of our language is based (for example, in giving descriptions).
>
> 241. 'So you are saying that human agreement decides what is true and what is false?' It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life.
>
> 242. If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments. This seems to abolish logic, but does not do so.-- It is one thing to describe methods of measurement, and another to obtain and state results of measurement. But what we call 'measuring' is partly determined by a certain constancy in the results of measurement.
>
> 243. 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...[?]

It seems that Wittgenstein's claims about rules, orders, and interpretation in §206-7 serve as grounds for conclusions which he explicitly draws and amplifies in §§240-243. The emphasis is again on regular connections between sign and action -- e.g., as in §240, between signs taken as rules and actions understood as in accord with these rules. Now, however, Wittgenstein discusses such regularity in relation to communication as well as interpretation. So in §240-2 he considers sign-action regularities among speakers who use language to communicate with one another; and then in §243 he considers the role of similar regularities in understanding persons who do not communicate, but rather use language to express thoughts to themselves.

In §206-7 Wittgenstein used the figure of the explorer/interpreter to illustrate his claim that the finding of regular connections between utterances and other actions is necessary for interpretation. Since communication requires interpretation, this would imply that the finding of such regularity is basic to communication as well; and this is made explicit in §240-2. In §243 Wittgenstein returns to consideration of the explorer/interpreter, to add that by reference to the same regular connections such an interpreter might also succeed in understanding and translating the utterances of people who speak only in monologue (do not communicate). The point of this is evidently to make clear that he holds that sign-action regularity is not only required for interpretation, but also capable of sustaining interpretation in the absence of communication. Agreement is required for communication, but communication is not required either for the use or the interpretation of language.

The sign-action regularities upon which Wittgenstein concentrates in §1 explicitly involve the practices or techniques of using signs, which, as we see, he takes to be remarkably varied. Those in §206-

7 and §240-2 concern signs which serve as specifications of what an agent is supposed to do -- e.g to comply with a rule, or to obey an order -- and actions which accord with these. These kinds of examples are related, for the actions which a person performs in following a rule or order can be taken as parts or aspects of the practice of using the sign by which the rule or order is given. In §243 Wittgenstein cites a further kind of instance, which he evidently takes to be related to those in §240-2: the regularity which enables the interpreter of the monologue people to predict their actions on the basis of their resolutions and decisions.[3] Resolutions and decisions are expressions of intention; so this suggests that correlations which hold as between expressions of intention, and actions in accord with these intentions, can be taken as a further example of the sign-action regularity Wittgenstein has in mind. This kind of correlation, moreover, clearly overlaps with that involved such other examples as practices, rules, and orders; for in general the linguistic expression of an order or rule of practice can also be regarded as a specification of the intention with which the person who follows the order or rule thereby acts.

This link with intention seems also to be foreshadowed in §1: for the shopkeeper who uses sign-action correlations for 'five', 'red', and 'apples' to generate a series of actions which accords with 'five red apples' thereby also fulfils an intention to select five red apples. The role of intention also gives a clear sense to the idea that an utterance and action which are correlated in this way are thereby assigned co-ordinate locations in a frame of reference: the correlation shows the utterance and action to be related to one another via their relation to the same intention, that is, to the same state of mind; and this also establishes the possibility that this utterance and action might also be related to others, via the relation of the intention which links them to further intentions. So for the moment we will concentrate on the role of intention in this kind of regularity, and relate this more fully to rules, orders, etc., below.

An utterance or other production of a meaningful sign is an episode in behaviour, as is an intentional action correlated with the production of the sign. So in order to find instances of the kind of regular connection with which Wittgenstein is concerned in §207 we (or the interpreter/explorer) must hypothesize interpretations: we must tentatively interpret, on the one hand, episodes in behaviour which we take to be utterances of sentences, and on the other hand, episodes in behaviour which we take to be actions related to the hypothesized utterances. We must interpret a pair of episodes in order to see them as co-ordinated by intention; and success in the interpretive co-ordination of such episodes is a requirement of interpretation itself. So we are concerned with what might be called *interpretive regularity* -- regularity which is grasped and understood through interpretation, and which, therefore, emerges as fully specified only in the activity of interpretation which such regularity also sustains.[4]

From Wittgenstein's claim about such regularity in §206, it evidently follows, as he concludes in §242, that if individuals are to succeed in understanding one another in this way -- that is, if they are to use language as a means of communication -- they must be able to agree in empirical judgments. For as these passages imply, communicators must agree in their interpretation of one another's utterances. (If A succeeds in communicating with B by an utterance of a sentence 'S', then B interprets that utterance as A intends, so that A and B understand that utterance of 'S' in the same way.) So if this requires a grasp of interpretive regularities, communicators must also be able to be in accord about these.

Wittgenstein thus holds that communicators must be able to agree in some such judgments as, that an utterance of a word or sentence is connected with a certain type of action, say by a rule or order; and hence they must be able to agree as to whether an action of that type has been performed. Wittgenstein puts the point in §458 by saying that 'If an order runs "Do such-and-such" then executing the order is called "doing such-and-such"'; so, clearly, agreement in understanding an order or rule will encompass agreement as to whether actions of the required kind have been performed. Again, likewise, comunicators must be able to agree that an utterance of a sentence specifies or expresses a particular intention, and that a certain action fulfils that intention; and hence also they must agree as to whether the world is as it would have to be, for that desire or intention to be fulfilled. (They must be able to agree, e.g., for appropriate values, that an uttered sentence 'S' specifies the rule or intention that P; that an action accords with the rule or intention that P; and so that it is the case that P.)[5]

In §240 Wittgenstein concentrates on agreement as to whether a rule has been followed, e.g. agreement as to whether an action is in accord with the rule 'Add 2'. This of course also constitutes agreement as to whether an action is an instance of (intentionally) adding 2, and hence whether it accords with an intention specified by 'Add 2'. These are judgments about interpretive regularities which, as Wittgenstein says in §242, those who use language to communicate must both make and agree in making; and he perhaps adds, more generally, that the use of a sentence 'S' for purposes of communication requires agreement among those who use it as to the circumstances in which 'S' is to be held true, that is, agreement as to the circumstances in which it would be correct to judge that S. These claims are plausible, and they indicate how the kind of regularity upon which Wittgenstein is concentrating is registered in commonsense interpretive judgments -- and in particular judgments about intention -- which link mind, language, and the world.

As the above remark about orders suggests, connections of this kind can also be described as *sentential*. For each instance of such utterance-action regularity not only concerns a putative sentence of an interpretee's language, but also is marked by a repeated use of a sentence of the *interpreter's* language, to describe alike the utterance, intention, and action of the interpretee, and hence also the world as the interpretee conceives it. This flows from our practice of describing the mind via the embedding of sentences which we also use for describing the world. Thus, as noted above, we describe mental states like intention, desire, belief, hope, fear, and so forth, by using sentences which specify the worldly objects or states of affairs upon which these motives are directed. Schematically, we may use a sentence 'P' to speak of the intention to (or that) P, the desire to (or that) P, the hope or fear that P, and so on; and this means that the worldly circumstances specified by the sentence 'P' -- that is, the circumstances in which 'P' is true -- are also those in which the intention is fulfilled, the desire satisfied, the hope or fear realized, etc.

This practice systematically implements our conception of the mind as possessing intentionality: that is, as directed on, or engaged with, the world. In consequence, Wittgenstein paid particular attention to it throughout his philosophical career. Thus in *Philosophical Remarks* he says

> I only use the terms the expectation, thought, wish, etc. that *p* will be the case, for processes having the multiplicity that finds expression in *p*, and thus only if they are articulated. But in that case they are what I call the interpretation of signs.[6] 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 expectation). [7]

The notion of interpretive regularity which Wittgenstein introduces in conjunction with that of an 'articulate' language in §207 is itself correlative with this practice of articulating thoughts by sentences. For an interpretive regularity is one in which, in accord with this practice, an interpreter can use a single sentence to describe both (the content of) a sentence of an interpretee's language, taken as specifying a rule, or again a desire or intention, and also a connected action of the interpretee, taken as in accord with that rule, desire or intention. So this same sentence will also articulate the interpretee's desire or intention in acting, by describing the circumstances in which this desire or intention would be fulfilled.

This means, among other things, that interpretive hypotheses covering the kinds of regularities with which Wittgenstein is concerned will have a particular methodological feature: they will characteristically be framed and tested by what we can regard as successive uses of the same sentences. Thus, for example, if an interpretee utters (sounds best construed as an instance of the sentence) 'I will tie my shoelaces', an interpreter may take this as specifying the interpretee's desire or intention that she tie her shoelaces, which the interpreter describes by the sentence 'she ties [+ tense] her shoelaces'. The kind of regularity Wittgenstein is discussing will be instantiated if this (behaviour best interpreted as) expression of desire or intention is accompanied by (behaviour best interpreted as) the interpretee's

intentionally tying her shoelaces. This action will again be described by the interpreter's sentence 'she ties [+ tense] her shoelaces', as will the worldly circumstances in which the interpretee's desire or intention is fulfilled. Finally, this repeated use of the interpreter's sentence should in turn parallel a potential for a comparable use of the interpretee's, as might be represented in terms of '[subj] ties [+ tense] [subj's] shoelaces, to describe her own intention, action, and circumstances.

The existence of this kind of 'regular connection' between utterances and actions is thus shown by regular success in the sort of interpretation -- the specification of articulate desires or intentions -- which the regularity makes possible. Regularity in utterance-action connection is evidenced by regular success, on the part of an interpreter, in framing judgments about desire and intention which connect utterance and action in the requisite way. This in turn requires the interpreter's regular success in finding world-describing sentences of his or her own language (used in interpretive hypotheses) which serve also to describe the intentions, utterances, and further actions of the interpretee. The interpreter must constantly succeed in finding sentences which as it were form a bridge, marked by repeated use of the same terms, connecting mind, language, and the world, and in a perspective common to both interpreter and interpretee. This is also the kind of success we take ourselves to have in understanding others with whom we share a common language. And since the common sound or shape of our sentences does not by itself guarantee that they bear common meanings, this sharing also appears to rest on an intuitive practice of interpretation such as that we are discussing.[8]

Finally, we should note that it is not necessary for the holding of interpretive regularity of the kind we are considering that the utterances of sentences which serve to specify the desires, intentions, or actions of interpretees should precede the actions with which these utterances are correlated. If Wittgenstein's monologue people were prone to comment on their own actions after performing them, the explorer could predict what they would say on the basis of what they had done, as well as vice-versa. Also the fuller an interpretee's commentary on action, the better interpretive evidence it provides, and actual comments would characteristically provide grounds for hypothesis about further comments which would be forthcoming if circumstances were different. Hence the role of such regularity is clearly counterfactual -- it extends beyond sentences which interpretees actually utter or hear, to others which they would use or respond to, for example if describing their actions more fully, if asked about what they want or intend, and so on. So in general, insofar as we interpret actions as in accord with sentences which we hold that interpretees themselves would offer or accept as specifying their motives, we thereby take our interpretations as answerable to sentential regularities of the kind Wittgenstein indicates. Seen in this way, the role of such regularity in interpretation appears both pervasive and fundamental.

II

Now clearly the use Wittgenstein makes of an explorer seeking to interpret an unknown language is comparable to that which Quine, and, following him, Davidson, have also made of a such a figure, described as a 'radical' translator or interpreter. (Wittgenstein emphasizes interpretation in §206 and translation in §242). Wittgenstein, that is, can be said here to be using the conception of a radical interpreter, to cast light on meaning, and in particular on both rule-following (§206) and intentionality (§205). So, for example, we are to supposed to accept that language need not involve communication by considering what such an interpreter might find (§243), and this also is supposed to resolve questions as to the right way to follow a rule (§207). And it is in this same perspective that we are to accept that language essentially involves an order to be found in human behaviour, which constitutes 'the system of reference by which we interpret an unknown language', and which, it would seem from §241, is also that of a 'form of life.' Overall Wittgenstein's idea in these remarks seems comparable to that put by Davidson in another context: 'The key to understanding all these mental phenomena is to see them from the point of view of an interpreter'[9].

Radical interpretation thus constitutes a link between Wittgenstein's account of mind and language and those developed later, but independently and more explicitly, by Quine and Davidson. A similar use of

the notion seems common to all three. In everyday life we discern motive and meaning in human behaviour spontaneously and without reflection: we simply see intention in the ordered series of movements constituting an agent's behaviour, hear meaning in a sequence of sounds which a speaker produces, and so on. Hence if we want to get a clearer view of understanding of this kind -- if, for example, we want to consider whether the behaviour towards which such understanding is directed shows an underlying order which makes this understanding possible -- we need to represent both the behaviour and our mode of understanding in a way which enables us to take account of them in a way which is more explicit and detailed.

One way to do this, as Wittgenstein suggests in his early remarks, is to imagine linguistic activities which are simplified, so that important features stand out clearly. Another, which emerges at §206-7, is to imagine that we are seeking to understand persons in a society with which we are entirely unfamiliar. In this case, unlike the simpler ones, we are required to take into account a range of behaviour which is clearly full enough to support the kind of ascriptions with which we are concerned, and also to contemplate seeking to understand this behaviour with a minimum of preconception. We are thus led to consider both the nature of the behaviour, and the steps we would have to take in order to understand it, in a more deliberate and reflective way.

Thus in imagining interpretation of this kind we naturally present ourselves with a distinction. We imagine the utterances of our interpretees as strange to us, but other of their actions -- and in particular those that are characteristically human -- as more or less recognisable. Hence, as Wittgenstein assumes, we are better able to see that we could make sense of their sounds only in the context of, and by connection with, their non-verbal behaviour and action. This, however, indicates a conceptual point about interpretation, which also bears upon our own unreflective practice. We see that we too must understand speech as co-ordinated with non-verbal activity; and hence we have reason to suppose that in finding the speech of others intelligible, we also regularly find (or somehow register) connections between their speech and other actions, whether we are aware of this or not.[10]

We take another step of the same kind, it seems, in following Quine, and considering interpretation in terms of explicit hypothesis; and yet further steps, if we follow Davidson and try to give a fuller and more explicit characterization of the relevant theoretical structure by reference to the step-by-step use of a theory which encompasses both meaning and decision. Such a strategy offers to provide a relatively detailed account of a way of thinking which would suffice for interpretive judgments; and this in turn should enable us to see more clearly what interpretive understanding is, and what in our behaviour enables us to be understood in this way.

The idea in this, we should note, is *not* that we actually come to understand one another by the steps or hypotheses which such radical interpretation describes; nor again that these are steps by which either babies or field linguists actually proceed, should ideally proceed, or so forth. Davidson plainly represents his interpreter both as constrained in ways actual people are not, and as thinking in ways in which actual people do not, and these features are intrinsic to his project.[11] His idea is apparently to cast light on what we actually do by describing an alternative way of doing the same thing, or something relevantly similar -- where we can describe the alternative in a more informative way than we can our own practices.

The imagined interpreter shares with us the basic grounds and conclusions of interpretation, so that she, like us, is set the task of passing in an empirically disciplined manner from observable behaviour to assignments of motive and meaning which serve to explain this behaviour by rendering it, as Wittgenstein says, logical and intelligible. So insofar as we can give a non-question-begging account of the way the interpreter effects this transition, we attain a further perspective on the task we actually accomplish, and upon features of the evidence which we use which render such an accomplishment possible. And the more the imaginary interpreter is constrained in respect of the evidence she can use,

the more the account enables us to see about what information can be extracted from that evidence, and in what ways.

This mode of explication is not conceptual analysis; so it may be worth trying to spell it out more fully. One way to do so (which is not Davidson's) is in terms of the philosophical tradition stemming from Kant, according to which we can understand concepts in terms of their role in judgment.[12] According to this kind of account concepts serve to enable us to make certain transitions in thought, which terminate in beliefs or judgments; so they can be explicated as *rules* or *functions* which yield these judgments. A concept used in everyday perceptual judgment, for example, may enable us to make a transition from a perceptual encounter with a certain kind of object, to a judgment that that object is of the appropriate kind. Thus someone who possess the concept of a horse is able (among other things) to judge of a perceived object that it is a horse. The role of the concept here includes that of taking perceptions stemming from an object (or perhaps the object itself) as input, and yielding the judgment that the object falls under the concept as output. (A similar account of concepts seems to have had a deep influence on Wittgenstein, probably via a formulation by Frege; for at 4.0141 of the *Tractatus* he hints that many of our symbolic processes -- encompassing thought, language, and action -- are to be understood in terms of input-output rules of this kind.)

In this perspective to gain understanding of a concept is to gain understanding of that function of judgment which, in our practice of forming judgments, the concept enables us to compute. So one way of casting light on a concept is to provide a further account of this function, e.g. by providing some alternative or fuller means for determining it. We can see traditional attempts to give definitions of concepts, or to provide necessary and sufficient conditions for their application, as of this kind. And we can see that such an account would tend to offer better and fuller explication of a concept the more closely the account approximated to an genuine algorithm for determining the function in question. A step-by-step procedure for effecting a conceptual transformation, so far as we can attain it, should enable us to gain a perspective on that transformation which is particularly clear, explicit, and practical.

Such attempts at explication have commonly proved to be instructive failures. The Kantian analogy enables us to see why failures of this kind should be instructive; for even an incomplete part of an alternative means of computation may give us a better view of a function we want to understand. Also we have become familiar with one particular source of difficulty, which is often described in terms of holism. As Kant stressed, we employ each empirical concept together with many others, so that our judgments result from the operation of an integrated system, which works upon the data of our perceptual experience more or less as a whole. In consequence, it seems, we cannot hope to provide explication of concepts by describing their roles in mediating input and output one by one. Rather we must somehow treat each concept as part of a larger system, and without circularity; and in many cases this has not proven easy to do.

Now it appears that the concept of meaning might be explicable along the lines considered previously; for our notions of meaning and motive seem to have a clearly discernable role. They can be seen as the concepts which enable us to impose upon behaviour and utterance the interpretive judgments in terms of which we understand behaviour and utterance as meaningful, that is, the judgments encompassed in our seeing the behaviour as intentional, hearing the utterances as particular acts of speech, and so on.[13] This, however, faces us with the problems of holism and circularity just mentioned. Thus take our notion of a *thought* or *proposition* : in assigning meaning to sentences, we relate sentences to thoughts or propositions; and in assigning motive to behaviour, we relate that behaviour to propositional attitudes like intention, belief or desire, and hence, again, to propositions. As Quine emphasized, so far as our concepts of meaning and motive are those of proposition and propositional attitude, they seem linked so inextricably as to prevent significant explanation of one in terms of the other.

We can see Quine and Davidson as seeking to meet this difficulty by a change in strategy which is apparently both required and minimal. We apply commonsense interpretive concepts to behaviour, to yield judgments which we can regard as explaining that behaviour; and we apply concepts in a similar way when we make use of explicitly formulated theories to explain empirical data. So it seems that we can replace the goal of explicating a single commonsense interpretive concept, or perhaps an interrelated family of these, with that of setting out a theory which would yield relevantly similar explanatory judgments on the basis of relevantly similar data. This will enable us to meet the problem of holism by employing a number of related concepts together in a theory; and then we can keep track of the distinct work done by these concepts via the assignment of explanatory roles which the theory will serve to make explicit. And we can still try to envisage the application of this explicatory theory in a way which is as nearly algorithmic as possible, that is, in terms of a series of clearly defined steps.

The aim thus becomes that of specifying as fully as possible a theory and a procedure for applying it which would enable an interpreter to proceed step by step from exposure to the behaviour which grounds interpretive judgments about meaning and motive to these judgments themselves, that is, to judgments which assign propositions, or propositional content, to utterance and motive together. This should provide the basis of a philosophical account of meaning and motive -- and perhaps further concepts which are related to these, and which figure in the theoretical procedure -- which is as close to traditional attempts at conceptual explication as the material allows, but which is also tailored to meet the conceptual interdependence upon which so many other attempts have foundered. This program clearly does not require that the 'theoretical algorithm' by which we seek to explicate our notions of meaning and motive should overlap with procedures employed in our actual interpretive practice, or again with actual input-output relations in our heads. But if -- as some linguists and philosophers argue -- there is further reason to suppose that this is so, then the theoretical work in question may play a role in explication which is fuller still.[14]

The notion of radical interpretation thus provides an interesting example of continuity, convergence and development in analytical philosophy. Since Quine and Davidson have articulated this conception in ways Wittgenstein did not, we can usefully see aspects of his work in light of theirs. For example, it is often noted that the later Wittgenstein equated meaning with use; and also that this, by itself, can seem a somewhat uninformative connection. In light of the above, however, we can see that Wittgenstein's conception is deeper and more rigorous, and in a way which is now familiar. The ascription of meaning is indeed fixed by linguistic use, together with other behaviour. But the link between use and meaning is made via interpretation, which for purposes of philosophical understanding is usefully considered as radical. So for Wittgenstein use fixes meaning, precisely because because use, together with other behaviour (broadly construed), exhausts the evidential basis of radical interpretation. The role of interpretation also imposes further constraints: e.g. that linguistic use must be part of an order in behaviour which is interpretable in virtue of regularities which are manifested in it. Meaning is both constituted and constrained by the possibilities for interpretation.

Agreement among Wittgenstein, Quine, and Davidson apparently extends from method through results. Thus the seeming abolition of logic of which Wittgenstein speaks in §242 appears comparable to Quine's attempted abrogation of the distinction between the analytic and the empirical. It seems that for Wittgenstein this is a consequence of regarding meaning as a construction which depends upon such empirical phenomena as sign-action regularities and agreement in judgments, and which can be represented in terms of the hypotheses of a radical interpreter, who 'will often have to guess the meaning...and will guess sometimes right, and sometimes wrong' (§32).

To see meaning in this way is, among other things, to abandon the notion of logic as a framework for judgment which is given *a priori*. Hence it is to give up the idea which Wittgenstein expressed at 5.552 in the *Tractatus*, that 'Logic is prior to every experience -- that something *is so*.' Still, to take logic as part of an interpretively imposed empirical theory is not to abolish it; for Witttgenstein apparently also holds that logic is encompassed in the agreement required for successful interpretation and mutual

understanding. Quine and Davidson make similar claims, e.g. in their use of the principle of charity. Their own approaches to meaning, moreover, are such as to yield the result that many sentences in the language of an interpretee will be analytic, in the sense that they will be true solely in virtue of the translations or meanings which are assigned them. So it might be said, paraphrasing Wittgenstein, that their accounts of meaning seem to abolish the analytic/synthetic distinction, but do not do so; rather they relocate the distinction, as part of an account of language which is ultimately empirical. What is abolished, as in Wittgenstein's account, is the pre-empirical *givenness* of the analytic, or of meaning itself.

Again, in discussing holism and the understanding of language Davidson remarks that Frege might have said that 'only in the context of the language does a sentence (and therefore a word) have meaning.'[15] But then in *Philosophical Investigations* §199 Wittgenstein did say (nearly) this, and while linking meaning with understanding: 'To understand a sentence means to understand a language.' (He also added 'To understand a language means to be master of a technique', and we shall see the importance of this below.) More particularly, Wittgenstein urges in §207 that if a system of sounds is not part of an order in behaviour which we can interpret we would not call it language; and this is comparable to Davidson's claim that 'a form of activity that cannot be interpreted as speech behaviour in our language is not speech behaviour.'[16] Similarly, Wittgenstein's conclusion that communication requires agreement in judgments is also drawn, and often stressed, by Davidson, as when he says that the purpose of radical interpretation is 'to make meaningful disagreement possible, and this depends entirely on a foundation -- *some* foundation -- in agreement.'[17] And like Wittgenstein in §241-2, Davidson takes this agreement to be a manifestation of a deeper consilience. As he puts it,

> ...The central point is that finding the common ground is not subsequent to understanding, but a condition of it. This fact may be hidden from us because we usually more or less understand someone's language before we talk with them. This promotes the impression that we can then, using our mutually understood language, discover whether we share their view of the world and their basic values. This is an illusion. If we understand their words, a common ground exists, a shared 'way of life'.[18]

Davidson's 'way of life' is less naturally associated with natural history or biology than Wittgenstein's 'form of life'; but their use of these similar phrases (which may be intentional on Davidson's part) registers a similar point. Agreement in language requires not only agreement in opinion, but something deeper, which deserves to be called agreement in life.

These relatively obvious points suggest that there may be deeper comparisons. In what follows I shall try to indicate that this is so, and that Wittgenstein's and Davidson's conclusions in particular can be seen as having a detailed common basis, in light of which they are mutually reinforcing.

## III

For this it will be useful to concentrate further on the theme emphasized by the placing of §205 and §206, namely the way Wittgenstein links the topics of rule-following and intentionality. Wittgenstein begins a familiar line of his discussion by considering learning to follow a rule like that for addition, with the example of adding 2 (the rule '+2'). He notes that someone may have been trained in the use of the rule as we have, and might seem to have learnt to follow the rule in the way we do, but not actually have done so. A learner might, for example, continue the series for '+2' correctly (as we would say) up to 1000, but then go on to write 1004, 1008, and so on. This would not necessarily show lack of an understanding on his part. It might be that going on in this different way is natural to him, and we might find an interpretation which explains this, and according to which it is 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 natural 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, however, raises the question as to how we know that we are supposed to follow the rule '+2' or 'add 2' 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 the one which is actually *correct*. And as Wittgenstein makes clear, this question can seem exceedingly difficult to answer. He has already noted that the infinite extension of the rule is in no sense 'present to the consciousness' of a person who follows the rule, and has considered various other ways in which one might be supposed to relate to it. Thus he has discused the supposed event of grasping a Fregean sense (§138ff), together with a version of his own early picture theory of understanding (§139ff); and also the state of understanding more generally (§147ff), and the physiological mechanisms which might be supposed to underlie it (§158). It is clear from this discussion that the present question cannot be answered by citing any of these things, since the question concerns the presumed (and so far unexplained) *correctness* of any such thing.

We may be inclined to say that we know we are to do as we do because we are to write what *follows from* the rule. But this again is no answer, because

> §186...that is just what is in question: what, at any stage, does follow from that sentence. Or again, what, at any stage, we are to call "being in accord" with that sentence (and with the *mean*-ing you then put into that sentence -- whatever that may have consisted in). It would almost be more correct to say, not that a new intuition was needed at any stage, but that a new decision was needed at any stage.

So we seem faced with a deep and general problem about meaning, put in §198 as: '...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?' This, clearly, is also a version of the question with which §206, quoted at the outset, begins. Suppose one person reacts in one way and another in another to the examples and training associated with the order or rule 'Add 2'. Which one is right, and why?

As Wittgenstein hints by assimilating rules to orders in §206, the problems he has raised about rules and language are also problems about intentionality, that is, about the relations of mental states to action and to the world. For at an intuitive level the question how we are to know what action really accords with the rule 'add 2' is just the same as the question how we are to know what action really accords with the order 'add 2', or again with the desire or intention to add 2. This again is a consequence of our practice, noted above, of describing intentions and other like states via sentences specifying the worldly circumstances towards which they are directed. In accord with this practice any rule or order 'R' expressed in our language will have a correlative desire or intention whose content is ascribed by 'R', that is, the desire or intention to (or that) R. (Intuitively, this will be the desire or intention to act in the way directed by the rule or order: thus to the rule or order 'add 2' there corresponds the desire to add 2 or the intention that I add 2, and so forth.) So any instance of such a question about the normative relation of language and reality as 'How do I know how act in accord with the rule "R"?' will be interconvertible with a corresponding question about the normative relation of mental states and reality, e.g. 'How do I know how to act in accord with the desire or intention to (or that) R?'

Clearly, as in the case of rules, we do know these things. Indeed, since knowing the contents of our intentions or sentences *is* knowing the actions or states of affairs which are supposed to accord with them in this sort of way, knowing these things is part of knowing what we intend or mean; and these are matters we take ourselves not only to know about, but to know more about that others characteristically do or can know -- matters in the sphere of our first-person authority. The problem Wittgenstein is raising -- and which, as Kripke[19] has made clear, could also be raised by someone taking the role of sceptic about our knowledge of what we mean or think -- is that we seem unable, in these cases, to give any real explication of what we know or how we know it. Thus in discussing the following of a rule in §197 Wittgenstein turns abruptly to the same sort of question, but now about intentionality and intention.

> ...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?

The original problem thus seems quite general. It relates to items or states which have what we might call sentential content -- content *that S*, which is assigned by the use of a sentence 'S' -- and any of the actions or items in the world which are supposed (or meant) to *accord with* these bearers of content. The same questions thus arise whether sentential content is taken linguistically, as in speaking of sentences as rules of language, or psychologically, as in speaking of desires (wants) or intentions which are articulated by sentences. So if we are to understand why in accord with the rule 'add 2' we *must* follow 1000 with 1002 (and not with any other number), then we have also to understand why in accord with the desire or intention to play chess we *must* play chess (and not any other game). The questions relate to thought and action as well as to language, and they concern both the constitution of the norms we take to govern these phenomena, and our knowledge of these norms. (Something is subject to a norm, in this sense, if it can succeed or fail to be in accord -- in correct accord -- with that norm.) 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. As he puts this in the case of intentionality 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.")

This remark again indicates the scope and depth of Wittgenstein's questions. Questions of normative accord extend to the circumstances which render *thoughts* true (a thought seems to know what makes it true), and so they arise alike for the conditions of satisfaction, fulfilment, realization, or whatever, which pertain to attitudes with content in general. This again can be seen as a consequence of the way questions about rule-following extend to sentences, and as they are used in the specification of the contents (or objects) of propositional attitudes. Thus 'How am I to know how to follow the rule "add 2"?' has the sentential version 'How am I to know the circumstances in which "I have added 2" is true?'. This evidently applies to connections between sentences and the circumstances in which they are true generally, and hence to connections between thoughts described by those sentences and the circumstances in which these thoughts would be true. (Whence, we may ask, concerning any sentence 'P', and so also any expressible thought that P, this determining of a connection with the situation that P, a situation which may not even obtain?) So again the questions hold for the connection between items with sentential content -- such as the wish, desire or intention that P -- and the related action and circumstances generally. Also, we see again that we cannot hope to answer Wittgenstein's questions by

reference to knowledge of the content of either thought or language alone. Since the same questions -- that of explaining the 'despotic demand' (or the 'superstrong connexion') and our knowledge of it -- arise in the same way for both, they must apparently be answered for both together.[20]

This is made plain in other remarks in which Wittgenstein both raises his questions and hints at answers. For example:

> 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 both these words would go for both sentences...
>
> 445. It is in language that an expectation and its fulfilment make contact.

Here Wittgenstein more explicitly considers our practice of describing the contents of thought and language, and links this practice with the questions under discussion. A sentence is used in the same way in assertion as in the description of an attitude like desire or expectation[21], and this is relevant to the problems we have been considering, that is, to how I can say that my desire or expectation has been fulfilled. In both cases we connect sentence and situation: for assertion we use the link to describe how the world has to be for the assertion to be true, and for desire or expectation we use the link to describe how the world has to be for the desire or expectation to be fulfilled. (This, apparently, is part of the idea that it is in language that an expectation and its fulfilment make contact.) So, again, problems about the content of a mental state, and that of the sentence used to describe it, arise as one: the problem, e.g., of the nature of the 'has to be' in the above formulations, and our knowledge of it. But here Wittgenstein also hints at an order of priorities in his treatment of these difficulties: it is in *language* that expectation and fulfilment make contact; and so, presumably, it is via language also that a wish knows what would satisfy it, or a thought what would make it true. Language is somehow to provide the key to the 'hardness of the logical must' and so to the other questions which Wittgenstein has raised.

We can sum this up roughly as follows: Wittgenstein raises two parallel questions, for both rule-following (meaning), and sententially ascribed attitudes such as desire, intention, expectation, and the rest. The first is the question as to what constitutes the normative requirements of language and thought: what makes it the case that we are constrained to mean (follow rules) and to think (to link thoughts, or again desires, intentions, etc., with action and reality) as we do? What makes any such thing the case, as, that 'He has added 2' is to be regarded as true just if someone has added 2? Or again what makes any such thing the case, as that a particular action accords with the desire or intention that one add 2, or the thought that this has been done. (What kind of superstrong connection do we find here? Alternatively, whence these despotic demands?) The second is an epistemic question related to the constitutive one, as to how we are able to know these normative requirements, that is, how we are able to know what we mean and think. How do we *know* any such thing as, that 'He has added 2' is true, just if someone has added 2; or again, how are we able know that a particular action accords with the desire or intention that a person add 2, or the thought that one has done so? (How do we know what these connections or demands are, as we must, in order to know that they are satisfied?)

As Wittgenstein raises these questions they seem at once pressing, and also to admit of no answer that we can give. 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. But trying to answer Wittgenstein's explicit questions, we can seem quite unable to elucidate either the basis of these requirements or the knowledge of them which comes so readily to us. (We can, of course, repeat or rephrase what we take ourselves to know, and we can affirm that we do indeed know it; but neither of these responses is to the point, and we seem scarcely able to go beyond them.) Now, however, we can see that Wittgenstein's point in both §205ff and §240ff is that we can cast light on these questions by considering radical interpretation, and the notion of interpretive regularity which he introduces together

with it. Just as Wittgenstein raised these problems by considering interpretation -- that of the putative rule-follower in §185 -- so he proposes to resolve them by considerting interpretation as well.

His overall idea can be indicated in a rough and preliminary way as follows. First, if we ask what makes it the case that there are rules which govern what we do, and which we can follow only by acting in certain ways, or again that we have intentions (etc.) which we can act in accord with, but only by acting in certain ways, one answer is the following: As a matter of empirical fact there is a complex order in our behaviour, which is part of our 'natural history', and characteristic of our 'form of life' (§25, §207). This order includes 'the common behaviour of mankind' and also the more specific linguistic practices interwoven with it. We naturally make sense of this order by interpretation; and interpretation proceeds by the imposition of norms, and requires the finding of correlations in behaviour, in which those norms are satisfied. So it is in interpretatively understanding our own actions and practices that we both lay down the requirements with which we are concerned, and also find these requirements regularly to be satisfied.

In interpreting we lay down norms: our interpretive hypotheses or judgments ascribe to ourselves customs, practices, rules, intentions, and other items or states which have normative content -- that is, which are such that particular actions can accord or fail to accord with them, and so can be termed correct or incorrect, succesful or unsuccessful, or the like. Also, in order to interpret people we must find empirical correlations: we must be able to correlate their linguistic and non-linguistic behaviour, and to correlate both with the environment. For Wittgenstein these features of intepretation -- the laying down of norms, and the establishing of correlations -- are interdependent. The correlations which interpretation requires us to establish are correlations whose elements are co-ordinated with one another in accord with the normative requirements which our interpretive hypotheses lay down. In particular, interpretation requires that we correlate the behaviour which we interpret as persons' use of words and sentences expressing their own intentions, rules, and so forth, with further behaviour, which we interpret as actions in accord with those desires, intentions, rules, etc. To correlate speech and action in this way is to see them as co-ordinated with one another, and also with the environmental objects and circumstances towards which we take such action to be directed, and in what we regard as the right way. The episodes in behaviour which successful interpretation requires us to correlate are episodes which we thereby find to be in normative accord -- in normative co-ordination -- both with one another and with the environment.

So, secondly, if we ask how we can knowingly behave in accord with such normative connections -- how, e.g., we can know that we are to act in a certain way, and as we do act, on a particular rule or motive -- a further answer is that our possession of this knowledge is already present, and also already manifest, in that order in our behaviour which we understand by interpretation, and which makes understanding ourselves in this way possible. Understanding one another as users of language requires correlating (behaviour best interpreted as) utterances specifying desires, intentions or rules with (behaviour best interpreted as) actions in accord with those desires, intentions or rules. Such interpretive understanding therefore already presupposes what we take to be the *correct* co-ordination of action with the linguistic expression of desires and intentions or the linguistic description of rules. (The required correlation has already been effected, and before we thought of it as a requirement, or wondered how such a requirement could be met.) The fact that we can understand one another as users of language already shows that the order in our behaviour is that of agents who regularly act (correctly) in accord with rules and intentions which they can specify, and hence that of agents who not only act in accord with intentions, but also speak so as to specify both their intentions and their actions in accord with them correctly. Thus if our actions are best interpreted in terms of intentionally following a certain rule, say, then although we use that rule blindly, still we follow it (and feel compelled to do so, to correct ourselves or accept correction from others in relation to it, and so forth) with sufficient accuracy to make interpretation in terms of that rule mandatory (that is, the best explanation of this behaviour we can give). Hence the way we grasp a rule, or again a phrase or sentence taken as an expression of intention, is ultimately shown, both to ourselves as to an interpreter, 'in what we call "obeying the rule" and "going against it" in actual cases' (§201).

This brief sketch of an answer, and the particular ideas in it, need to be filled out more clearly. We can do this by relating the main points of the argument to Wittgenstein's text in more detail.

IV

As noted, Wittgenstein stresses (§199, §202) that following rules, giving or understanding orders, and so forth, are customs, practices, or techniques. Taking this in light of §207, we can see that his claim is that these are *interpretable* practices -- that is, practices which admit of understanding in virtue of interpretable regularities holding over the use of language and other behaviour in them. These regularities, as we have seen, are correlations which are observed or postulated as holding over episodes in behaviour, where these episodes are described as the linguistic expression of rules, orders, or motives on the one hand, and actions in accord with these rules, orders, or motives on the other.

As has been made clear, we interpret a pair of expression-action episodes in this way by hypothesizing a further element, correlated with the others and linking them -- a desire or intention, whose content is specified by the interpretation assigned to the linguistic expression, and which is therefore directed to the type of action or situation specified in the same terms. So we can see that interpretive hypotheses, and the natural (causal) regularities which we understand through framing them, span precisely those cases of normative connection -- between the expression of a rule, and the behaviour which counts as action in accord with it, between an intention and the action which counts as fulfilling it, and so on -- which Wittgenstein poses us the problem of explicating. That is: the 'superstrong connections' we want to understand are normative connections which hold among the naturally correlated elements of interpretive regularities, as these are understood and co-ordinated via interpretive hypotheses; and the 'despotic demand' which an item or state with content makes on another, or upon reality, is the demand of one element of such a correlation, interpretively understood, for another, which would, as it were, render the correlation -- the interpreted co-ordination of mind, language, and reality -- more complete, and hence more fully intelligible.

A part of Wittgenstein's proposal is thus simply that we should see a range of regular connections -- between intention and action, between the expression of a rule and the activity of following it, among the drawing of successive conclusions in the course of a deductive argument, and so forth -- as at once naturally causally ordered, and also (in this order) naturally interpretable and so understandable in normative terms. The fact that the regularities are part of the natural causal order means that we can take them as subject to causal explanation in familiar ways -- say by reference to experience and training, and those aspects of our natural history which serve to explain why these have the effects they do. The same regularites, however, are also ones which we interpret, and so ones which we also describe and understand in a normative framework, for the imposition of which, again, regularity is required. So, overall, the instances of 'regular connection' of which Wittgenstein speaks in §207 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 regularities, but as regularities which correctness demands, and which support the kind of interpretation in which correctness or the lack of it is ascribed.

We can see this triple status, and its relation to the matters with which we are concerned, in the examples with which Wittgenstein both illustrates his questions and attempts to dissolve them. To bring this out it will be worth quoting at some length from §198 and §201-2.

> "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.
>
> ...It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each contented us for a moment, until we thought of yet another standing behind it. What this shows is that there is a way of grasping a rule which is *not* and *interpretation* but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.
>
> Hence there is an inclination to say: every action according to the rule is an interpretation. But we ought to restrict the term "interpretation" to the substitution of one expression of the rule for another.
>
> And hence also 'obeying a rule' is a practice...

The notion of 'custom' or 'practice' which Wittgenstein introduces here is plainly that which he links with regularity via radical interpretation in the remarks with which we started. So 'practice' here can be understood as 'interpretable practice' above, that is, practice manifesting interpretable regularity.

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 interpretee; but still it can be regarded as a concrete instance (token) of one of an interpretee's sentences, so we can treat this practice as involving interpretive regularity in the sense described above. In particular 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 reguired 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.[22]

Such an understanding of going-by-the-rule also yields answers to the 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 straightforeward. 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 Wittgenstein has already discussed, of the person 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 occassion 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.[23]

This remark gives grounds to fill out Wittgenstein's view more fully, and also to note some features of the rhetoric and argument of the *Investigations* more generally. Although the remark is an early one, it has a clear bearing on the themes we have been considering, which are made fully explict 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 interpretive 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, 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.[24]

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 can make use of the interpretee's own 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.

Also Wittgenstein's talk of hypotheses in this case, and his reference to natural laws in §54 above, make quite clear that we are not going wrong in describing his conception as one in which the observer who interprets a language (or game) can be described as framing hypotheses, which serve to explain and predict the behaviour of persons in a way which is partly comparable to the use of hypotheses dealing with other natural regularities. In this case, however, the hypotheses serve to explain in a particular way, that is, by specifying intentions on which persons act, and rules or norms to which they adhere, and seek to adhere.

Thus these remarks suggest, as noted above, that what distinguishes interpretation from other forms of empirical explanation is partly that the regularities and hypotheses with which an interpreter deals have a normative character, to which interpretees themselves are sensitive. Moreover, in cases of interpretive regularity of the kind we have been considering, the same norm characteristically figures in two ways: that is, both as a norm governing behaviour in accord with a sign (phrase or sentence), and also as a norm governing an action in accord with an intention or other psychological state. Thus in interpreting a sign as meaning 'turn left', we lay it down that an action will accord with the sign just if it is an action of turning left. In understanding action in accord with the sign in this way we also explain such action by reference to the intention to turn left. This hypothesis too is normative, for in framing it we lay down that an action which results from the intention will accord with the intention just if it is an action of turning left. The norm for acting correctly in accord with the sign, and that for acting successfully on the intention, are the same; and in both cases the norm can be regarded as linguistic, or as relating language and action -- that is, as embodied in a connection between the phrase 'turn left' and turning left.

This is not an accidental feature of the instance. Rather, it is a consequence of the conception of interpretation in terms of which I think Wittgenstein's remarks are best explicated that what we have been calling the norms of language, thought, and action can be seen as the same, and can be regarded as embodied in the practice of language itself. (This is part of what Wittgenstein means in saying that it is in language that an expectation and its fulfilment make contact.) This again is a consequence of the role of sentences in articulating motives, and hence in interpretive hypotheses. In framing interpretive hypotheses we use sentences to specify the meanings of sentences or the contents of motives. We use a sentence 'P', say, to characterize an interpretee's sentence as true just if P, an interpretee's rule as followed just if P, an interpretee's intention as fulfilled just if P, and so forth. In this we in effect use the sentence 'P' as a standard-bearer: the sentence by which we frame an interpretive hypothesis serves to lay down an hypothetical norm, which we apply to a sentence or motive of the interpretee, and hence to actions of the interpretee which can accord or fail to accord with that sentence or motive.

The norm or standard laid down by such an sentential hypothesis, moreover, is always an analogue of truth, as applied to the sentence in terms of which the hypothesis is framed. As is familiar, each of us knows an endless series of what Davidson calls 'snowbound trivialities' about his or her own language, such as:

'Snow is white' is true just if snow is white.

The series, and our knowledge, can be represented more schematically, as of the form

**NL**: 'P' is true just if P[25]

Let the label **NL** remind us both of natural language, and of norms of language: for the knowledge of language thus schematized can also be taken as knowledge of a set of norms, namely those in accord with which a particular individual, or a community of individuals with closely comparable idiolects, aims to speak. Our *that P* mode of description of content thus has the consequence that when we use our **NL** in interpreting others, we map the conditions of truth for our *own* sentences into those for the conditions of truth, satisfaction, fulfilment, or whatever, of the sentences and propositional attitudes which we ascribe to the *other*.[26]

This form of description thus serves to integrate the relational and normative structure of the concept of truth which holds for the sentences by which we describe motives into the sytem of causal explanation by which we represent the working of the motives described in this way. In this integration the relation of truth holding (let us say for convenience, and without ontological committment) as between sentences and situations reappears as description of causal relations holding as between motives articulated by those sentences and those same situations; and the normative character of truth then reappears as evaluation of these relations as ones which satisfy the causal norms which we take to be standards of causal coherence or success. The same sentential interpretive norm or standard thus appears in one guise when we speak of the conditions in which a speaker/interpretee's sentence is true, and in another when we speak of the conditions in which an interpretee's intention articulated by that sentence is fulfilled, and so has worked -- produced (caused) the requisite situation -- as it was supposed to do.

Since this claim will be important for our argument, it will be worth spelling it out in a way which links Wittgenstein and Davidson. For Wittgenstein interpretation involves finding actions intelligible or logical, as for Davidson it involves finding actions rational. This in turn can be partly explicated in terms of understanding actions as resulting from desires and beliefs which are appropriately related.[27] Thus, for example, if a person utters 'Snow is white' intending to say that snow is white, we may take him to have desired to say that snow is white, to have believed that if he uttered those words he would do so, and so to have wanted to do this. Such ascriptions have a familiar pattern:

A desires that P [that he say that snow is white]
A believes that if Q then P [that if he utters 'Snow is white' he says that snow is white.]
A desires that Q [that he utters 'Snow is white.']

Read from the bottom up, the sentences which articulate a reason of this type have the pattern of *modus ponens*, that is, a transition which is a paradigm of logic. This makes clear that if the agent succeeds in satisfying the final desire in the pattern, then, provided the belief in the pattern is true, the agent must also satisfy the desire which heads the pattern. To understand action in accord with such a pattern is therefore to find it logical and rational. Let us call this pattern **PR** (for the pattern of reasons, or practical reason). Then using 'A' to stand for the agent, and 'des' and 'bel' for 'desires' and 'believes', we may write it schematically as:

**PR**: A des that P & A bels that if Q then P -[causes]-> A des that Q

**PR** deals in desires, and we take it that action can be explained as resulting from desire. In intentional action the causal role of a desire that P is to bring about the action or situation desired, that is, to bring about P; and this in turn should lead to the agent's forming a belief that P, which serves (perhaps together with the desired situation itself) to *pacify* the agent's desire, that is, to cause it to cease to operate. This gives a pattern which we may call **D**, which shows the life-cycle of a single desire in intentional action. We can represent this uniformly with PR as follows:

**D**: A des that P -[causes]-> P -[causes]-> A bels that P -[causes]-> A's des that P is pacified.

Now this pattern contains within itself a further one concerning belief, which we may lable separately as **B**.

**B**: P -[causes]-> A bels that P

We find instances of this form not only in persons' awareness of what they have done, but more widely in the case of belief based upon experience or perception. In general, and as indicated above, to perceive that P is to have perceptual or experiential reason to believe that P, which is caused by the situation which renders 'P' true. Hence we might represent a situation described by **B** more fully by

**B*** P -[causes]-> A has an experience (or perception) as if P -[causes]-> A bels that P

These patterns illustrate the claims made above. Each is a paradigm of the normative. **PR** is that of the rational formation of desire; **D** that of desires which are acted on successfully, and thereby satisfied and pacified; and **B** that of beliefs which are true, and rightly caused by what renders them so. Each can be regarded as implemented via the repeated use of a sentence (or sentences) of the interpreter's language, namely those that appear in actual examples in place of the repeated P's and Q's; so that the particular norms which the use of a pattern imposes are those borne by sentences of the interpreter's **NL**. We can take it that in understanding others we make tacit use of such patterns; so these norms systematically reappear as the conditions of satisfaction and truth of the desires and beliefs which we ascribe, as a matter of empirical hypothesis, to those we interpret. An hypothesis that an interpretee's behaviour is rational action in accord with **PR** entails that this behavior falls under **D** (and hence **B**) at least twice, that is, in respect of both the initial and derived desires which figure in the pattern. Where an interpreter applies **PR** to actions as they unfold, each instance of **D** specifies a series of demands which are at once normative and predictive, that is, which constitute predictions contingent on the interpretee's success in action, proper function in the formation of belief, and so on; and we regard our interpretations as disconfirmed by the failure of these predictions, and confirmed by their success.[28] So considering these patterns enables us to regard interpretation as a process in which each of us systematically re-finds the norms of his or her *own* idiolect in the causal patterns in *another's* behaviour, and in this discovers that these causal patterns conform to these norms, and so are intelligible.[29]

As Wittgenstein also points out, the norms employed in interpretive explanations of this kind are ones which the interpretees themselves accept; so that interpretive regularities are ones to which interpretees both naturally conform and also seek to conform (or feel they ought to) -- towards which they will accept correction, try to correct themselves, and so on. We not only tend to be such, but also strive to be such, as to be interpretable in these ways. Interpretive understanding is thus a system which, among other things, partly serves to shape and regulate (control) the behaviours which it describes; but a system of explanation is no worse for a confluence of vectors which tend to ensure that the things it represents gravitate towards the way it represents them.

This also makes clear that the case of intention upon which Wittgenstein concentrates is closely comparable to that of other motives which also relate to action, but less directly. As well as linguistic expressions of intention, there are linguistic expressions of expectation, hope, fear, belief, desire, and so forth. Just as we interpret by correlating linguistic expressions of intention with the actions and situations which accord with them, so also we correlate linguistic expressions of desires, beliefs, hopes, fears, etc., with behaviour related both to the possession of these motives and to the situations in which these motives are fulfilled, satisfied, realized, verified, or whatever. These correlations too are made by successive uses of the same sentence in interpretive hypotheses, and so are both sentential and normative. By considering the remainer of Wittgenstein's remarks quoted above, we can see that he takes such correlations to be akin to those we have so far considered.

## V

The remarks we have taken are mostly concerned in one way or another with intentionality, the capacity of thought to relate to its object. As Wittgenstein says in first introducing the issue, this can seem mysterious and remarkable:

> 95. "Thought must be something unique" For when we say, and *mean*, that such-and-such is the case, we -- and our meaning -- do not stop anywhere short of the fact; but we mean: *this - is - so*.' But this paradox (which has the form of a truism) can also be expressed in this way: *Thought* can be of something which is *not* the case.

In line with the explication above, Wittgenstein's remarks seem partly meant to show that this remarkable-seeming feature of thought can be regarded as a relatively unmysterious consequence of our practices in using language, together with our concomitant practice of interpretive explanation, that is, our practive of finding the actions of persons to be intelligible or logical, and the conseqence of intentions and other motives regularly expressed in speech. Thus in §43, as we saw, Wittgenstein notes that a wish seems to know what would satisfy it, or a thought what would make it true; and also that this can seem a 'despotic demand'. This now appears as a reflection of our linguistic and interpretive practice: wishes and thoughts 'know' what would render them satisfied or true, in the sense that we describe them by sentences which also describe what would satisfy or verify them. This form of description serves to fix the normative requirements, and hence the predictive claims, implicit in explanations of behavioural regularities linking the expression of wishes and thoughts with their satisfaction, verification, etc. If we use a sentence 'P' to describe a hypothesized wish or thought, we thereby lay it down that the wish will be satisfied, or the thought verified, just if P; and this, as Wittgenstein says, is a demand which nothing else can satisfy, and for something which may not even obtain. Seen this way, however, the demand is not mysterious, but a straightforeward aspect of the working of a particular mode of explanation.

Reference to this form of description and explanation also provides an account of the 'super-strong' connection between intention and action discussed in §197. Such a connection, as we have seen, is one which is both causal and in accord with (or demanded by) a norm; so it can appear as both real and efficacious (casual), and yet also having as having a strength of connection with its effect which surpassess anything actual. Thus the intention is a cause of this action, and it is an intention for *just this* kind of action: so it is a kind of cause which seems already to contain its effect within it, or to be connected with its effect in a super-empirical way (cf §194). But given that the connection between intention and action, like that between sign-post and action, is causal, one should also ask, as Wittgenstein does in the case of an intention to play chess, how we know about this connection: 'Is it experience that tells me that this sort of game is the usual consequence of an act of intending? so is it impossible for me to be certain what I am intending to do?' As Wittgenstein says, this may be nonsense; but then how is the connection between intention and action made and known? Part of the answer, again, seems clear from the discussion so far. Intention and action are naturally (causally) connected in behaviour which manifests interpretable regularity, that is, in which there is correlation between (behaviour interpretable as) the linguistic expression of rules or intention and (behaviour interpretable as) action in accord with these rules or intentions. We make the connection explicit by interpreting the regularity. So my natural certainty that I intend what I do -- my first-person authority about intention -- is an integral part of my (interpretable) form of life, and a condition of my being such that others can understand me.

This, however, points to a further connection between intention (and other such states of mind) and linguistic practice. In §205, as noted, Wittgenstein asks how the rules of chess are 'present in the mind' of someone intending to play chess, and this is a reminder that what someone intends, means, expects, etc., may go beyond anything which could plausibly be said to be a content of that person's consciousness at the time. The same point is also clear in the case of such an intention (or rule) as that to add 2, or the variety of instances of some descriptive word, which could not be borne in mind in this sense. We tend to link intentionality with consciousness. But since the capacity of our minds to reach out to reality -- our capacity to intend or mean -- goes beyond what we are conscious of, reference to consciousness cannot explain it.[30]

Rather, Wittgenstein holds, this capacity requires also to be explained by reference to language and practice. Thus as Wittgenstein both asks and answers the same question 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." 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, and the many co-ordinate actions which constitute the teaching and learning of chess, the playing of actual games, and so forth. Interpretation exhibits the links among spatially and temporally disparate constitutents of such a practice by co-ordinating them 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 part of such a practice. So an intention naturally reaches beyond consciousness, via the linguistic practice in terms of which the intention is expressed.

Thus having emphasized in §199 that to understand a language means to be master of a technique, and having urged in §205 that such technique (interpretable practice) serves to explicate the connection between intention and its object, Wittgenstein can add in §445 that it is in language, thus understood, that expectation and fulfilment make contact. This short remark, I think, can be seen as a culminating point in his discussion, summarizing much that has gone before. In holding that the expressions and objects of such propositional attitudes as intending, wishing and expecting are linked as potentially correlated parts of interpretable practices, we in effect hold that the link between these attitudes and their objects is partly consitituted by the use of language itself.

Wittgenstein leads up to this point by considering self-ascription. Thus in §441 he takes the example of wishing:

> 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.

Wittgenstein's reference to nature and training indicates that he is here treating the linguistic practice of expressing wishes and other propositional attitudes in a way analogous to that in which he treated the practice of following a sign-post, that is, as involving the holding of regularity between utterance and other behaviour which is subject to interpretation. His idea, I think, is that wishing and other attitudes are connected with a particular kind of sentential regularity, which also sustains their interpretive ascription.

These states are marked by the presence of a natural basis for the development of an expressive verbal disposition, namely, a disposition to utter certain sentences ('By nature...we are disposed'). The development of this disposition in the course of learning language provides a kind of spontaneous behavioural expression of the state, which may replace more primitive expressions, or come to be co-ordinated with them (Cf. §244: '...words are connected with the primitive, the natural expressions...and used in their place'; and the application of this notion to states like intention, e.g. at §647.). Such an expression serves also to specify further events and circumstances related to the state, namely those in which the wish is fulfilled, the expectation satisfied, or whatever; and so the agent's behaviour relating to these further circumstances also becomes relevant to the ascription of the state.

This constitutes further regularity of the general kind which Wittgenstein mentions in §207, as holding between utterance and other behaviour. For it means that the expressive utterance, and hence

the state it expresses, require to be interpreted as in co-ordination with further behaviour, which is related to the further events or circumstances specified by the utterance itself; and this further behaviour, as Wittgenstein notes, may itself include, or be correlated with, another utterance of the same kind. So this is in effect an extension, or a fuller specification, of the original claim about interpretive regularity made in §206-7. What enables us to take utterances as expressions of motives or mental states is the way such utterances can be correlated with further behaviour which is to be understood as issuing from the same motives, and hence related to the situation upon which that motive is directed, and hence also to the agent's further utterances regarding that situation.

In §441 Wittgenstein concentrates on the link between such a disposition and first-person authority. Then in §444 he moves the discussion foreward, asking what constitutes the link between the behaviour which expresses such an attitude and the events which fulfil or satisfy it:

> 444...
>
> 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.

In focussing on 'I expect he'll come in' Wittgenstein is again considering utterances which serve at once to express an attitude and to specify further events related to it, namely those which would fulfill the attitude, satisfy it, or whatever. His point now is that this potential for relation to the same type of utterance (sentence) can itself be seen as constituting the similarity which links the attitude and the satisfying events. In §444 he stresses that a person who uses the sentence "I expect he is coming" thereby uses the words "he is coming" in the same sense as in the assertion "He is coming", since this is the way these words are used in saying that the expectation has been fulfilled. So the similarity which binds expectation and fulfilment is that both are potentally correlated with successive uses of the same sentence.

In taking the use of language as naturally related to other expressive behaviour, Wittgenstein takes a speaker/interpretee to use sentences in two connected ways: first to express expectations or other propositional attitudes, but in a way which relates them to events which may be remote in time and space from the expression (the use of 'he is coming' as in 'I expect he is coming'); and secondly to register that such expression- and attitude-related events have or have not occurred (the use of 'he is coming' as in assertion). In terms of our discussion, this is an ablilty on the part of a speaker to create interpretive regularity: that is, (i) to behave (speak) in a way which at once specifies the content of an element in such a regularity and relates the future behaviour of the utterer to a norm, and (ii) again to behave -- perhaps later, at a distance, etc. -- so as to indicate that this norm has or has not been met. This in turn makes it possible for an interpreter to make a parallel ascription of attitude and object, this time by the repeated use of the same sentence (or one with the same meaning) in framing and testing an interpretive hypothesis. The expectation that P and the situation that P thus meet in the potentially repeated use, by both interpreter and interpretee, of a sentence which means that P. The capacity to use sentences in this way renders our articulate propositional attitudes possible by providing for their objective ascription: for the capacity both consititutes a speaker's ability knowingly to link such attitudes with their objects, and enables an interpreter to make the same links in the course of discerning sentential regularities.

Thus on Wittgenstein's account the capacity of thought -- expectation, wishing, etc. -- to seize upon an object (even one which is not there), and with the grasp of the logical must, reflects the fact that people so behave that their behaviour can be explained in the way we have been considering, and that both the behaviour and the explanation involve the use of language, and in agreement, by interpreter and

interpretee. It follows that we cannot credit such present- and consciousness-transcending thoughts to animals or to infants; for in the absence of an interpretable practice which spans time and space in the manner of human language, the question as to how these attitudes relate to their objects cannot be answered (cf §650). Here again, Wittgenstein and Davidson draw similar conclusions, and on the basis of requirements relating to interpretation which are closely related.[31]

VI

Wittgenstein's remarks presuppose that we have practical capacities to express and describe desires, intentions, and the like, and to follow rules. They say little further, however, about how we manage to do this. Indeed on Wittgenstein's view there is little further to be said. These capacities are to be accepted[32] in philosophy as part of the natural order, and hence perhaps as having no explanation apart from what explains order of this kind. So in our own case, as Wittgenstein stresses, we ascribe ourselves particular intentions, or feel bound to act in particular ways in accord with rules, but on no basis of which we are aware. We do what we are supposed to do, but blindly. Hence when Wittgenstein asks the corresponding question about rules -- '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' -- we are particularly liable to feel at a loss for an answer.

Now, however, we can see that Wittgenstein's point in asking this question is to make us aware that answer to it neither can nor need be given; and also that our puzzlement at this point in his text is partly a consequence of failure to grasp an aspect of his account of these matters which is relatively straightforeward. We assign meaning by interpreting actions and utterances, and in so doing we constantly both lay down norms and register that thought and behaviour is proceeding in accord with them; but this is something we can do in the main only for the use of language by others, and not in our own case. It follows that our assignments of meaning can be justified, as empirical interpretive hypotheses; but that the interpretive justification for an assignment of meaning to a particular person's utterances is not available to that person in the prespective of his or her own case.

We have seen that Wittgenstein takes the question 'How do I know what rule a person is following?' as one to be answered by framing interpretive hypotheses and testing them by their behavioural consequences, which may be predictive. (Such hypotheses go beyond the data, and so are fallible; but this is a general feature of hypotheses, and no objection to any form of understanding which proceeds by framing and testing them.) He applies this alike to the case of figuring out the rules of a game (§54) and the rules governing someone's use of words (§82), to settling the question which way of going on from examples and training for a rule is the right one (§206), and to interpreting the behaviour of people alien to us, even if they speak only in monologue (§243). But of course if interpretation thus enables us to understand the rule another is following, it also enables us to know what that other ought to do at a particular point, in order to act in accord with the rule. So the difficulty is not with answering such questions in general, and by reference to interpretation; but only with doing so in one's own case.

The assymetry here -- which plays a partly comparable role in Davidson's account of first-person authority[33]-- can be brought out further as follows. An interpreter establishes what an interpretee means by framing hypotheses which in effect map sentences of the interpreter's language (used in content-specifying hypotheses) on to the utterances and actions of the interpretee, so as to yield an assignment of meaning or content to both. Clearly, however, an interpreter cannot take this same hypothetical attitude towards the meaning of his or her own sentences. For, as Wittgenstein repeatedly argues, each sentence used in such an hypothetical assignment of meaning (each 'substitution of one expression of the rule for another') would still require to have its meaning determined in the same way; so no such interpretation could fix the meaning of any sentence. One cannot apply interpretation in one's own case without massively presupposing the results of interpretation, that is, without presupposing that one does, after all, really understand what one is saying and doing. So in one's own case, as Wittgenstein puts it, 'any interpretation still hangs in the air with what it interprets, and cannot give it any support'.

But of course this does not show that one does not grasp the meaning of rules or sentences in one's own case. (The fact that my use of language cannot certify its own correctness to me in this way does not show that it requires certification of this kind.) As a matter of fact we naturally and spontaneously use language both for understanding the world and other persons, and in such a way as to be interpretable by them. So each person's use of language both proves it worth in making sense of others and the world, and also is certifiable for correctness by a potential for interpretation. Such certifying interpretation, however, is something that one can provide for others, but not for oneself. In understanding one another in practice we each thereby ratify the other's spontaneous self-expressive and descriptive uses of language, and by a process of interpretation which discerns in these uses an order which is genuinely intelligible, and objectively there. This, I think, is the only form of ratification which such a practice could require.

Thus what consideration of one's own case makes particularly salient is precisely, as Wittgenstein says at §201, that 'there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.' If I can make myself understood as a speaker of language then my account of my own motives and rules is by and large correct, despite the fact that further interpretation *on my part* of my use of language would hang with what it interprets. For another's interpretation of my account of the rule, and of my motives and practice in following the rule, does not hang in the air at all. Rather another's interpretation can be seen as a form of empirical hypothesis, grounded in what must be the strongly correlated descriptions of my uses of language and other actions in relation to objects and circumstances in the world, and confirmable as such. So Wittgenstein's point is not that we face a problem in justifying our assigments 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 assymetric 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.

This account is of course very schematic, and leaves out much of what Witgenstein says. Still, it suggests that the questions which he poses seem so perplexing partly because we fail to consider them in light of the kind of account of meaning which he -- and also Quine, Davidson, and of course very many others -- puts foreward. In this account judgments about mind and meaning are to be regarded as ultimately answerable to the interpretation of ordered behaviour in the world. The most straightforeward paradox, as one might say, is simply that the first-person perspective in which we are most authoritative about the phenomena of mind and meaning -- that in which we find no doubt, and in which we discern no indeterminacy -- is not that in which an account of these phenomena is ultimately to be grounded, or in which such an account is to be regarded as justified objectively. Like Descartes, we tend to assume that where clarity and certainty are, there ground and justification must also be, and this is an error. So tenacious is this conception, however, that when we meet Wittgenstein's arguments against it we may feel that he is denying the existence of the mental, abolishing meaning, speaking incoherently, or the like. This, however, is part of the dialogue in which he engages us: for here as elsewhere Wittgenstein's remarks are so framed as to make us cleave to the inarticulate assumptions in which the confusions he takes us to share with him[34] are rooted, at the same time as he exposes these assumptions as untenable, and indicates an alternative.

## VII

So far we have tried to explicate Wittgenstein's views by emphasizing the role of interpretation in §206-7 and such associated remarks as §54, §82, and §243. These remarks indicate that Wittgenstein took interpretation to involve the detection of what we have been calling interpretive regularities in behaviour, and hence to be a counterpart of the notions of use and practice in whose terms he sought to explicate meaning. By themselves, however, such remarks give little further information as to how Wittgenstein thought of interpretation, or indeed as to whether he gave the matter much explicit

consideration at all. It seems to me, however, that there is some textual evidence suggesting that Wittgenstein both had a particular conception of interpretation and made some claims as to its limitations in casting light on meaning. It will be worth following this out, not only because it facilitates a more informative comparison with Davidson, but also for its own sake.

As is well known, Wittgenstein introduces his later ideas by contrast with a quotation from Augustine; and the passage which he quotes is, as he notes, closely connected with the topic of interpretation.

> When they (my elders) named some object, and accordingly moved towards something, I saw this and I grasped that the thing was called by the sound they uttered when they meant to point it out. Their intention was shown by their bodily movements, as it were the natural language of all peoples: the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of voice which expresses our state of mind in seeking, having, rejecting, or avoiding something. Thus, as I heard words repeatedly used in their proper places in various sentences, I gradually learned to understand what objects they signified; and after I had trained my mouth to form these signs, I used them to express my own desires.

Wittgenstein relates this passage to his own *Tractatus Logico-Philosophicus*, and refers to it often; his summary comment, however, is the following:

> 32. Someone coming into a strange country will sometimes learn the language of the inhabitants from ostensive definitions that they give him; and he will often have to *guess* the meaning of those definitions; and will sometimes guess right, sometimes wrong.
>
> And now, I think, we can say: Augustine describes the learning of human language as if the child came into a strange country and did not understand the language of the country; that is, as if it already had a language, only not this one. Or again: as if the child could already *think*, only not yet speak. And "think" would here mean something link "talk to itself."

Thus Wittgenstein first introduces the figure of the explorer/interpreter, whose practice he will later take as a touchstone in his claims about meaning, as part of a comment on Augustine's description of language-learning; and his comment is precisely that Augustine treats the child who is learning languge as such a figure. At the end of section V above we considered the line of thought which shows why Wittgenstein takes this to be a significant criticism of Augustine. Now we can also note something further. In saying that Augustine describes the child as if he were an explorer/interpreter, Wittgenstein must surely be implying that Augustine's description can *also* be taken as an account of the activity of such a radical interpreter, in the use of this figure which Wittgenstein himself later makes. On this reading, Wittgenstein would here be more or less explicitly connecting what Augustine says in §1 with the conclusions which he will draw by reference to the explorer/interpreter in §206, §207, and §243.

If this is so, then Wittgenstein's comment at §32 should be seen as including, as well as criticism, a tribute to the philosophical penetration of Augustine's description of infancy, taken as concerned with interpretation. For as §32 implies, the task which Augustine assigns to the infant in §1 is in fact the same as that which Wittgenstein himself assigns to the explorer/interpreter in §207 and §243: namely, that of proceeding from observation of utterance and other behaviour to the interpretive understanding of language. Further, and as Wittgenstein seems to be indicating, Augustine's remarkable description of the fulfilment of this task presages many of Wittgenstein's own later conceptions and conclusions, including those of §206-7, in some detail. That is, Wittgenstein seems here to be highlighting the way Augustine's description in fact characterizes the behaviour -- including the interwoven roles of utterance, action, and the natural expression of intention -- which Wittgenstein himself takes as basic to the understanding of language. In particular it seems that Augustine's account of 'bodily movements, as it were the natural

language of all peoples' is to be seen as anticipating, and perhaps as inspiring, both Wittgenstein's conception of the 'natural expressions' of sensations and various propositional attitudes, and also that of 'the common behaviour of mankind' which provides 'the system of reference by which we interpret an unknown language'.

In this light Augustine's description of the infant appears particularly suited to the purpose for which Wittgenstein employs it, that is, as an introduction to his *Philosophical Investigations* as a whole. For Wittgenstein's reader can see Augustine's description as embodying not only errors from Wittgenstein's early work, but also insights from his mature philosophy; and the reader can see this by imposing on Augustine's description the same change in perspective -- roughly, the shift from a Cartesian view of psychological and semantic concepts to one which emphasizes their ascription of the basis of interpretable practice -- as Wittgenstein seeks to effect in the course of the *Investigations* itself.[35]

In any case the similarity between problem and solution in §1 and §§206, 207, and 243 is obvious; and a literal reading of §32 requires us to look to Augustine's description as an possible account as to how the kind of interpreter Wittgenstein has in mind in the later remarks might be supposed to proceed. Once we do so the relevance of Augustine's description again seems unmistakable. For Augustine represents his infant/interpreter as attempting to find just the kind of 'regular connections' between behaviour understood as utterance and behaviour understood as action related to utterance, as Wittgenstein mentions in §207, and which we have discussed at length above. Augustine's interpreter, that is, proceeds by trying to correlate the utterances of articulate sentences ('words repeatedly used in their proper place in various sentences') with intentional actions ('seeking, having, rejecting, or avoiding something'); and this correlation is mediated by the interpreter's natural understanding of intention ('our state of mind in seeking, having, rejecting, or avoiding...') as this is shown in naturally expressive behaviour, that is, in the 'bodily movements, as it were the natural language of all peoples' by which this state of mind is expressed and understood ('the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of voice...'). (The 'something' thus sought, rejected or avoided is the object or target of intention, the situation at which intention, and hence intentional action, is directed.)

Thus, as §32 would lead us to expect, Augustine's description in §1 merges seamlessly with what Wittgenstein says about the explorer/interpreter at §243, as well as with what we have said about interpretation in explicating §206-7. As Wittgenstein describes matters, the interpreter is sometimes able to guess that he is hearing expressions of desire or intention (In §1 as in §243, it seems, the interpreter 'hears them making resolutions and decisions'), and also sometimes able to guess the actions or situations intended. So, as we may take it, the interpreter tries to increase this nascent understanding, by connecting such hypothesized utterances and actions (or action-related situations) in a systematic way. The interpreter does this by framing further hypotheses, which would serve to link the language and actions of the interpretee with the objects and situations constituting their common environment more generally. Thus, as it might be, the interpreter hears an utterance of 'I will tie my shoelaces', and also sees that the utterer ties, or tries to tie, her shoelaces; and the interpreter takes this utterance and action to be connected, as expression and fulfilment of intention.

On the basis of such data, the interpreter seeks to construct hypotheses which will connect further sentences 'S' which the interpretee could use as an expression of intention with further situations that P which the interpretee intends. The interpreter seeks, that is, to construct a projectable utterance-action correlation, holding as between sentences 'S' and intended actions or situations that P, and hence of the form 'S'...P; and hopes eventually to exploit and test this correlation, by uttering an appropriate 'S' (in a suitable intention- or desire-communicating tone of voice) as an expression of desire that P. Augustine's interpreter supposes that mastery of such a correlation constitutes an important part of mastery of language, and one which enables persons to make themselves understood.

The correlation, however, is not a simple one, for Augustine's interpreter in §1, like Wittgenstein's in §243, is dealing with an 'articulate' human language. Hence the interpreter must understand sentences on the basis of understanding the words from which they are composed. Accordingly, the interpreter seeks to discern the syntax of the sentences, by concentrating on the 'proper places' of the words as they occur in them; and at the same time the interpreter tries to link these words with something further, which will serve to determine the action or situation (related to 'seeking', 'having', etc.) which the combinations of words in sentences are used to specify. This, as we can now see, locates more precisely the point at which the later Wittgenstein begins to diverge, both from Augustine and from his own early work.

Augustine takes it, as did Wittgenstein in the *Tractatus*, that what someone seeking to understand language must link with words in order to understand the sentences from which they are composed are 'the objects [those words] signified'. So as Augustine represents matters the aim in coming to understand a language is to frame hypotheses about (i) the referents of words and (ii) the way words are used to form sentences, which in turn yield hypotheses about (iii) what (desired or intended) situations these sentences are used to describe. A natural way to think of this is to regard (i) and (ii) as determining the situations in which sentences are held true, and hence those in which intentions are fulfilled. And this, of course, is the way Wittgenstein thought of the matter in the*Tractatus*, in which he repeatedly stressed that we understand sentences in an articulate language through a grasp of the 'rules of projection' which map the logical forms of sentences and the referents of their words onto the conditions in which they are true; and he also linked this with translation, and the way that translations of sentences proceed via those of their constitutent words.[36]

Wittgenstein's quotation of Augustine in §1 thus evokes an account in which understanding and meaning are ultimately explicated in terms of reference and truth. As is familiar, and noted above, Wittgenstein immediately opposes this account with another. What the shopkeeper must connect with 'five', 'red', and 'apples', in order to understand their combination in 'five red apples', are not *objects*, but rather the distinct kinds of *activities* -- the putting of things into one-to-one correspondence with numerals, the comparing of things with standards of colour, etc. -- which constitute the practices of using those words. These are the practices linked with words, by what Wittgenstein calls rules for their use; which he takes it that an interpreter can sometimes 'read off from the practice' of language, as from a game, 'like a natural law governing the play.' (§54, §82). Discerning such rules enables an interpreter to understand utterances of sentences as actions which are interwoven with others to constitute the practice of using language, as in the example of 'five red apples' with which he begins. This makes it possible to understand these sentences as instances of interpretive regularities, and thus to satisfy the requirement on interpretation urged in §207; so it enables an interpreter to proceed from 'Hearing words repeatedly used in their proper places' (§1) through finding 'regular connection between what they say, the sounds they make, and their actions' (§207) to 'translating their language into ours' (§243).

This picture differs significantly from Augustine's, for Wittgenstein holds that the practices involved in the use of words are no more to be grasped through the consideration of objects which the words signify than the rules of chess are to be discovered by contemplating chess pieces (§31). Only someone who is already competent in the relevant practices involving objects can gather the use of a word by coming to know the object for which it stands; and it is in terms of practice that concepts like reference (§37) and truth-conditions (§437) require ultimately to be understood. (This particularly applies to the normative aspect of truth, stressed here; for a norm is essentially something by reference to which a human performance can be assessed as correct or incorrect.) Thus although these practices, or rules, can be thought of as linking words with their referents, or again sentences with the conditions in which they are true, the situation is not, as Augustine suggests, that we learn the practices through grasping these links. Rather we can think of ourselves knowing the links -- as relating words to objects or sentences to situations -- only because we have mastered these practices. For we come to possess the concepts and capacities for thought which we connect with words and sentences in the course of learning to use the words and sentences themselves (§208, §381, §384). So, for example, it is not that we come to understand the word 'red' by thinking of it as linked with the colour red; for we can think of word and

colour as connected in this way in this way only if we possess the concept red, and we come to do this in the course of learning to use 'red'. What Wittgenstein wishes to stress is that the infant learns to *do* as the elders do, and *therein* comes to think as they do; and this is why explanations in terms of the way users of language act provides a terminating point in an account of language.[37] This is precisely what Augustine obscures, in representing the infant as an interpreter, that is, as someone who brings the understanding of one language -- and hence the capacity to think in terms of a range of concepts, including those of reference and truth -- to bear upon the understanding of another.

VIII

Now much of Wittgenstein's discussion of Augustine and the *Tractatus* does not serve to relate him to Davidson; for the latter is perfectly clear, for example, as to the necessity for distinguishing between the conceptual resources of infant and interpreter. But in part Wittgenstein is criticising the employment of the notion of reference in explicating the use of language more generally, on the grounds, e.g. that it leads us to assimilate linguistic practices which are utterly different, and hence to give an account which is misleading or vacuous (see, e.g. the forceful critique at §§10 - 14). This evidently carries over to the Tarskian notion of satisfaction which is at the heart of Davidson's account. For as Davidson explains, we can regard satisfaction as 'a generalized form of reference', and so take Tarski to have shown how the truth of sentences depends upon the reference of their parts.[38]

Bearing this in mind, we have the following preliminary comparisons, rough and schematic as they are. Both Wittgenstein and Davidson regard the interpretation of language and action as proceeding together, and as involving empirical hypotheses whose nature can be elucidated by radical interpretation; and both in consequence see a speaker's use of sentences as underwritten by a potential for interpretation which links the speaker's use of sentences with other actions, and which that speaker does not perform in his or her own case. Again, both take it that the abilities of someone who can use or interpret the sentences of an articulate language can be described, but only partially, via hypotheses which link words and their referents with sentences and the conditions in which they are true. Wittgenstein thought of concentration on the function relating reference to truth as promoting an account of the relevant abilities which assimilated fundamental distinctions in practice; and Davidson holds that a Tarskian theory which serves partly to specify such a function requires to be supplemented by another, which treats of the agent's preferences and choices.

This brings us to an obvious and significant contrast, namely, Davidson's use of explicit formalized theories of truth and decision. Hypotheses linking reference and truth concern 'grammar' in Wittgenstein's sense of the term, which encompasses both the syntax and semantics of natural language and also many other aspects of linguistic practice. Wittgenstein was willing to consider that grammar in this sense might be the subject of an explicit empirical theory, e.g. one which described language 'as part of the psycho-physical mechanism'; and indeed modern work in lingusitics, e.g. that by Chomsky, aims to encompass much that Wittgenstein seems to have had in mind (as well, of course, as much that he did not).[39] But Wittgenstein did not attempt to frame such theories, nor did he contemplate using them in a conceptual exercise aimed at casting light on the practical capacity for understanding language. (Thus despite the emphasis which Wittgenstein lays upon utterance-action regularity, or other general features of what he takes to be the grammar of language, he goes no further towards any systematic representation of such generalizations than indicating, e.g., that it is a matter of grammar that if an order runs "Do such-and-such" then executing the order is called "doing such-and-such", and that something similar holds for rules, intentions, and the like.) By contrast, as noted, Quine discussed radical translation in terms of empirical theory from the beginning, and following him Davidson saw that Tarski's theory of truth could serve in an account of meaning at which a radical interpreter might aim, and by which, therefore, the ability to understand linguistic behaviour might be explicated more systematically. In consequence Davidson assigns the notion of truth a theoretical role in the elucidation of meaning (or what Wittgenstein called grammar) which goes beyond anything which even the author of the *Tractatus Logico-Philosophicus* could have envisaged.

This difference, however, is also bound up with an important similarity. For Davidson's emphasis on the role of action bears comparison with that of the later Wittgenstein. In his more recent work Davidson has taken the fundamental notion in radical interpretation to be that of an interpretee's *preference for the truth of one sentence to another*. In this connection he asks us to 'observe that every utterance that can be treated as a sincere request or demand may be taken to express the utterer's preference that a certain sentence be true rather than its negation';[40] and the same, of course, applies to an utterance which expresses a desire or intention. Thus Davidson, like Wittgenstein, assigns a primary role to an interpretee's utterances of sentences which can serve to describe, and hence must accord with, the interpretee's preferences and choices; and since, in this framework, action constitutes a choice among alternatives, this means that the accord must hold over the interpretee's actions generally. So Davidson also assigns a primary role to what he treats as the interpretee's 'evaluative attitudes' to sentences, among which he includes desire and intention as well as preference. According to Davidson, 'it would not be wrong to say that the evaluative attitudes, and the actions that reveal them, form the foundation of our understanding of the speech and behaviour of others.'[41]

We saw that Wittgenstein emphasized what he took to be a fundamental and sententially described connection between an interpretee's utterances and other actions by stressing the pervasive role of commonsense action-related analogues of truth such as the fulfilment of intention. It now appears that Davidson in effect places a comparable emphasis, by re-describing these analogues as attitudes to the truth of sentences, and thus representing them in terms of truth itself. Indeed it seems that through this representation Davidsonian radical interpretation comes to embody an assumption of language-action regularity of much the same kind as Wittgenstein emphasized and linked with the notion of a linguistic rule. For if an agent is to be interpretable in accord with the principles Davidson espouses, then what the agent *says* which relates to preference, choice, and intention must systematically accord with what the agent *does* in actual instances of decision and action. This correlation, again, will be sentenial in the sense indicated above, for it will naturally be registered in the actual or potential use of the same sentences for describing both evaluative attitudes and actions in accord with them, and in both the language of the interpreter and that of the interpretee.

Such a sentential correlation, however, cannot be known to hold unless there is evidence that the interpretee actually knows how to act in accord with the sentences which are taken to characterize her evaluations and actions; and this evidence must surely come from action itself. (Thus Davidson stresses the basic role of both 'the evaluative attitudes' and 'the actions which reveal them'.) So it seems that a Davidsonian interpreter will regard it as a condition of intelligibility that an interpretee should *act* in accord with his or her sentences wherever appropriate, and should also know, for very many others, what it would be to do so; and it will be a condition of the successful application of the kind of theory that Davidson envisages that this should be shown to hold for the speakers to whom the theory applies.

This means that although Davidson's highly theoretical programme contrasts sharply with that of the later Wittgenstein, it would be quite wrong to argue, from a Wittgensteinian perspective, that Davidson is liable to criticisms such as Wittgenstein implicitly levels at Augustine as well as the *Tractatus* -- e.g., of failing to see that action and practice are more basic notions in the explanation of meaning than those of reference and truth. For Davidson's tacit assumption of sentential regularity as between utterance and action entails that what Wittgenstein called practice enters his account at a basic level. To act in accord with sentences is to act in accord with words, and this is also to engage in whatever rule-governed practices of using words and sentences there happen to be. So the requirement that agents be able to act in accord with their sentences wherever appropriate -- e.g., that a speaker know how to act so as to render 'x gets five red apples' true (Davidson), or again how to act on an intention articulated by 'x gets five red apples' (Wittgenstein) -- must tend to ensure that agents are competent in precisely those aspects of the practice of using words and sentences that Wittgenstein took to be illegitimately assimilated by a generalizing use of the concept of reference. There is thus a case for saying that the theoretically expanded and action-engaging role which Davidson assigns to truth actually serves ends

which Wittgenstein himself might (or ought to) have regarded as desirable. For this can be seen as enabling Davidson to integrate the general reference- and truth- based approach to meaning of the *Tractatus* with the emphasis upon the basic and endlessly diverse role of interpretable practice characteristic of the *Investigations*; and also to represent general facts about grammar, including those which Wittgenstein took to be required for interpretation, in an explicit and systematic way.[42]

Davidson also stresses the role of norms. What makes interpretation practicable, he says, is 'the structure the normative character of thought, desire, speech, and action imposes on correct attribution of attitudes to others, and hence on interpretation of their speech and explanations of their actions.'[43] On this point, however, we again find an apparent contrast with Wittgenstein. For although Davidson mentions the normative character of speech, he appears to distance himself from much that is said in a Wittgensteinian vein about rules or conventions of language. Thus, e.g., he urges that 'Conventions and rules do not explain language; language explains them', and stresses the way communication can be envisaged to proceed in the absence or abeyance of linguistic convention; and he goes so far as to say that most language learning 'is accomplished without learning or knowing any rules at all.' [44]

It is unclear how far Davidson here means to contrast his views with Wittgenstein's; but on the points we have been discussing the contrast is more apparent than real. For again, a requirement that agents act in accord with their sentences imposes a notion of accord between sentences and actions to which correctness or the lack of it is applicable, and which seems indistinguishable from that discussed in connection with Wittgenstein and rules above.[45] So it seems that for Davidson as for Wittgenstein interpretation requires a link between sentences and actions which deserves to be regarded as normative. Such sentence-action links will, moreover, be customary for Davidson in the same sense as for Wittgenstein: that is, they will require to be displayed with a degree of regularity which renders interpretation in accord with them possible. (And where communicators share knowledge of sentence-action connections which make their interpretation of one another possible, there will also be a case for regarding such connections as having the status of rules or conventions among them.)

Some final details for comparison are provided by recent essays in which Davidson touches explicitly upon the question of rule-following.[46] Thus in an account published after the present essay was nearly finished, Davidson says

> there certainly must be a way of distinguishing between correct and incorrect uses of a sentence... What is required, I think, is not that people speak alike, although that would serve. What is required, the basis on which the concepts of truth and objectivity depend for application, is a community of understanding, agreements among speakers on how each is to be understood. Such 'agreements' are nothing more than shared expectations: the hearer expects the speaker to go on as he did before; the speaker expects the hearer to go on as he did before. The frustration of these expectations means that someone has not gone on as before, that is, as the other expected. Given such a divergence there is no saying who is wrong; this must depend upon further developments or additional observers. But the joint expectations, and the possibility of their frustration, do give substance to the idea of the difference between right and wrong, and to the concept of objective truth. They therefore provide an answer to Wittgenstein's problem about 'following a rule'...
>
> For a speaker to follow a rule is, as I am interpreting it, for the speaker to go on as before; and this in turn means for the speaker to go on as his audience expects, and as the speaker intends his audience to expect. (A finer analysis must allow for cases in which the speaker goes on in the way the audience does not anticipate, but in which the audience nevertheless detects the anomaly as intended by the speaker)...

We saw that Wittgenstein's questions about the correct use of rules or sentences extend to the motives which sentences articulate, and so cannot be answered by reference to motives like intention and expectation alone. Although Davidson does not attend to this feature of Wittgenstein's dialectic, he

nonetheless invokes expectation and intention in the context of an account which satisfies Wittgenstein's implicit constraints, by attempting the interpretive specification of the content, and hence the normative status, of sentences and motives together. The expectations and intentions which figure in Davidson's account of the interpretation of sentences here will thus coincide closely with those involved in the interpretive regularities which Wittgenstein emphasizes via consideration of the interpreter-explorer in §206, §207, and §243; and taken in this way, the treatment of rule-following which Davidson suggests above coincides in essentials with that indicated by Wittgenstein in the passages with which we began. Both relate the imposition of norms to the kind of interpretive explanation which renders human speech and action logical and intelligible as a whole. Wittgenstein eschews theory, and specifies the evidential basis for such radical interpretation only in terms of regular connection between action and utterance; and this is an important, albeit tacit, aspect of Davidson's far more explicit and theoretical approach.[47] It thus appears that on the points about interpretation and rule following which we have been discussing our philosophers are in agreement. In particular, for both 'there is a way of grasping a rule [or sentence] which in *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" [or acting in accord with the sentence] and "going against it" in actual cases'; and for both this practical grasp, and our ability to recognise it in intepretation, are basic to our understanding of one another.

IX

Both Wittgenstein and Davidson advance a conception in which everyday interpretation rests upon a naturally based correlation between utterances and other behaviours, in which the utterances are such as to serve an interpreter for specifying the motives which render the correlated behaviours intelligible, and thereby also such as to embody the norms of truth and rationality in terms of which such behaviour is explained. As Davidson makes explicit, such a correlation seems best understood as one which holds between the effects of a set of common causes, where the causes are articulated motives. In these terms -- which are of course not Wittgenstein's -- the claim is that interpretation enables us to specify the causes of utterance and action by a kind of triangulation from effects to causes; and that the precision of such interpretive triangulation depends upon the extent to which the effects themselves admit of correlation.

This conception entails a close connection between the precision and congency of interretation and first-person authority. The notion of interpretation, or radical interpretation, presupposes an interpreter who has a degree of interpretive cogency, that is, who has the ability to frame hypothesis about desires, beliefs, and other motives which make sense of the behaviour of interpretees. The argument of §206-7 is that this ability can attain the especially impressive results which we find in our understanding of one another only if it is directed upon behaviour of a certain kind, that is, behaviour which the interpreter can understand as exhibiting regular connections between the interpretee's utterances and actions. Thus on Wittgenstein's account, such regular connections provide the key to interpretation which is precise, accurate, and fully grounded. Such regular connections appear at a maximum in behaviour which can be understood as the exercise of first-person authority. There is therefore a potential dialectic between the cogency of an interpreter and the first-person authority of an interpretee, which enables them to be mutually reinforcing, and also mutually explanatory.

We can bring this out by imagining that we are in the position of Wittgenstein's explorer/interpreter, tying to understand people strange to us by using our own **NL** to explain their behaviour (including utterances) as action in accord with **PR, D**, and **B**. For this we need to establish for their language an infinite correlation corresponding to our own **NL**, since such a correlation specifies the semantic beliefs which we would require to construe interpretees' utterances as intentional sayings and other speech acts in accord with **PR.** Let us call this langauge 'F' (for 'foreign'), and use '∑' to schematize foreign sentences. Then we can indicate the kind of (non-homophonic) linguistic correlation which we wish to specify as

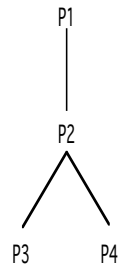
**FNL:** '∑' is true (in F) just if P

Thus schematically, if we take 'P1', 'P2', 'P3', etc., as sentences of our own language, and '∑1', '∑2', '∑3', etc., as sentences of the language we are aiming to understand, then we need data for hypotheses of the form:

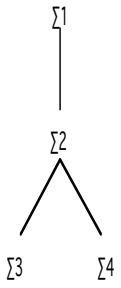
'∑1' is true (in F) just if P1
'∑2' is true (in F) just if P2
'∑3' is true (in F) just if P3
etc.

Given preliminary hypotheses of this kind as a starting point we may hope to begin to abstract out the semantic role of the words from which the F sentences are constructed, and so to frame a theory of truth for F, which will have as theorms instances of a correlation which suffices for interpretation.

§206-7 suggest that we can collect the required data if we can construe our interpretee's utterances as expressing desires and beliefs which we also assign in understanding their actions in accord with **PR**, **D** and **B**. To represent the situation more vividly and realistically, let us represent our applications of these principles by means of trees, in which we show an interpretee's hypothesized goals (desires) by sentences, and the structure imposed upon these by the interpretee's beliefs by lines connecting these sentences. Thus suppose we hold that an intepretee desires that P1, and believes that if P2 then P1, and in consequence desires P2; and also believes that if P3 and P4 (in that temporal order) then P2, and in consequence desires P3 and P4 (in that order) as well. This means that we ascribe a tree of goals as follows:




Here P1 - P4 specify desires or goals related by **PR**, and as the agent acts we will expect each of P1- P4 to be satisfied and pacified in an instance of **D**; thus if the agent performs P3 and P4 in that order, this should result in the satisfaction and pacification not only of those desires but also of P3, and this in turn in the satisfaction and pacification of P1, which was the agent's overall goal.[48] The idea that we can construe an interpretees' utterances as expressions of goals and beliefs, or as about the situations wanted or believed, is the idea that we can use these utterances to construct a parallel tree in the language of the interpretee (or part of such a tree), which we can represent as:



Such a tree constitutes an hypothesis both as to the interpretee's motives in acting and what he or she would say about them; so it also constitutes an hypothetical specification of a set of connections which hold between the interpretee's utterances and actions. Evidence in favour of this hypothesis, however, is also evidence for mapping the sentences in the trees, so that we have:

'∑1' is true (in F) just if P1
'∑2' is true (in F) just if P2
'∑3' is true (in F) just if P3
'∑4' is true (in F) just if P4

These are the kind of hypotheses required to begin the interpretation of F; and we may suppose that such comparison of trees enables us to start to formulate a theory of truth which yields a hypothesized **FNL** correlation with these instances. Then the process of relating the interpretee's utterances and actions to our own can be accelerated. Having set out a hypothetical tree in our own language for a particular non-verbal action on the part of an interpretee, we can use our theory of truth-conditions to construct a counterpart tree in F. Hence insofar as we can get an interpretee to draw upon first-person authority, and set out a tree for his or her own action, we can test our understanding of the interpretee's verbal and non-verbal behaviour together, by comparing these trees.

In this case interpretation yields predictions which are precise and powerful. For if both our initial tree and the theory by which we translate it are correct, then our translated tree should match that provided by the interpretee sentence for sentence. Failure at any point in such matching will give us reason to revise our initial tree, our translation, or our assumption as to the correctness of the tree provided by the interpretee. Success, on the other hand, should tend to confirm all three together. A match, that is, should raise our confidence in our initial tree, towards whatever level we associate with the interpretee's first-person authority in the case. Also, however, the same match should tend to confirm the interpretee's possession of this authority, by showing a correlation between self-ascription and the results of interpretation by another. Finally, the match offers support for the theory of truth -- the source of the **FNL** correlation which we are using -- with regard to all the sentences which figure in the tree. For the match indicates that our hypothetical correlation maps the interpretee's sentences on to the same actions and situations as the interpretee does in actual practice; and this, according to the argument of the preceeding sections, is the most basic arbiter of what the interpretee actually means.

It thus appears that insofar as we hold that this kind of match with the speech of an interpretee obtains for an interpreter's trees generally, we thereby hold (i) that the interpretee has a degree of well-founded first-person authority about her goals and beliefs; (ii) that the interpreter can aspire to a degree of confidence in ascribing these goals and beliefs which tends to approach that of the interpretee; and (iii) that the interpretation of any action on the part of the interpretee can be turned to the testing, and hence to the confirmation, of the interpreter's understanding of the interpretee's idiolect. For on the assumption of match the confirmation or support associated with each tree by which the interpreter understands an action of the interpretee stands to be inherited by the theory which yields the correlation which registers the interpreter's understanding of the interpretee's sentences. So systematic utterance-action regularity of the kind we have been considering can tend simulatneously to render interpretation cogent, first-person authority credible, and our interpretive grip on the meanings of sentences as firm as any we possess.

The idea in this is of course *not* that we simply assume that others can describe or express their states of mind correctly, and so by and large take their word for they think and feel. Taking another's word presupposes understanding it, and we are trying to cast light on how we understand the words of others in the first place. The idea, as taken from §206-7, is rather that we understand another's words by taking them to express states of mind which also serve also to explain other behaviour; and so we can understand another the better, the more he or she puts such motivating and explanatory states of mind into words. Hence also we need not assume that all the interpretee's verbal expressions show all the

features of first-person authority. Freud has made it plausible that there are circumstances -- including standard psychoanalytic free association -- in which interpretees regularly produce utterances which correlate with and serve to specify their own motives, but are unaware that they are doing so. These verbal expressions are used in the framing of hypotheses about motives for which the interpretee's utterances provide crucial specifying evidence, but in respect of which the interpretee's exercise of first-person authority may otherwise prove defective. Then further utterance may provide further evidence bearing on the ascription of these motives as the interpretee's authority is brought to bear in considering these hypotheses, considering further evidence which bears on them, and so on.[49]

Thus on the present account first-person authority is central to the success of our interpretive practice. Each of us, as interpreter, aims to map the sentences of his or her language onto *both* the utterances and non-verbal actions of an interpretee, and so onto the whole field of the interpretee's behaviour in its relations to the environment. So far as an interpretee has first-person authority, that interpretee can *also* map his or her own utterances with non-verbal actions, and thence also with the environment, so as to provide a comparison with the mapping which an interpreter hypothesizes. This provides data which make it possible for an interpreter's mapping to approach an interpretee's mapping, or to be co-ordinated with it, with constantly improving accuracy. Of course an interpreter cannot always interpret accurately, and an interpretee's first-person authority may fail. But an interpreter can still correct bad interpretations in light of the evidence manifest in the interpretee's actions, and also check and, if relevant, try to correct the interpretee's own account. This process allows of continual repetition and refinement. So the fact that each of us in *both* a potentially accurate interpreter *and* a potentially authoritative interpretee would appear to allow us to calibrate our interpretations of verbal and non-verbal behaviour continuously and cumulatively, and so as to give both something like the degree of precision and accuracy which we observe them to enjoy.

This indicates that within the conception advanced by Wittgenstein and Davidson we should regard precise and accurate interpretation and first-person authority as interlinked phenomena, each of which renders the other possible. (And this suggests that they might well have been made for each other by evolution.[50]) To be able to express desires, beliefs, and other attitudes, is to have the ability to supply part of the behaviour -- correlable utterance -- which accurate interpretation requires; and to have first-person authority is to be able to supply such utterance as and when required. In the conception we are considering, therefore, first-person authority does not seem solely or primarily a form of self-understanding. Rather it appears as a complement to the ability to interpret: it is the ability to manifest maximal interpretive regularity, and thereby to make oneself understood. If the preceeding argument is correct, then to see first-person authority in this further perspective is to see it rightly. For the argument indicates that our capacity to think and speak about ourselves is constituted as knowledge by a possible relation to others, and in particular by our being such as to be interpretable by them.

X

Davidson is one of a number of philosophers -- including Field, Lewis, Churchland, and Dennett -- who have recently sought to compare interpretation with empirical measurement.[51] These philosophers hold that just as we use numbers to keep track of comparisons and relations among objects in assigning weights to them, so we use utterances or sentences to keep track of comparisons and relations among states of mind. Something like this idea also seems an integral part of Wittgenstein's treatment of interpretation. We have already seen that in §242, just before introducing the notion of private language, Wittgenstein explicitly compares the way both intepretation and measurement alike require constancy in the results of employing standards. He makes a similar comparison earlier, just before the introduction of the language-game of writing numbers in accord with a rule; and again he refers to regularities in the behaviour which we take as expressing mental states.

> 142. It is only in normal cases that the use of a word is clearly prescribed; we know, are in no doubt, what to say in this or that case. The more abnormal the case, the more

> doubtful it becomes what we are to say. And if things were quite different from what they actually are -- if there were for instance no characteristic expression of pain, of fear, of joy; if rule became exception and exception rule; or if both became phenomena of roughly equal frequency --- This would make our normal language-games lose their point.-- The procedure of putting a lump of cheese on a balance and fixing the price by the turn of the scale would lose its point if it frequently happened for such lumps to grow or shrink for no obvious reason. This remark will become clearer when we discuss such things as the relation of expression to feeling, and similar topics.

Here the reference is to the measurement of weight in particular, rather than to measurement in general; but the relation to §242 is clear. This is another example of the way interpretation is tacitly considered in the *Investigations* prior to §206-7; and also of the connectedness in Wittgenstein's philosophical thinking generally. In remarks prior to §142 Wittgenstein discusses the way in which the rule-goverened use of a classificatory word depends upon the holding of 'normal' circumstances, taking for example the use of the word 'chair' (§80). Then in §142 he turns briefly to the use of words which describe mental states, and hence, implicitly, to the use of words and sentences in interpretation; and he compares the regularities upon which such use depends to those which sustain a practice of measurement. This apparently introduces into the discussion of interpretation in the *Investigations* a central idea of the *Tractatus* : that meaningful sentences -- sentences in their systematic normative connection with the range of possibilities which constitute logical space -- are the instruments by which we map the actual world; so that we can think of a proposition as 'laid out against reality like a measure' (2.1512). The remark therefore serves a double purpose: it compares interpretation to a process of measurement; and at the same time (as we have seen above) it is part of an explication of the normative relation of sentences and thoughts to reality, which the *Tractatus* use of the same analogy took for granted.

In light of the exposition above it seems that the metaphor of measurement can be regarded as having a relatively precise application in this case. Generalizing on §142, we can say that in measuring we employ empirical techniques, such as putting one object in balance with another (or against a scale calibrated to reflect such balancings) or, as in the example Wittgenstein most frequently uses, the laying one object against another to determine length[52] These techniques compare objects with one another, and thereby assign them to classes in respect of the relation used for comparison; and in this a particular set of objects, or a particular technique of comparing objects (a meter rod, the scale on a balance) may acquire the status of a norm or standard, by reference to which other objects are classified.

The techniques of comparison involved in a process of measurement must be capable of being applied regularly, consistently, and in harmony with one another. (For example rules like 1 mile = 5280 feet, or c = 186,000 miles per second, clearly presuppose that the results of measuring with small units, or by one technique, coincide with those of measuring by large units, or by other techniques.) This means that the objects being measured must relate in regular ways to the techniques and standards of measurement, and also that they must relate in regular ways to one another. The application of a technique of measurement thus characteristically presupposes that an extensive and systematic regularity holds over the items taken to share the property measured. (The ascription of weight rests upon the distribution of mass; and hence upon the vast system of spatio-temporal regularities described via the scientific use of that concept). The existence of such regularity is a contingent matter; we can imagine that objects might vary unexpectedly in the property being measured -- or again might fail to behave in the ways which underlie the ascription of this property -- so that no technique could determine a stable and useful classification. Such lack of regularity, as Witttgenstein remarks in §142, might render measuring of the kind we now practice pointless, or again impossible.

In such cases we could say, paraphrasing §207, that there was 'not enough regularity' for us to speak of the measurement of length, weight, or whatever; and the idea of §142, §207, and §242 is evidently that interpretive regularity is likewise necessary for the assignment of content, that is, as we may say, for the measurement of sense or meaning. As we may put the point to bring out the

comparison, the sentences of an interpreter's language which are used in framing interpretive hypotheses can be compared to measuring rods, in that they embody standards -- contents, normative relations to particular situations expressed in terms of truth -- with which episodes in an interpretee's behaviour must coincide in various ways to count as rational, logical, or intelligible, that is, as successful intentional actions, correct followings of rules, and so forth. Just as empirical measurement requires compatability among the results of laying instruments against objects in a variety of ways and circumstances, so interpretation requires compatability in the results of laying sentences against episodes in behaviour in a variety of ways and circumstances. If a rod is to be a standard embodiment of a length, and so be capable of measuring a spatial interval, it must be capable of being laid against other objects in a regular way; and hence both it and the objects which it is used to measure must share the systematic order in spatio-temporal behaviour which we chart by speaking of objects as rigid. Likewise if an sentence is to embody a meaning, and so to be capable of measuring (articulating) the mental, it must be capable of systematic comparison with speech and other behaviour; and so episodes in speech and behavior themselves must fit together as parts of a comparable overall regularity, which we chart by speaking of behaviour as rational. The required compatability includes the kind of fit between utterance and other behaviour mooted in §207 and §444-5, and that discussed in the last section.

Thus both philosophers compare the regularities which must hold over human behaviour for it to be interpretable with those which hold over physical things for them to be understood in terms of the measurement of length, weight, and other fundamental physical properties. This comparison enables us to sharpen our sense of what both mean in speaking of the 'foundation in agreement' which Davidson describes in terms of 'way of life', and which Wittgenstein says 'is not agreement in opinions, but in form of life.' It is of course true, and central to their common account, that people judge and speak, and hence tend follow rules, or use sentences, in similar ways. This, however, does not support to the idea that agreement as to how a rule should be followed itself *constitutes* correctness, i.e., that to use a sentence or follow a rule correctly is just to do so as others do, *however that might happen to be*. (This is the account frequently ascribed to Wittgenstein as the 'community view' of rule-following, which he articulates and rejectsin §241[53] via the question ''So you are saying that human agreement decides what is true and what is false?'; and indeed Davidson might have felt the need to add a similar caveat to the remarks quoted at the end of section VIII above.) In Wittgfenstein's terms, such agreement would be a form of agreement in opinions (i.e. in opinions as to how a sentence is to be acted in accord with at this or that point); and such opinions might be right, wrong, or inevaluable.

For Wittgenstein, as I think for Davidson, such agreement in opinion about norms should be seen the surface manifestation of a natural co-ordination among human begins, or order in human behaviour, which is the basis of normative accord, or, in Davidson's terms 'the basis on which the concepts of truth and objectivity depend for application'. Just as there is an order in spatio-temporal behaviour which constitutes objects as rigid and hence measurable, so there is an order in human behaviour which constitutes actions, including utterances, as intelligible and hence interpretable. On this view a successful action in accord with a rule or sentence is one part of a systematic concordance, which extends through the whole field of behavioural dispositions and relations to the environment upon which linguistic use, empirical and psychological judgment, and sensitivity and deference to norms all supervene. It is in this natural co-ordination with one another -- in which each speaks and acts in ways which others can interpret, and therefore in accord with the norms which each imposes via his or her language on others in making sense of what they say and do -- that our practices of judgment, and the phenomenon of normative accord which they exhibit, are to be regarded ultimately as based. On this account the objectivity of rule-following -- and the 'hardness of the logical must' -- is a reflection of a behavioural and cognitive harmony which goes deeper than judgment, and which renders it possible.[54]

XI

Both Wittgenstein and Davidson also use the analogy between interpretation and measurement to enforce the idea that language must be public.[55] Wittgenstein draws this conclusion as an explicit consequence of his claims about rule-following:

> 202. And hence also 'obeying a rule' is a practice. And to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately': otherwise thinking one was obeying a rule would be the same thing as obeying it.

Wittgenstein is here making use of a familiar form of distinction between subjective and objective. In general we distinguish between *it seems to such-and-such a thinker that P* and *it is the case that P*; and how far we regard a judgment or class of judgments as objective as opposed to merely subjective depends upon how this distinction is drawn. In these terms Wittgenstein's claim in §202 is that the following of rules or the application of concepts is also an objective matter. We are required, that is, to distinguish between *it seems to such-and-such a thinker that she is correctly following rule R (applying concept C)* and *it is the case that such and such a thinker is correctly following rule R (applying concept C).*

This requiremement clearly applies to judgment or concept-application quite generally: bringing any phenomenon under a rule or concept requires an application of that rule or concept which, as Wittgenstein stresses at §265, is *actually correct*; and hence one to which this has application. So, as Wittgenstein argues, this kind of objectivity must also pertain to judgments which might be regarded as 'private', namely, a person's judgements about his or her consciousness or subjectivity itself. Thus, as we may say, Wittgenstein is here concerned with the objectivity of the subjective: with what is required for judgments about the subjective to have objective validity, that is, to qualify as objective applications of words or concepts. His claim is just that if judgments about the subjective are to have such validity, the practice of making these judgments, like that of making others, must admit of interpretation. The claim is not epistemic but constitutive: it is that the *notion* of correctness as opposed to seeming correctness applicable to a particular person's practice of rule-following (or judgment) is that which would be applied in the course of finding the practice intelligible or logical, and hence in interpreting it. A purported practice which could not be found intelligible in this way would not be part of the order which constitutes the space of normative accord, and so would not be one to which this distinction had application.

As is familiar, Wittgenstein argues this point for the mental by reference to someone trying to keep a diary about a sensation which is private, in the sense that it 'can only be known to the person speaking'; and so giving himself an inner ostensive definition of a sign 'S' which is to stand for this sensation. Such imagined inner ostension, as Wittgenstein remarks, is supposed to be a vehicle of normativity, that is, to bring it about 'that I remember the connection *right* in the future.' But in such a case there is actually no criterion of correctness: the process of ostension itself can create no rule or norm against which further thought or behaviour on the part of the diarist might be assessed; so no norm yet covers the case. In consequence there is so far no determinate practice of judgment associated with 'S', and hence no determinate content to the diarist's supposed concept *sensation 'S'*. As Wittgenstein says: 'One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'. (§258) He emphasizes the point by imagining the diarist saying "Well, I believe that this is the sensation 'S' again", to which he replies 'Perhaps you *believe* that you believe it!' (§260). The diarist's claim that he believes that this is the sensation 'S' presupposes that he has rendered himself subject to a norm for use of a concept *sensation 'S'*; but this is what is in question.

Now of course Witttgenstein does not hold that we ordinarily need a criterion for the use of words for sensations or conscious states; as subjects we naturally use such words without criteria, but in ways which others can make sense of, and so find to be correct. The question arises only 'if we cut out human behaviour, which is the expression of sensation' (§288); for we thereby cut the first-person use of sensation words adrift from the order in behaviour which renders them intelligible. So the point here is

that same as Wittgenstein also makes about descriptions of the subjective more generally, in terms of the analogy with measurement.

> 278. “I know how the colour green looks to *me*” -- surely that makes sense! -- Certainly: what use of the proposition are you thinking of.
> 279. Imagine someone saying: “But I know how tall I am!” and laying his hand on top of his head to prove it.

A criterionless self-ascription counts as a measurement of a state of mind -- a genuine laying of a proposition against it -- only insofar as such an utterance enjoys the potential connection with further behaviour which consititues interpretive regularity. To try make such a description answerable to a subjective reality which is private is therefore to cease to use it in accord with a norm.

Wittgenstein has an explanation for our tendency to try to make the subjective a locus of pseudo--normative description of this kind. He urges that we are hostage to a 'picture' of the mental which 'forces itself upon us at every turn' (§425), and which can seem irresistable (§299). In this we represent the mind as an enclosed space, or sealed container, and ourselves as detecting the mental items which appear within this container on the basis of something like sight. Thus as he says

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

This is also a picture which we apply in our own case, as Wittgenstein illustrates by comparing our situation to one in which we

> ...Suppose everyone had a box with something in it: we can call it a 'beetle' No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle. -- Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing...(§293)

Wittgenstein points out that to take this metaphor concretely is to render our conceptualization of the mental unintelligible. The picture, as we may put it, represents my world as a world which is private; it converts my perspective on the world into a world which could be seen and known only by me. Such a world, however, could not be related to language, or to norms of description, at all. For on the one hand, as already observed, this conception cuts my self-ascriptive uses of mental descriptions adrift from the interpretive norms which enable others to make sense of them, and which make it possible to regard them as correct; and on the other, as is obvious, a private world could not be described in a public language -- that is, one in which speech was interpreted in accord with public norms -- at all. Since the box might contain anything or nothing, the supposedly private objects would drop out of consideration as irrelevant.[56] The picture represents our language for the subjective as uninterpretable and the phenomena of consciousness as indescribable, and hence, so far as communication is concerned, as scarcely better than nothing at all.

Of course Wittgenstein does not adhere to this conception. Rather he says that the phrases in terms of which he describes it -- as that of 'the sensation itself' which is represented as 'a something about which nothing could be said' and hence seems 'not a *nothing*, but not a *something* either'[57] -- constitute a paradox, which disappears only when we free ourselves of the idea that language always serves to convey Fregean thoughts, 'which may be about houses, pains, good and evil, or anything else

you please.' (§304) This emancipation, in turn, he links with gaining a better understanding of the picture itself. As he says:

> 423. *Certainly* all these things happen in you. -- And now all I ask is to understand the expression we use. -- The picture is there. And I am not disputing its validity in any particular case. -- Only I also want to understand the application of the picture.
> 424. The picture is *there*; and I do not dispute its *correctness*. But *what* is its application. Think of the picture of blindness as darkness in the head or the soul of the blind man.

A first thing to note about this picture is that it is utterly phantastic. The temptation to think of the mind as an enclosed space may, as Wittgenstein says, be constant and overwhelming; but the thought itself cannot be credited for a moment. There simply is no such enclosed space, no such impenetrable container, as this picture represents the mind as being. Yet despite this -- and his own satire in terms of the beetle in the box notwithstanding -- Wittgenstein thinks that the picture is not just misleading, but should also somehow be regarded as correct. Also Wittgenstein describes the picture as having a certain systematicity, as indicated in the example of blindness as darkness 'in the head or in the soul', or again in the idea that in thinking in this way of processes in the head 'we are not thinking of brain-processes, but of thought-processes.'

Attending to the systematic role of the picture enables us, I think, to see both what it is and how it can be at once correct and misleading. This picture, as we have seen, represents subjective or conscious events as objects of a kind of imaginary sight in a kind of imaginary container, rather than as events in the brain or nervous system. ('We are not thinking of brain-processes...') If we suppose that these events actually are bodily, neural events, then we are supposing that they do occur in us, hidden from the sight of others; so in this we can regard the picture as correct. But these are events which occur in the physical space internal to our bodies, and this is precisely what the picture obscures. Wittgenstein's 'picture' of 'head or soul', or of things 'in the head' which are not brain-processes but thought processes, should be seen as what Lakoff calls a cognitive metaphor;[58] and like other such metaphors, this one should be seen as having a genuine representational function.

In this case we may hypothesize that the 'picture' serves as a natural, preconscious, and prescientific way of representing the internal bodily processes which constitute or realize mental events. Accordingly, we should acknowledge that the image of the mental space or realm within, and the contrast between the inner and the outer which goes with it, are actually to be regarded as potentially misleading representations of the body, including the brain. It is to be expected that we should form representations of this kind, since we think in terms of such metaphors generally, and in doing so tend to assimilate all forms of awareness to sight. So the 'picture' here is a primitive representation of neural events, which represents them as other than they are. (Here, as perhaps with interpretation in the sections above, we are taking Wittgenstein's formulations somewhat further than he did; but the direction seems that in which they already point.)

Construing the picture of which Wittgenstein speaks in this way enables us to see the rest of his remarks as both correct and consistent with a Davidsonian account of the mental which he seems to have approached but did not formulate (see Appendix 1). Above we associated interpretive notions of cause-locating triangulation with both Wittgenstein and Davidson. If we bring these to bear in this case it seems that we will regard a pain -- 'the sensation itself' -- as a neural event, located by interpretation as the common cause of the self-ascriptions and other behaviour which, in any particular case, we take to manifest or express pain. A pain is thus both an event which occurs in a container which we do not penetrate in order to determine reference, and also a public physical event which engages with our norms and practices in using the word 'pain'. In imposing our natural *metaphor* for the containment of pain, however, we do *not* represent this event as it is, but instead construe it as something non-neural and private. And now even though we have in effect thought away the actual event which engages with our

descriptive norms and practices, we may still *think* we have a 'sensation itself' for which we might frame an ostensive definition even if we 'didn't have any natural expression for the sensation, but only had the sensation...' (§256) Conceived in terms of the metaphor the sensation also appears from outside as 'The thing in the box [which] has no place in the language game at all; not even as a *something* : for the box might even be empty...'. Yet of course we do not actually suppose that the head of a person in pain might actually be empty, or that the role of events in the brain in mediating the causes and effects of pain might somehow be cancelled out. Here we see the creation of a paradox. The public and bodily event which both self- and other-ascription serve to locate as the reference of 'pain' is represented as beyond the reach of language, and hence beyond the reach of thought, simply by being thought in terms of a metaphor for its bodily container.

In putting the matter this way, moreover, we evidently approximate to Wittgenstein's own view; for this seems precisely the lesson of some of his most compressed remarks:

> 296. "Yes, but there is *something* there all the same accompanying my cry of pain. And it is on acount of that that I utter it. And this something is what is important -- and frightful." -- Only whom are we informing of this? And on what occasion?
> 297. Of course, if water boils in a pot, steam comes out of the pot and also pictured steam comes out of the pictured pot. But what if one insisted on saying that there must also be something boiling in the picture of the pot?

It is not to be doubted that there are real events on account of which persons express pain, nor that such events are important and frightful. Also there is no problem in talking about these events, for like other of the basic objects of speech they tend to be causes of the utterances which describe them, although in this case, as it happens, causes which are inside our bodies. But once we think of these events as ones which we perceive in an enclosed space which is *not* the body, they come to seem indescribable and incommunicable. Things happen in our bodies which cause the verbal and other behaviour through which we express pain, as things happen in a boiling pot which cause the expression of steam. But in this case, remarkably, we form a picture of these internal events as internal to a space, and then insist that the events we picture are occurring, not (or not only) in our bodies, but in this pictured space. We insist that the internal events are occurring not in the pot, or not only in the pot, but in the picture of the pot.[59]

Wittgenstein takes this metaphor to be the source of a conviction that when we have pains, say, we *recognise them as such* (as they appear to us in the inner realm, and not as bodily or neural events); and despite his critique this conviction seems to have lost little of its hold on philosophers. He seems to have thought that once we take a sensation word as used in accord with a bodily correlate the notion of inner recognition simply drops out of account as irrelevant;[60] and the notion does seem unjustified apart from our determination to think in terms of it. In thinking this way, however, we impose a whole series of categories at once, taking pain as an object from which we derive its description via a concept applied within, and hence which appears under an inner mode of presentation, manifest in a place which is not public space. (Hence Wittgenstein's reference to the model of 'object and designation' in §293, and to Frege in §304) This is 'the decisive move in the conjuring trick', for in this we espouse the 'yet uncomprehended process in the yet unexplored medium' of consciousness (§308), and hence the feeling of 'an unbridgeable gulf between consciousness and brain-process' (§412) which goes with it.[61] This is an account of the source of the problem of consciousness which seems both naturalistic and plausible. It may be that Wittgenstein was wrong, and that the notion of introspective recognition which he took to give rise to the problem is actually legitimate, or again that the problem has some different origin. This, however, has not been shown.

A final point may be worth mentioning. It is in Wittgenstein's discussion of consciousness that his method most closely approaches that of Freud, as he hints in saying at §255 that a philosopher treats a problem as does a physician. His idea is that philosophical problems quite commonly arise as a consequence of representations of which we are unaware. (Cf §112: 'A similie that has been absorbed into

the forms of our language produces a false appearance, and this disquiets us...'). In order to dissipate the hold of such representations, we must bring them to the surface and examine their working; and this is what he attempts in his discussion of private language (§374). In this case, however, the unconscious image is that of the interiority of the body; and this makes the link with Freud deeper than Wittgenstein will have realized. For according to Freud's conception of the bodily ego, the image of the body pervades our unconscous representations of mental processes generally, yielding what Richard Wollheim calls 'the corporealization of thought.'[62] Thus for example the rejection or projection of something is characteristically represented unconsciously in terms of expelling something from a container which represents the body, acceptance or incorporation in terms of taking something into such a container, and so forth; so that unconsciously we constantly model our states of our minds and relations to others in terms of various containers which are also, in one way or another, symbols of the body. Thus the cognitive metaphor central to Wittgenstein's discussion of consciousness lies at the core of our apprehension of the mind generally; and Wittgenstein attempted to analyse the working of this metaphor in a place Freud did not, namely our conception of consciousness itself.

---

## Appendix 1: Wittgenstein and Anomalous Monism

In *Zettel*[63] §608 and following Wittgenstein compares structures in the brain which give rise to the mental phenomena of speech and writing with seeds which give rise to plants of various kinds. He clearly regards these structures as physical, and as causes of thought and behaviour, as well as effects of learning and the like; so he here holds (i) that physical events (or structures) cause mental events, and vice-versa. Also he explicitly assumes (ii) that causality, even in the brain, has a nomological character. He specifies that the seed-structures he considers give rise to behaviour in accord with laws, that is, that 'a seed always produces a plant of the same kind as that from which it was produced...' (§608); and he stresses that there might be 'a natural law connecting a starting point and a finishing state of a system, but not covering the intermediate states' (§613). Finally, his main concern in these remarks is to affirm a version of (iii) the anomalism of the mental. He urges in §608 and a number of the remarks which follow that there is a no set of lawlike psycho-physical connections, and again no form of linguistic encoding in the brain, which would make it possible 'to read thought-processes off from brain processes' (§608). He criticizes the notion of 'psycho-physical parallelism' involved in this as 'a fruit of a primitive interpretation of our concepts' (§611), and urges that speech need not be conceived as the rendering of information stored in the nervous system in the form of 'a translation with another symbolism.' (§612)

These remarks thus show Wittgenstein's adherence to the three seemingly inconsistent principles -- of interaction, nomologicality, and anomalism respectively -- which Davidson discusses and reconciles in 'Mental Events'.[64] Wittgenstein himself, moreover, seems clearly to regard the principles as inconsistent. He assumes, that is, that if he is to affirm his versions of (i) and (iii) together, then he must deny something ordinarily encompassed in (ii). So in these remarks he seeks to deny that the brain exhibits what he calls *causal efficacy* (§613). A physical structure or mechanism has this kind of causal efficacy, it seems, insofar as the effects which the mechanism produces can be explained by reference to its intrinsic physical features together with the laws in accord with which it operates.

Wittgenstein sees that if mechanisms have such efficacy then they will be subject to a further principle, which we may label (ii)*: that mechanisms which are intrinsically alike must operate in like ways in like circumstances. This is a version of the maxim that like causes have like effects, and here is a direct consequence of the notion that the mechanisms in question are alike in respect of the features by reference to which their output can be fully explained in accord with laws. Hence Wittgenstein seeks to deny that brain mechanisms have causal efficacy by denying that they are subject to principle (ii)*; and he presents his seeds as providing an alternative model which is both metaphysically possible and empirically plausible. The seeds he imagines are supposed to be intrinsically alike, but to produce different effects, and in accord with laws; for the laws to which they are subject relate to their history rather than their physical structure. Hence, as he puts the point, 'nothing in the seed [neural mechanism] corresponds to

the plant [thought and behaviour] which comes from it; so that it is impossible to infer the properties or structure of the plant [the properties or struture of thought and intentional behaviour] from those of the seed [underlying neural mechanism] ...this can only be done from the history of the seed [the history of the mechanism]' (§608).

Thus in these remarks Wittgenstein manages to find an account which allows him both to affirm anomalism with respect to the brain, and also avoid contradiction, by holding that while the output of the brain, like that of the seeds, is subject to law, it is not fully explicable by reference to the physical structure of the brain, as opposed, say, to its history. So he can hold that this output 'might come into being out of something quite amorphous, as it were causelessly' and urges that 'there is no reason why this should not really hold for our thoughts, and hence for our talking and writing.' (§608). His 'as it were' here is important, for in fact he is still ascribing a causal and nomological role to the brain, comparable to that of the imaginary seeds; it is just that this role has been so qualified -- as it were, by putting history in place of intrinsic physical features -- as to render the contribution of the mechanisms of the brain ambiguous, and hence not necessary to the detail of the effects they produce.

This line of thought combines insight with confusion. On the one hand, Wittgenstein seems right to insist on the anomalism of the mental, and also to hold that the mechanisms in the brain which give rise to the systematic production of symbols in speech or writing need not themselves be thought of as embodying some parallel symbolic form. (This latter point seems borne out by connectionist research.) On the other, he surely errs in suggesting that there is reason to expect that intrinsically similar neural mechanisms may well work like his imaginary seeds, that is, so as to produce output in systematically different ways. This is a scientific question, and it seems that all investigations have tended to support the claim that like mechanisms do behave, or tend to behave, in like ways in like circumstances. No mechanisms like the imagined seeds have ever been found.

It thus seems that Wittgenstein had no reason to deny (ii)* besides his version of (iii) and his sense that this would otherwise be inconsistent with (i) and (ii). Hence his arguments on this point are forced, and the line he constructs is artificial and *ad hoc*. He did not include these arguments in the *Investigations*, and it is hard to suppose that he would have felt impelled to frame them if he had seen that there was a simpler way foreward. For once linguistic confusions (e.g between the role of types and tokens) are cleared away, it becomes plain that an absence of strict law in the application of psychological concepts or vocabulary (in the mental language-game) is consistent with causal lawfulness described in terms of the physical. Davidson's account of this matter seems in better accord with the general lines of Wittgenstein's philosophy than Wittgenstein's own. In particular, his resolution of the seeming inconsistency provides a clear example of Wittgenstein's claim that philosophical confusions can be resolved through understanding the working of the language which gives rise to them.

---

## Appendix 2: Davidson on Measurement and Indeterminacy

Davidson also uses the analogy with measurement to argue that the indeterminacy of interpretation is consistent with the precision and objectivity of content. Here his claim is that just as we can keep track of the same facts about weights using different numbers if we measure them in carats as opposed to grams, or the same facts about temperatures in Farenheit or Centegrade, so we can keep track of the same facts about mind and meaning while 'measuring' them using different words (different words or sentences of the interpreter's language). As he says,

> We know there is no contradiction between the temperature of air being 32 farenheit and 0 celsius; there is nothing in this "relativism" to show that the properties being measured are not "real". Curiously, though, this conclusion has repeatedly been drawn. John Searle, for example, finds it incomprehensible that either of two quite different

> interpretations might correctly be put on the same thought (or utterance) of a person. Yet in light of the considerations put foreward here, this comes to no more than the recognition that more than one set of one person's utterances might be equally successful in capturing the contents of someone else's thoughts or speech. Just as numbers can capture all the empirically significant differences among weights or temperatures in infinitely different ways, so one person's utterances can capture all the significant features of another person's thought and speech in different ways. This fact does not challenge the "reality" of the facts or meanings thus variously reported...does not suggest that the states of mind of the speaker or thinker thus captured are somehow vague or unreal.[65]

Davidson's reference to 'two quite different interpretations' makes this a challenging claim, and one which may well seem counter-intuitive.[66] For we are surely inclined, as one might say, to suppose that a particular sentence correctly used in interpretation gives the essence of a thought or meaning. It seems that if I hold that 'P1' and 'P2' are non-synonymous sentences, then I cannot also hold that someone else's non-vague thought or utterance is really equally well interpreted by either. Indecision on this point, it seems, must indicate that I am not understanding the non-vague thought of the other fully or correctly, just as the comparable indecision the other's part would indicate that the other had not rightly understood me.

This is a strong intuition; but on the analogy with measurement, as Davidson points out, it is simply mistaken. 'P1' may be best on one scheme of interpretation (one scale of measurement), and 'P2' on another; and this possibility of multiple description no more impugns the precision and reality of the content thus ascribed than in the case of weight or temperature. The different indices assign the same content (weight, temperature), but via a different scale; and either scale represents the phenomena with equal empirical adequacy. But then if this is so why are we so unaware of the limited precision which our thoughts and utterances possess?

It is of course not enough simply to confront our intuitions about the definiteness of what we mean with the analogy of measurement; for these intuitions might as well be used to argue against the analogy itself. Davidson thus supports his claim by two further lines of argument. The first flows from his conception of the inscrutability of reference in the context of a theory of truth, the second from the bipartite nature of a theory of content.

In arguing the inscrutability of reference Davidson observes that if we have a theory of truth for a particular language, then we can always produce another theory which is empirically equivalent to the first by exploiting a permutation of the universe, that is, some one-to-one mapping ø of each object x in the universe on to another object ø(x):

> If we have a satisfactory scheme of reference for a language that speaks of this universe we can produce another...whenever, on the first scheme, a name refers to an object x, on the second scheme it refers to ø(x); whenever, on the first scheme, a predicate refers to (is true of) each thing x such that Fx, on the second scheme it refers to each thing x such that Føx...Here is a simple illustration. Suppose every object has one and only one shadow. Then we may take the ø to be expressed by the words 'the shadow of'. On a first theory, we take the name 'Wilt' to refer to Wilt and the predicate 'is tall' to refer to tall things; on the second theory we take 'Wilt' to refer to the shadow of Wilt, and 'is tall' to be true of the shadows of tall things. The first theory tells us that 'Wilt is tall' is true if and only if Wilt is tall; the second theory tells us that 'Wilt is tall' is true if and only if the shadow of Wilt is the shadow of a tall thing. The truth conditions are clearly equivalent. If one does not mind speaking of facts, one might say that the same fact makes the sentence true in both cases.[67]

Even if we accept that such permutations yield theories specifying equivalent truth-conditions in different ways, this does not mean that the theories are equivalent in every empirically relevant respect. On Davidson's account they are not equivelent *in use*, that is, equivalent from perspective of an actual or radical interpreter. Davidson stresses that 'the events and objects that cause a belief also determine the contents of that belief,' so that an interpreter must 'correlate his own responses and those of the speaker by reference to the mutually salient causes in the world of which they speak.'[68] Wilt will be a mutually salient cause both of utterances in which 'Wilt' figures and also of the beliefs and desires which an interpreter takes such utterances as expressing; Wilt's shadow will not. So an interpreter who proceeds as Davidson says will with good reason correlate responses by reference to Wilt, and so to link utterances of 'Wilt' with Wilt, and not the shadow. The fact that this correlation might be subjected to permutation does not alter its status as naturally preeminent in the empirical practice of interpretation. So even if we accept that permutation enables us to provide 'two quite different interpretations' of a single utterence -- and even if one takes this to apply to one's own utterances -- this is not a claim about interpretation as an ongoing empirical practice. Nor, in this respect, are the different interpretations comparable to assignments in Farenhiet and Celsius: the practice of measuring temperature does not itself render one scale central.

Davidson also stresses that "No causal theory, nor any other 'physicalistic' analysis of reference, will affect our argument for the inscrutability of reference...For the constraints on the relations between reference and causality (or whatever) can always be equivalently captured by alternative ways of matching up words and objects." Thus

> ...suppose, as before, that ø is a permutation of the universe, and that Cx,y is an appropriate causal relation between a word and an object. One good theory says that 'Wilt' refers to Wilt only if C 'Wilt', Wilt...while another empirically indistinguishable theory says that 'Wilt' refers to ø(Wilt) only if C('Wilt', ø(Wilt))...

This, however, seems liable to the same answer as above. The relation C which holds between 'Wilt' and Wilt's shadow will not be the causal relation which Davidson describes as central to the empirical practice of interpretation, but an artefact defined in terms of this relation. Theories cast in terms of such artefacts remain distinguished from that taken in terms of the causal relation itself.

Davidson has a further basis for his claims about interpretation, and therewith for his use of the analogy with measurement, in his bipartite account of content. As he says:

> There are often cases, I believe with Quine, where the totality of relevant evidence in a person's behaviour is equally well handled by each of two theories of truth, provided we make compensating adjustments in our theories of his belief and other attitudes...[69]

An interpreter assign contents to an interpretee's sentences and attitudes via what we can regard as two distinct theories: a theory of truth for the sentences, and a decision theory for the attitudes. These two theories will be answerable to the totality of relevant evidence jointly. There will, however, be many utterances which the interpreter regards as false, and which can be explained equally well in either of two different ways: by assigning to the interpretee a belief which the interpreter thinks false, or a pattern of linguistic usage (concept) which the interpreter does not share. Different choices in such cases may yield differing truth-theory/decision-theory pairs, that is, pairs in which the interpreter uses different sentences of his language to characterize particular sentences of the interpretee's language, or particular nodes in trees which explain the interpretee's actions. These differing theory-pairs will consititue schemes of interpretation which are equivalent in overall empirical significance; for by hypothesis they explain both everything relevant to the assignment of content, and they explain it equally well. Davidson thus seems to be illustrating the source of such indeterminacy in the following passage:

> If you see a ketch sailing by and you companion says 'look at that handsome yawl', you may be faced with a problem of interpretation. One natural possibility is that your friend has mistaken a ketch for a yawl, and has formed a false belief. But if his vision is good and his line of sight favourable it is even more plausible that he does not use the word 'yawl' quite as you do, and has no mistake at all about the position of the jigger on the passing yatch. We do this sort of off the cuff interpretation all the time, deciding in favour of reinterpretation of words in order to preserve a reasonable theory of belief. As philosophers we are peculiarly tolerant of systematic malapropism, and practiced at interpreting the result. The process is that of constructing a viable theory of belief and meaning from sentences held true.[70]

This argument does lodge the indeterminacy which it explicates within interpretive practice; but it is not clear that it renders indeterminacy analogous to choice of scale. The account appears to turn on lack of agreement between interpreter and interpretee, and this is consistent with Davidson's use of charity to eliminate indeterminacy elsewhere.[71] But then so far as an interpreter's uses and opinions coincide under interpretation with those of an interpretee, the interpreter can have no reason *of this kind* to hold that a non-vague thought or utterance on the part of the interpretee can be equally well described in incompatible ways. But if indeterminacy is really just a matter of scale, it is puzzling that it should be related to disagreement in this way: differences of scale, it seems, should hold and show in all cases alike.

Further, the analogy with measurement is supposed to show that utterances or states of mind which can be assigned more than one content-sentence are not thereby rendered 'somehow vague or unreal.' But insofar as interpretive precision depends upon agreement, it seems that indeterminacy resulting from ostensibly false utterance might best be regarded as a form of vagueness consequent upon the lack of agreement which such utterance reveals. Such lack of agreement, it seems, may be about the use of words, or about the world, or about both. But then there is a further point: an interpreter faced with such lack of agreement can neither dismiss the possibility that there is something to be learned from the interpretee, nor be certain of being able to learn it. So it may also be that indeterminacy which begins where agreement ends is not so much an effect of scale of measurement as an index of failure to understand.

We can make this idea more precise by returning to Wittgenstein's metaphor[72], in which a radical interpreter who as it were constantly lays her own sentences against the utterances and actions of an interpretee is comparable to someone who uses measuring rods to survey a space. (In terms of this comparison the interpreter is trying to take the measure of the sentences by which the interpretee takes the measure of the world.) We can treat spatial measurement as establishing a set of coincidence relations, as among measuring rods and spatial objects and intervals generally. As is familiar, we can describe such coincidence relations in terms of measurement and geometry in more than one way. In particular, the same underlying relations will yield one set of measurements and one geometry if we treat the interval realized by a standard rod as everywhere the same, and another set of measurements and another geometry if we take the interval to vary with the position and orientation of the rod. These differing sets of measurements will represent different sets of intervals as congruent or equivalent; and with one system of equivalences the geometry may be Euclidean, and with another, non-Euclidean.

Since the two geometric descriptions report and summarize the obtaining of the same set of coincidence relations among rods, objects, and intervals, they can be regarded as empirically equivalent. We can think of the differences in the geometric properties and relations assigned in one representation as systematically offset by differences in the measurements assigned in the other. So the establishing of a geometry, in this account, involves an ultimate empirical 'cancelling out' of what might have seemed to be empirically significant differences, which is comparable to that which figures in Davidson's discussion above.[73]

This enables us to gain a further perspective on the second kind of indeterminacy which Davidson considers. Let us imagine for the sake of argument that we have two agents, A and B, both of whom are realistic and unconfused, but who have (some) different concepts, as shown in differing patterns of linguistic usage. To make the situation vivid let us suppose that A and B use the same words (sound-patterns), but in different ways. Thus although they inhabit the same world and describe it by the same sounds, they partly see and map their common world differently, in the sense that some of their words and concepts determine different classes of objects, and hence types of situations, as equivalent. In this they are like surveyers who measure the same objects and intervals with rods which are physically indistinguishable, but who make use of different geometries, and so assign the same things systematically differing lengths, shapes, etc.

In such a situation it seems that indeterminacy will arise from differences which are conceptual but not ontological, and in proportion to the number and extent of such differences as there are. Thus the situation which Davidson discusses above with 'ketch' and 'yawl' might arise and be dealt with as he describes. Taking the matter in this way, however, neither A nor B will be able to map their utterances so as to represent the other as realistic and rational overall; and each will be faced with a range of choices as to whether to take the other as mistaken about facts or as deviant in use of language. (Likewise each of two surveyers might think the other mistaken as regards the size of certain objects and intervals -- wrongly equating unequals in some cases, and wrongly distinguishing equals in others -- and choose between ascribing mistakes in measurement or reinterpreting the signs or procedures by which measurement is conducted.)

In this kind of case it would clearly be inappropriate to understand indeterminacy solely in terms of scale of measurement. Each interpreter would be wrongly seeing *difference from* the other as a form of *defect in* the other, which defect could be distributed in varying ways as between empirical error and linguistic or conceptual inadequacy. But the differences which each was seeing in this way would be neither defects nor genuine disagreements; they would be ways of representing things which, although not the same, were equally satisfactory, and which a case could be made for regarding as empirically equivalent overall. In such a situation indeterminacy would be a mark of failure to understand, and a sign that the interpreter's account of the interpretee was genuinely imprecise. An interpreter who thought of this as an effect of scale alone would in effect be assuming, in Wittgenstein's phrase, 'that certain concepts [the interpreter's own] are absolutely the correct ones, and that having different ones would mean not realizing something that we realize.' To get things right each interpreter would need, again to quote from the same passage, to make 'the formation of concepts different from the usual ones intelligible'.[74] This would make it possible to obtain the overall agreement which would render interpretation precise, and allow each better to understand both the other and the world.

---

[1] Oxford: Basil Blackwell, 1958. Further quotations from this work will cited by the number of the remark. I have altered Anscombe's translation of §205 by replacing 'queer' with 'remarkable'.

[2] The same theme is also continued in §208, where Wittgenstein considers elucidating 'order' and 'rule' by means of 'regularity'. Again he apparently has in mind the kind of interpretive sign-action regularity discussed in the this section.

[3] Wittgenstein evidently means his citation of the monologue people to block what has come to be called a 'communicational practice' interpretation of his remarks on rule following. Still something of this view remains, in the fact that Wittgenstein holds that interpretation requires a measure of agreement between interpreter and interpretee. The interpreter of §243 must agree in judgments with

those whose language he translates, and the interpreter of a language used for communication among persons must find agreement in judgments both among them and between them and himself. Thus Kripke's statement that if we understand a solitary speaker we take that speaker 'into our community' can be seen as partly correct.

[4] Although I was not aware of the connection, the interpretive regularities with which Wittgenstein is concerned are closely related to those which I took as providing a foundation for the cogency of commonsense psychology, and by extension for that of psychoanalysis, in my introduction to Wollheim and Hopkins, eds., *Philosophical Essays on Freud* (Cambridge: Cambridge University Press, 1982; cf. pp viiff and xviff); this is partly reprinted, together with another essay, 'Irrationality, Interpretation, and Division' in C. & G MacDonald, eds, *Philosophy of Psychology* (Oxford: Blackwell, 1994), which re-presents these ideas in relation to Freud, and also to Davidson's discussions of irrationality and the division of the mind, in the form adopted in section IX below. Further application to psychoanalytic theory is contained in 'The Unconscious' in S. Guttenplan, ed, *A Companion to the Philosphy of Mind* (Oxford: Blackwell, 1994); 'Patterns of Interpretation: Language, Action, and Dream' in L. Marcus, ed., *Cultural Documents: The Interpretation of Dreams*, (Manchester: Manchester University Press, forthcoming 1995); and 'Wittgenstein, Interpretation, and the Foundations of Psychoanalysis', forthcoming in *New Formations: Special Issue on Psyconanalysis and Culture*, Summer 1995.

[5] For ease of exposition I use quotation marks instead of a variety of more serious devices for referring to sentences. Phrase markers would be required to do the whole job thus swept out of sight.

[6] I am inclined to think that Wittgenstein here calls expectation, thought, and the like 'interpretation of signs' because at the time he wrote this passage he held that in taking it that 'P' expresses my thought I thereby interpret 'P' -- as Wittgenstein says in the *Tractatus*, I thereby 'use the perceptible sign of a proposition (spoken or written) as a projection of a possible situation', that is, of the situation about which I am thinking. (*Tractatus Logico-Philosophicus*, London: Routledge, 1961, 3.11.) This would be a version of the view which he rejects at *Investigations* §201, discussed below.

[7] Oxford: Basil Blackwell, 1967, p. 70. The 'multiplicity' Wittgenstein speaks of here has two aspects: as the sentence which expresses the thought is part of a system of sentences, so the thought is part of a system of thoughts; and as the sentence (and the system of which it is a part) relates to reality, so does the thought and the system of which it is a part. Both ideas are present, e.g., in the

*Tractatus* idea that both sentences and thoughts are representations (models) which have the same 'multiplicity' as what they represent.

[8] The considerable role played in interpretation by what might seem to be merely syntactic considerations is discussed in my 'Psychoanalysis, Interpretation, and Science' in Hopkins and Savile eds, *Psychoanalysis, Mind and Art: Perspectives on Richard Wollheim* (Oxford: Blackwell Publishers, 1993.)

[9] Davidson, 'Expressing Evaluations', *The Lindley Lecture*, University of Kansas, 1982, p 3.

[10] Wittgenstein elsewhere emphasizes the importance of general connections which underlie our use of concepts, and which we may remain unaware of, partly because of their very generality. See for example Investigations, II, xii.

[11] Thus actual interpreters are not constrained to base their hypotheses upon the single observable notion of preference, nor might they be supposed to find sentential connectives by the method which Davidson outlines. The specifics of his approach, as Davidson says, are 'not...meant to throw any direct light on how in real life we come to understand one another, nor on how we master our first concepts and first language.' Rather his is 'a conceptual exercise aimed at revealing the dependencies among out basic propositional attitudes at a level fundamental enough to avoid the assumption that we can come to grasp them -- or intelligibly attribute them to others -- one at a time. Performing the exercise has required showing how it is in principle possible to arrive at all of them at once. Showing this amounts to presenting an informal proof that we have endowed thought, desire, and speech with a structure that makes interpertation possible.' ('The Structure and Content of Truth', *The Journal of Philosophy*, 1990, p. 325).

If this way of trying to cast light on meaning seems unacceptably indirect, it is worth bearing in mind that it may be the only one which is possible. Thus it might be a fact, for example, that evolution has built our brains alike, and so that under the impact of experience they all tend to home in on more or less the same assignments of content, in terms of which we naturally understand speech and other action. Also it might be that we could have no other way of understanding one another -- that the problem of assigning appropriate contents on the basis of behaviour and context was otherwise intellectually intractable. Still it would remain that one way of understanding the phenomena of content, and also of understanding what it was that our brains were doing, would be to try to construct a theory which would enable someone who knew it to perform something of the content-assigning job which the brain performs naturally.

These same reservations clearly apply also to the idea that a theory which suffices for interpretation makes explicit tacit knowledge which is actually used in interpretive practice. It may, but it need not. In the *Tractatus* Wittgenstein

seems to have taken the line that human beings *embody* the abilities in terms of which we make understanding and meaning explicit, and that this kind of embodiment need not be understood as a form knowledge in the ordinary sense. As he 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 part of the human organism and is no less complicated than it.

[12] The sketch of this kind of view which follows is meant only to be illustrative. For a recent discussion of concepts which has a number of Kantian features see Peacocke, *A Study of Concepts* London: MIT Press, 1992.

[13] It is not part of this conception that we think of ourselves as passing from judgments which we make about 'raw behaviour', conceived in terms distinct from those about meaning and motive, to judgments about intentional action. We can regard the input to our interpretive faculties as raw -- e.g. as consisting in the visible displacements of body surfaces, audible productions of sound, etc., which our senses register -- without prejudice as to how our faculties represent this input in the course of yielding these judgments. Precisely what the actual input is, and how it is registered, is a further theoretical question. 'Obervable', as used in this connection, means something like 'capable of making an impact on our cognitive faculties via the mechanisms of the senses', and so includes, e.g. what we observe unconsciously, subliminally, etc.

[14] See for example Higgengbotham 'Linguisic Theory and Davidson's Programme in Semantics' in LePore, ed, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* , Oxford: Blackwell Publishers, 1986; Neale, 'Logical Form and LF' in c. Otero, ed, *Chomsky: Critical Assessments*; Cambridge: MIT Press, 1993 p. 789-838; Segal, 'Priorities in the Philosophy of Thought' *The Aristotelian Society Supplementary Volume* LXVII, 1994, p 107- 130; and Larson and Segal *Knowledge of Meaning: An iIntroduction to Semantic Theory* Cambride: MIT Press, forthcoming.

[15] *Inquiries into Truth and Interpretation*, Oxford, Clarendon Press, 1984, p 22.

[16] *Inquiries*, p185.

[17] *Inguiries*, p197.

[18] 'Expressing Evaluations', p 20.

[19] Kripke's *Wittgenstein on Rules and Private Language* (Oxford: Blackwell, 1984) provides a pioneering and powerful account of the problems discussed in this section. The text makes it clear, I think, that these problems have not been met by the solutions proposed in the literature which has resulted from Kripke's exposition. (On this point see also Boghossian 'The Rule Following Considerations' *Mind*, October 1989. For further discussion of Kripke's account see McGinn, *Wittgenstein on Meaning*, Oxford: Blackwell 1984; Baker and Hacker, *Scepticism, Rules and Language*, Oxford: Blackwell 1984; and Budd, *Wittgenstein's Philosophy of Psychology*, London: Routledge, 1989))

Although the account here differs from Kripke's in many respects, a number of features are carried over. Thus for example Kripke describes Wittgenstein's solution to the questions he raises as follows: 'The solution turns on the idea that each person who claims to be following a rule can be checked by others. Others in the community can check whether the putative rule follower is or is not giving praticular responses that they endorse, that agree with their own.' (p. 101) On the present account the first sentence of Kripke's statement remains true. The difference §206-7 shows is that Wittgenstein regards checking whether someone is following a rule as part seeking to understand the putative rule-follower's linguistic and non-linguistic behaviour overall, in such a way as to intepret the former by finding regular connections between it and the 'frame of reference' provided by understanding of the latter. This does entail, as Wittgenstein stresses, that interpreters will find agreement between themselves and someone whom they take to be following a rule. But Wittgensetin's account turns upon a substantial and empirical notion of interpretation, and the consequences of this as regards agreement, rather than upon agreement (or agreement in responses) itself. This difference is spelt out more fully in section XI below.

[20] This can be seen a further version of the interdependence of our notions of proposition and propositional attitude noted in connection with Quine and Davidson in section II above; and this also makes clear that the 'Gricean' solutions proposed by many commentators, or others cast in terms of the notion of intention, do not address the questions Wittgenstein raises. The point can be put by saying that Wittgenstein's questions concern the source of our first-authority as to *both* how we should follow a rule, and the contents of our propositional attitudes like belief and intention; such questions are not answered, but begged, by citing again our knowledge of these contents (= conditions of satisfaction, etc.), as they figure, e.g., in the intentions with which we use sentences.

[21] Wittgenstein evidently retains what Davidson calls 'our pre-Fregean semantic innocence' (*Inquiries*, p 108.) Davidson's paratactic theory attempts to make this

good. For a recent discussion see Ian Rumfitt 'Content and Context: The Paratactic Theory Revisited and Revised', *Mind*, July 1993.

[22] Wittgenstein's emphasis in §198 upon seeing the sign-post with its associated training and practice both in terms of causes and of interpretable practice seems to anticipate Davidson, in the sense that Wittgenstein apparently assumes that the language-game (vocabulary) of cause and effect, and that of rule, intention, and practice, will here apply to the same objects and events, and in harmony. (The one language-game doesn't, as it were, tend to crowd the other out.) This seems suggested in other remarks as well. Thus in §493 Wittgenstein asks whether 'if it were shown how the words "Come to me" act on the person addressed, so that finally, given certain conditions, the muscles of his legs are innervated, and so on -- should we feel that that sentence lost the character of a sentence?' The same question could of course have been asked of the sign-post ('if we were shown how the words 'turn left' written on the sign act...should we feel that the post lost the character of a sign?'), and it seems that the answer is No. Since we assume that sentences and sign-posts are parts of a psychological causal nexus set up by the training, etc., which we regard as part of learning language, we have no reason to suppose that they stand to lose their linguistic character if this causal role is spelt out more fully. And see also §495, in which Wittgenstein apparently regards it as clear and unproblematic that we can regard learning language as 'adjusting a mechanism to respond to a certain sort of influence.'

This sensible view is inconsistent with the anti-causalism commonly ascribed to Wittgenstein on the basis, e.g., of his remarks distinguishing reasons and causes. There is indeed uncertainty in this area in Wittgenstein's writings, and Appendix 1 shows how this can also be related to Davidson.

[23] Compare §506: 'The absent man who at the order "Right turn!" turns left, and then, clutching his forehead, says "Oh! Right turn" and does a right turn. -- What has struck him? An interpretation?' Clutching the forehead here is comparable to the behaviour Wittgenstein cites below, characteristic of correcting a slip. And although according to Wittgenstein the agent himself is not interpreting the order, but simply acting to correct himself, this is something which we might interpret, in the course of reading the 'rule' for 'Turn Right!' from his behaviour, including behaviour in which he turned left.

[24] Thus consider also §31, in which Wittgenstein distinguishes learning to behaving in accord with rules from learning the rules themselves: '...One can also imagine someone's having learnt the game without ever learning or formulating rules. He might have learnt quite simple board games by watching, and have progressed to more and more complicated ones....' The first paragraph of this remark is explicitly about learning language, and I take it that Wittgenstein regards

the learning of natural language as comparable to this, and interpretation of the kind he describes in §54, §82, and §206-7 as the framing of hypothesis about the rules with which such behaviour accords. Tim Williamson points out that in his 'Vagueness and Logic' (*Philosophy of Science*, 6, 1939,163-80) Hempel 'compares abstracting a language from behaviour to learning the rules of chess from watching it played.' See *Vagueness* (London: Routleged, 1994), p 79.

[25] This use of quotation marks, which I shall retain for ease of exposition, is of course criticized and replaced by Tarski in 'The Concept of Truth in Formalized Languages,' in *Logic, Semantics, Metamathematics* (Oxford: Oxford University Press, 1956.)

[26] Wittgenstein puts essentially this point at Tractatus 5.542, where he says " 'A believes that P', 'A has the thought that P' and 'A says P' are of the form ' "P" says P' ". Extending this to desires or wishes requires accepting that a desire or wish that P is satisfied just if P, and so denying that the satisfier is whatever quells the feeling of nonsatisfaction, or again that satisfaction must make one feel better (Cf. §440 and §441). This point is discussed further in section IX below.

[27] This was also Wittgenstein's view, although he was not concerned to put it in print. Thus he is reported as saying in his 1936 lectures that **GET REF**

[28] The topic of the confirmation and disconfirmation of such interpretive hypotheses is discussed more fully in my 'Interpretation, Cogency, and Self-Knowledge', forthcoming, and in relation to psychoanalysis in 'Patterns of Interpretation: Speech, Action, and Dream' cited in footnote 4.

[29] Considering the role of these patterns also enables us to see the non-causal notion of truth specified in **NL** as an interpretive concept, and one which is anchored in, or which abstracts from, the causal notions of the satisfaction of desire (des P -[causes]-> P) and verification of belief (P -[causes]-> bel P) which appear in **D**. In consequence we can see our capacity for understanding persons as integrating the understanding of the truth-conditions of our own sentences with that of motives for action more generally, and in such a way as to render **NL** a specification of content which spans language and mind.

[30] This, I think, indicates that Wittgenstein's understanding of the relation of consciousness and intentionality is different from that of philosophers such as Searle, who take consciousness as more basic.

[31] Thus at *Investigations* 174e Wittgenstein writes that 'One can imagine an animal angry, frightened, unhappy, happy, startled. But hopeful? And why not...Can only those hope who can talk? Only those who have mastered the use of a language. That is to say, the phenomena of hope are modes of this more complicated form of life. (If a concept refers to a character of human handwriting, it has no application to beings that do not write).' Davidson's view is related to

interpretation more directly: see 'Thought and Talk' in *Essays*, and 'Rational Animals' in LePore, ed, *Truth and Interpretation* .

[32] 'What we have to accept, the given, is -- so one could say -- forms of life.' (*Investigations* 226e) But of course what is accepted in philosophy can still be explained elsewhere, and in the normal way.

[33] Davidson's treatment of first-person authority also makes many of the same points, and turns on this distinction. See 'First Person Authority', *Dialectica*, Vol 38, No 2-3 (1984); 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, 1987; and 'What is Present to the Mind', in Brandl and Gombocz, eds, *The Mind of Donald Davidson* Amsterdam: Rodophi 1989.

On this point it is worth recalling again Wittgenstein's §506. Wittgenstein clearly intends that we should say that the absent-minded man who responds to the order to turn right by turning left and then corrects himself has not been struck by an interpretation, in the sense of a further and perhaps clarifying verbal expression of the order 'Turn right.' Hence in correcting himself and turning right he illustrates Witgenstein's claim in §201, that there is a way of grasping a sentence which is not an intepretation -- which does not consist in knowing an alternative verbal formulation -- but which shows in being able to act in a correct, as opposed to an incorrect, way. Also the example bears out Wittgenstein's claim that we should not say that every action in accord with the rule is an interpretation, for this is to confuse finding the correct verbal formulation of the rule, which is one thing, with acting in accord with the rule, which is another. Interpreting a rule someone else is following is a matter of fitting language to action; following the rule is one of fitting action to language.

Still it must be noted that we do not fully explain first-person authority by citing the fact that others must interpret what I say and do, while I need not do so, or again the fact that I must know what I think and mean if I am to be interpretable. These assymetries tend to explain why I know my own mind and language better than others, but not why either I or others should know such things so well. I try to say something more about this in section X below.

[34] As his use of the picture theory at §139 indicates, the views which Wittgenstein criticizes in the *Investigations* are mostly to be found in the *Tractatus*.

[35] It is not mandatory to interpret Wittgenstein in this way. It is possible that he was unaware of any connection between, e.g., (i) his mentioning the explorer/interpreter in referring to Augustine in §32, and his later use of the figure; (ii) Augustine's description of intention as shown in 'the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of voice' and his own concept of natural expression; and (iii) Augustine's 'natural

language of all peoples' and his own 'the common behaviour of mankind'. But such unawareness seem unlikely to me.

In this context it is worth noting that other aspects of §32 can be understood by reference to later remarks as well. Wittgenstein says that Augustine treats the child as if it could think, but not yet speak, where 'think' means something like 'talk to itself.' At §32 it seems quite premature to connect thinking with talking to oneself; and no such connection actually appears in Augustine, except that the infant's thoughts are presented as fully articulate. But Wittgenstein's argument through §445 supports the claim that articulate thought presupposes an interpretable practice of language, so that thinking of the kind Augustine ascribes to the infant really would have to mean something like talking to oneself -- as in the case of the monologue people of §243.

[36] *Tractatus* 4.025.

[37] This need not imply that the infant's understanding of what adults do (of what they say, for example) may not outpace the infant's ability to do the same; for Wittgenstein's remarks assign a role to the infant's observation of the practice as a part of learning it. The idea is that the learning of the practice of putting things into words is not simply based upon, but develops together with, the formation of articulate thought.

[38] 'The Structure and Content of Truth', *The Journal of Philosophy*, 1990, p. 296.

[39] Wittgenstein, *Philosophical Grammar*, Oxford: Blackwell Publishers, 1974, p 189. The work by Segal and Larson cited in footnote 14 is an example of using a theory of truth together with Chomskian grammar to describe 'part of the psycho-physical mechanism.'

[40] 'The Structure and Content of Truth', *The Journal of Philosophy*, 1990, p325n.

[41] 'Expressing Evaluations', p. 17-18.

[42] The emphasis upon action related to sentences can also be used to explain how a Davidsonian approach to meaning meets Dummett's requirement that a speaker should be able to manifest understanding of sentences; and also why a Davidsonian theory is not modest in the sense (or senses) which Dummett deplores. It will also, I think, serve to explain a language-user's possession of concepts in accord with a 'dovetailing' of the requirements for ascription of the concept with linguistic practice, as discussed by Peacocke (see his 'Linguistic Understanding' in Dancy and Sosa, eds, A *Companion to Epistemology*, Oxford: Blackwell Publishers, 1992.) In these terms the view of both Wittgenstein and Davidson would be that possession of many concepts can be secured only via linguistic or interpretive practice.

[43] 'The Structure and Content of Truth', *The Journal of Philosophy*, 1990, p325.

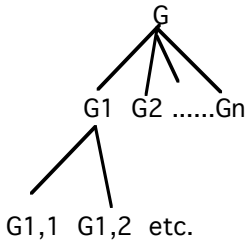
[44] These quotations are from 'The Structure and Content of Truth', p 316; and 'The Second Person' *Midwest Studies in Philosophy* XI, 1992. The role of convention is also pursued in 'Communication and Convention' in *Inquiries* and 'A Nice Derangement of Epitaphs' in LePore, ed, *Truth and Interpretation*.

[45] Imposition of a theory of truth will in turn resolve these into connections between words and things -- e.g. that 'Socrates' denotes Socrates, or the predicate 'red' is satisfied by red things -- which have a related normative status; so the picture is closely comparable to Wittgenstein's at this level also.

[46] See in particular 'The Second Person' and 'Davidson, Donald' in S Guttenplan, ed, *A Companion to the Philosophy of Mind,* Oxford: Basil Blackwell, 1994.

[47] 'Davidson, Donald', p 234. Davidson's reference to 'a finer analysis' apparently takes in Wittgenstein's example in §185. Also, since Davidson is discussing the case in which interpreter and interpretee are actually communicating, his remarks cover some of the same ground as Wittgenstein's §240-42; Wittgenstein's §243 involves a possible difference. In discussing people 'who speak only in monologue' Wittgenstein makes clear that the coincidence of interpreter's expectations with interpretee's intentions (and hence expectations) which he thinks required can arise in the absence of communication. Davidson seems to reject this possibility; but since he does not consider Wittgenstein's kind of example explicitly, some uncertainty remains.

[48] This kind of representation is intutively fairly clear, but let us spell out what is involved a little more fully. Suppose we have a goal G connected by branches to sub-goals G1 to Gn, and these by further branches to further sub-goals G1,1, G1,2, etc., as in the following:




Here the top tree corresponds to a desire that G and a belief that if G1 and G2 and... Gn (in that order) then G. So this tree constitutes a complex instance of **PR**, as does the tree down from G1, which correspondes to a desire that G1 and a belief that if G1,1.... then G1; and so on down the tree. Where we use a tree to represent the forming and ordering of the desires (goals) which consititute a plan, the lines connecting goals can be read as instances of '-[causes]->' as this appears in **PR**. Then insofar as this plan is translated into successful action, the desires will operate in the sequence indicated by the diagram, and will successively bring about the circumstances in which they are fulfilled, which are described by the same

sentences, so that each will figure in an instance of **D**. Thus each desire in the diagram should be pacified just when the rightmost desire subordinate to it is registered as satisfied, and this holds sentence by sentence from the top through the bottom of a whole tree. The ordering of goals manifest in the tree thus corresponds to series of instances of **D**, nested in accord with **PR**, in which each desire in the diagram has a particular place. We can think of a branch of a tree of this kind as extending to each of intentional bodily movements which an interpretee executes in performing an action, and of the interpretee as working through the tree in action by performing the movements in the required order from left to right. This mode of representation is discussed further in my 'Patterns of Interpretation' and 'Interpretation, Cogency, and Self-knowldege' cited above.

[49] It is a part of commonsense understanding that first-person authority tends to be defective in the case of motives like jealousy or envy, with which psychoanalysis deals. The considerations in this and the next section may help to make clear why the attainment of first-person authority is a goal of psychoanalysis. Some further points about this topic are made in 'Patterns of Interpretation' and 'Wittgenstein, Interpretation, and the Foundations of Psychoanalysis', cited in footnote 4.

[50] That is, it seems reasonable to suppose that both the ability to interpret and the ability to make oneself understood might confer reproductive advantage on a creature who possessed them; so that once evolution got it hands on a system of animal communication it might cull and save in favour of both.

[51] For a recent discussion of this topic, and some references to the literature, see Robert J. Matthews, 'The Measure of Mind', *Mind*, April 1994. The ascription of the view to Lewis is discussed in Field's postscript to 'Mental Representation' in N. Bloch (ed) *Readings in Philosophical Psychology* Cambridge, Massachusetts: MIT Press. Davidson first makes the comparison of concepts applied in interpretation with weight or length in 'Mental Events': see *Essay*s p 220 ff; his most explicit discussion is in 'What is present to the Mind?' in J. Brandl and W. Gombecz, eds, *The Mind of Donald Davidson,* Amsterdam: Rodophi 1989.

[52] Compare for example his discussion of the standard meter at §50, and of the determination of length in his brief remarks on 'judging other people's motives' at 225e).

[53] There is a more explicit rejection of the idea that agreement determines truth for mathematical statements at *Investigations* 226e.

[54] Thus on the present account interpretation -- including the independent interpretability of authoritative avowals -- provides a basis for the objectivity of meanings, or the requirements imposed by rules, which is both appropriately strong and consistent with our sureness of grasp on them. This may constitute a difference between the present approach and that taken by Crispin Wright to the same range of questions in a recent and interesting series of papers. (See, e.g., his 'On Making Up One's Mind: Wittgenstein on Intention' in P. Weingartner & G.

Schurtz, eds., *Logic, Philosophy of Science and Epistemology: Proceedings of the 11th Annual Wittgenstein Symposium*, Vienna: Holder-Pickler-Temsky; 'Wittgenstein's Rule-following Considerations and the Central Project of Theoretical Linguistics' in A. George, ed., *Reflections on Chomsky*, Oxford: Basil Blackwell, 1989; 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention', *The Journal of Philosophy*, 1989; and other articles cited and discussed in Peter Sullivan, 'Problems for a Construction of Meaning and Intention' *Mind*, 1994).

Very roughly, Wright argues that it is *a priori* credible that a person believes that he intends to ø just if he intends to ø, and that this is best explained by regarding first-person judgements about intention as *extension-determining*, that is, as serving to fix the extension of the notion of intention, rather than as reflecting an extension -- a set of intentions -- which is independently constituted.

In the present account self-ascriptions are elements in a correlation between utterances and other actions which is central for the interpretation of language and hence for the precise interpretation of action. So self-ascription does have role which might be regarded as partly constitutive of the phenomenon of articulate intention. Still, the role is that of one part of a correlation, which holds as among phenomena having common causes (verbal and non-verbal action) and the common causes (intentions, desires, beliefs) themselves. In this account the role of self-ascription is balanced by that of potential and independent ascription by others, regarded as a form of empirical hypothesis; and such independent interpretation locates the same range of causes as the avowals, and is required for their ratification. So, as sketched above, first-person authority is no less important than cogent interpretability, and each, indeed, provides the possibility for the other.

Wright's account would seem to admit of a construal along these lines, since he takes it that self-ascription must 'cohere with the subject's behaviour'; but his formulations seem mainly to concentrate on the 'best opinion' of the self-ascriber, and hence to leave the relation of avowals to the rest of behaviour, as stressed here, out of account.

[55] Thus in 'The Second Person', *Midwest Studies in Philosophy*, XVII (1992), p 264, Davidson claims that the kind of cause-locating triangulation which he describes as holding among interpreter, intepretee, and environment is a requirement of interpretation; and this leads him to make a direct comparison between his view and Wittgenstein's, saying that his argument shows that 'there cannot be a private language, that is, a language understood by only one creature'. This is slightly different from Wittgenstein's claim, which concerns a language which *could* be understood by only one creature.

[56] I take it that these arguments are familiar enough not to require detailed repetition. There is a fuller treatment in my 'Wittgenstein and Physicalism' *Proceedings of the Aristotelian Society*, 1972.

[57] I should stress that this account of what Wittgenstein means in this remark contradicts that given by many commentators. Thus to take one of the most distinguished and thoroughly researched commentarys as an example, in *Wittgenstein, Meaning, and Mind*, Oxford: Blackwell 1990, P. M. S. Hacker writes that Wittgenstein holds that 'pain is neither a something nor a nothing...he denies nothing other than the misleading picture we associate with the inner process.' Hacker fails, I think, to observe that ascribing to Wittgenstein the *interlocutor's* interpretation of his remarks -- namely that the 'sensation itself' is not a something -- he acribes more than a denial of the misleading *picture* of the inner process; he denies the *process* itself, in denying that the process really is something, really is a real thing. (cf. §308: ...And now it looks as if we had denied mental processes. And naturally we don't want to deny them.') Wittgenstein's aim is to discredit the picture *without* impugning the processes of which the picture is a picture. Hacker's interpretation reflects the inability which Wittgenstein is attacking, to separate picture from process, and so contradicts Wittgenstein's own explicit claim in §423 that *certainly* all these *things* happen in us.

On this point note that Wittgenstein's italicized uses of 'something' and 'nothing' in §304 clearly refer back to uses by his misunderstanding interlocutor in previous remarks: thus compare the 'Yes, but there is *something* there accompanying my cry of pain' of §296, which refers back to the 'not even as a *something* ' of §293; and the 'that when he writes 'S' he has *something* ...' of §261, which in turn arises out of the 'did the man...make a note of *nothing whatever?* ' of §260. This makes clear that Wittgenstein's use of 'not a *something*, but not a *nothing* either' is not an expression of his own view of the matter, but rather his characterization of his interlocutor's conception, which he takes to be erronious and paradoxical, but a consequence of the too concrete adoption of the cognitive metaphor under discussion.

[58] See for example Lakoff, 'The Contemprary Theory of Metaphor' in Ortony, ed, *Metaphor and Thought*, Second Edition (Cambridge: Cambridge University Press, 1993), and other papers referred to there. Reddy's classic paper is particularly interesting, for both describing the metaphor and taking it literally.

[59] See §282: '"But in a fairy tale the pot too can see and hear!" Certainly, but it *can* also talk.'

[60] See the 'And now it seems quite indifferent whether I have recognised the sensation *right* or not; let us suppose I regularly identify it wrong, it does not matter in the least.' of §270.

[61] For a survey of recent discussion of this 'gulf', in the form of a supposed 'explanatory gap' between consciousness and brain-process, see Block, 'Consciousness' in Guttenplan, ed, *A Companion to the Philosophy of Mind* , Oxford: Blackwell, 1994.

[62] On this see Wollheim, 'The Bodily Ego' and 'Psychology, Materialism, and Sexuality' in *The Mind and Its Depths* Cambridge: Harvard University Press, 1993; and my 'Synthesis in the Imagination', cited in footnote 4.

[63] Oxford: Blackwell Publishers, 1967.

[64] *Essays on Action and Events*, p 207ff.

[65] 'What is present to the Mind?' p.16. Davidson's reference is to John Searle, 'Indeterminacy and the First Person', *The Journal of Philosophy* 84 (1987) pp 123-146.

[66] Davidson makes clear that equally acceptable accounts may vary not only in the way they translate a given sentence, but also in the truth value they assign it (see *Inquiries* p 239); the view, as he put it earlier, is that meaning 'is what is invariant as between different acceptable theories of truth' (*Inquiries* p. 225).

[67] Davidson, 'The Inscrutability of Reference', *Inquiries*, p 230.

[68] Davidson, 'Meaning, Truth, and Evidence' in Le Pore, ed, *Truth and Interpretation*, Oxford: Blackwell, 198

[69] Davidson, 'The Inscrutability of Reference', *Inquiries*, p 239.

[70] Davidson, 'On the very Idea of a Conceptual Scheme', Inquiries, p.196.

[71] See, e.g. the remarks in 'The Inscrutability of Reference', Inquiries p 228-9; the larger role of charity and agreement is to reduce a range of indeterminacy which would otherwise be intractable.

[72] I omit discussion of Wittgenstein's comparison of interpretive hypotheses to formulae covering the development of a numerical series. This yields, among others, a conception of indeterminacy analogous to Davidson's, in which alternatives are compared to alternative formulae (perhaps connected with alternative procedures of computation) which can be used to compute the same function. Wittgenstein also uses the comparison with temperature, to indicate the importance of differences which are verbal. Thus §508:

> I say the sentence: "The weather is fine"; but the words are after all arbitrary signs -- so let's put "a b c d" in their place. But now when I read this, I can't connect it straight away with the above sense. I am not used, I might say, to saying "a" instead of "the", "b" instead of "weather", etc. But I don't mean by that that I am not used to making an immediate association between the word "the" and "a", but that I am not used to

using “a” *in the place* of “the” -- and therefore in the sense of “the”. (I have not mastered this language)

(I am not used to measuring temperatures on the Fahrenheit scale. Hence such a measure of temperature ‘says’ nothing to me.)

[73] There is some discussion of this topic in my ‘Visual Geometry’, *The Philosophical Review*, January 1972.

[74] *Investigations* 230e. This view is consistent with Davidson’s discussion, e.g. at p.197 of ‘On the Very Idea of a Conceptual Scheme’. Davidson does not deny that the optimizing of agreement on which understanding rests can include learning to think as the interpretee does.