ON MAKING UP ONE'S MIND: WITTGENSTEIN ON INTENTION

Crispin Wright
University of St. Andrews, GB

This paper is a sequel to my discussion of Kripke's *Wittgenstein on Rules and Private Language* which appeared in the *Journal of Philosophy* in 1984.[1] That paper concluded with some somewhat compressed remarks about the differences between Kripke's Wittgenstein and Wittgenstein. My principal purpose here is to relieve some of that compression: to explain in more detail how, as it seems to me, the two are related and what is the real moral of Kripke's argument.

I. *Kripke's argument*

We shall need, as briefly as maybe, to outline the sceptical argument which *Investigations* §§ 184–202 suggested to Kripke. The conclusion of the argument is that there are no states of affairs for whose description it is necessary to have recourse to the concepts of meaning, understanding or any other concepts cognate to them. So there are, accordingly, no facts about what any particular expressions meant or mean, and, consequently, no facts about what constitutes correct use of an expression on any particular occasion. The argument proceeds, as is familiar, in the setting of a debate with a Sceptic who doubts your knowledge of what you formerly meant by some randomly chosen expression, E. The facts which constitute your erstwhile understanding of E, it is assumed, will have to be found in aspects of your former behaviour and/or mental life. Accordingly, if you are idealised to have perfect recall of the events in both provinces, then you ought to be in a position to know the relevant facts and hence to see off the Sceptic. Conversely, if the Sceptic wins his debate with you, then the requisite facts cannot be there to be found in those two provinces—so, on the assumption that there, if anywhere, is where they must be, they do not exist at all. Since the same debate can be constructed for any particular agent and any particular expression, it follows that there are no facts about what anyone formerly meant by any expression. Since the debate will have the same result tomorrow with respect to your claim then to know what you meant by E today, it follows that there are no facts about whah anyone *presently* means either. But the, as it were, impersonal meaning of an expression must supervene upon what individuals mean by it: so it follows, finally, that there are no facts about what any expression means or meant.

So much, then, for the setting of the sceptical argument—a setting for which there is, Kripke would have to admit, little if any indication in Wittgenstein's text.[2] A prima facie successful argument within this setting would, however, be of the greatest philosophical interest in any case—and Wittgenstein's preoccupation with, if not exactly Kripke's problem, then anyway a cluster of related questions concerning the nature of determinacy of meaning,—a preoccupation which is pivotal to so much of the thought both of the *Investigations* and of the *Remarks on the Foundations of Mathematics*—lends urgency to the question: does his mature philosophy contain the resources for a satisfactory response to Kripke's Sceptic, even if that character is never sharply delineated in Wittgenstein's own work?

The actual contest with the Sceptic, you will remember, is a somewhat one-sided affair. He points out, reasonably as it seems, that the totality of your previous behaviour, a fortiori your actual historical uses and responses to uses of E, must be reconcilable with indefinitely many alternative interpretations of your former understanding of E, since it is a merely finite pool of data. (The underlying assumption is that finite data may always be subsumed under

indefinitely many distinct generalisations.) So the sought-after facts must be found, it seems, —if they can be found at all—in the pageant of episodes of your former mental life: your imaginings, sensations, dreams, thoughts and so on. But understanding an expression in a particular way is a state of a certain *generality*: a state whose content bears on the use of that expression in an open-ended set of circumstances. So the aspects of your former mental life, which you are to recall in response to the Sceptic's challenge, have to subserve that generality: what you recall has to be capable of constituting a state which can have something to say about the proper use of E in no end of situations. This would seem to exclude images, visualisations, sense experience—for no such process in consciousness has the requisite general content. Minutely faithful documentation of all such processes could, it appears, be added to the description of your total previous behaviour without in any way restricting the range of available unwelcome interpretations of your former understanding of E. All, it appears, that can stand a chance of effecting such a restriction would be a certain kind of general *thought*: a thought somehow encompassing all of the unwelcome interpretations and branding them as foreign to your intentions with E. From here the Sceptic wins in three moves. First, it is not at all clear how such a general thought might be formulated. Second, the strategy of seeking such a thought gets the intuitive epistemology of understanding wrong: your confidence that you know what you formerly, or presently mean by an expression is not at all contingent upon your ability to recall, or conjure a thought which somehow uniquely nails down that meaning. Third, and most important, the strategy is question-begging. Thoughts are thought in symbols. To assume the ability to recall a thought is therefore to assume the ability to recall the significance you attached to certain symbols. That is just to take it that you have the capacity which the Sceptic is disputing—the capacity to know your former meanings.

II. *Three possible responses*

The sceptical argument is more suggestive than coercive, involving various challengeable assumptions. One such explicit assumption was that facts about your former understanding of E must somehow be constituted by aspects of your former behaviour and episodes in your consciousness. Is not understanding better construed as a *disposition*? That you mean, for instance, *green* rather than $grue_{2000}$ by 'green' may be manifest only in your responses to certain future and/or counterfactual situations—there may be no recovering the fact from anything you have so far actually said or thought.[3] Kripke discusses this response in some detail, bringing to bear upon it two powerful-looking objections. First, there is the question whether our dispositions are not altogether too improverished to provide the basis for an account of meaning. The meaning of the plus sign in arithmetic, for instance, has—in conjunction with other relevant semantic facts—something to say about the truth-values of literally infinitely many simple arithmetical equalities about which we have no dispositions of judgement at all. In fact, our dispositions for arithmetical judgement are presumably finite. So the original problem, of extrapolating a unique account of your understanding of E by consideration of a finite pool of data, remains—it is merely that the originally finite pool is finitely enlarged by the addition of certain indicative and counterfactual conditionals. Second, meaning is normative: the meaning of an expression determines how it ought to be used, not necessarily how it actually is. Indeed I am, in a large class of cases, disposed to use expressions in ways that actually fail to fit my understanding of them—because I tend to make mistakes, or misperceive, or whatever. Yet if the meaning I attach to an expression were constituted by how I am disposed to use it, it would make no sense to suppose that I could be disposed to behave in ways that ran counter to it. If I were so disposed, I would not have the original disposition.

There has been some discussion whether a dispositionalist account may not have the resources to meet Kripke's objections.[4] I shall not try to decide that issue now, but will interpolate two considerations which will suggest that the dispositionalist can at best aspire to

a revisionary conception of meaning and understanding, even if Kripke's objections can somehow be surmounted.

Note that understanding cannot straightforwardly be a disposition, any more than belief and desire can, and for the same reason: understanding interacts holistically with other psychological states, including beliefs and desires, in the explanation of (linguistic) behaviour. How I am actually disposed to use an expression on a particular occasion will depend not just on how I understand it, but on the way I apprehend the circumstances, my other background beliefs, my intentions in so using it, and my other background desires. It is no exaggeration to say that *any* sort of performance with the expression can be reconciled with a subject's possessing some specified understanding of it, if one is willing to make sufficiently elaborate adjustments in these other parameters. It will be felt, no doubt, that while this point occasions some complications for the dispositionalist response to Kripke's Sceptic, it involves no serious compromise: we are still at liberty to see somebody's understanding as dertermined by their dispositions of use, *modulo* some fixed set of assumptions about them in the other relevant parameters. But there is a further point which now becomes salient, and which may well have been what Kripke, in invoking the idea of normativity of meaning, fundamentally had in mind. Assume some such background. Then the dispositionalist is taking it, in effect, that *whatever* the subject does, he cannot but behave in a way that is appropriate to expressing his understanding, relative to that background. That is, while sense may be attached to the idea of a speaker's using an expression in a way which does not strictly accord with the way he understands it, this failure of accord must—in the dispositionalist view—be explained by appeal to what is true of him in respect of the *other* background parameters. He misuses the expression, for instance, because he intends to deceive his audience, or is attempting irony, or is mistaken about the prevailing circumstances. Thus if, in particular, the subject correctly cognizes the relevant facts, and intends a literally informative use of the expression in question, the dispositionalist view would have it that he *must* then be disposed to use the expression appropriately (as far as his own understanding of it is concerned). Now, whatever the merits of this view, it is unquestionably a fundamental adjustment in traditional thinking about meaning. The *objectivity* traditionally associated with meaning would have it that what sort of use of an expression is appropriate, *modulo* a suitable set of background assumptions, is settled *independently* of the subject's response. What one ought to say, *modulo* the background assumptions,—the requirement imposed by one's understanding—is something to be *cognized*. There is a question of fitting one's behaviour to antecedent semantic tracks. It makes no difference whether we think of these tracks, after the style of mediaeval realism, as constituted in the nature of platonic universals, or whether we think of them as a man-made contract. Either way it cannot, on this way of thinking of the matter, be taken for granted that, *ceteris paribus*, one will be disposed to do what, semantically, one ought.

This idea of the objectivity of meaning cuts immensely deep in our ordinary thinking. It is unclear, for instance, that, unless we appeal to it, any good sense can be made of the idea that currently undecided issues, in whatever sphere of inquiry, may *already* possess determinate answers settled by the content of relevant statements and the state of the world in relevant respects. A fortiori, it is unclear whether it can be permissible to think—in the style of realism as conceived by Dummett—that statements may be determinate in truth-value which, in the nature of the case, we cannot appraise. The dispositionalist account thus brings into question our picture of an external world which is determinate in no end of respects for which we have adequate descriptive resources, irrespective of our cognitive achievements to date and of the limits of principle on such achievements in the future.

I have suggested elsewhere, and will suggest again later in this paper, that this is something which is at issue anyway. My immediate purpose is merely to urge on those who would favour a dispositionalist response to Kripke's sceptic the recognition that *they* are bringing it into question.

That was the first point. The second point is an application of something which have in effect, already noticed and which will be important to us in the sequel. Self-ascriptions of a

specific mode of understanding of some expression, like self-ascriptions of a large class of beliefs, intentions and sensations, are a kind of *avowal*. There is much to be said about the characterisation of this notion,[5] but here it will suffice to recall that subjects are credited with a special authority for their avowals, that we think of the knowledge which they thereby express as groundless and immediate. It is hard to see what justification there could be for this practice if what one ascribed, in self-ascribing a particular understanding of some expression, was a *disposition*. How can I know without evidence that I have a particular disposition or complex of dispositions, and why should I be credited with any kind of authority on the matter? What substitute could there be for *empirical* appraisal, of a kind which anyone else could as well engage in as myself? The dispositionalist account, as much as the move which tried to hold off the Sceptic by appeal to the occurence of some appropriately general thought, misrepresents the intuitive epistemology of understanding.

The same complaint is to be levelled against another, otherwise intuitively fair response to Kripke's Sceptic: namely, that at the very first stage of the debate he presupposes the cogency of what is, in effect, Goodman's version of inductive scepticism. The fact that no end of alternative generalisations are consistent with any finite accumulation of behavioural data is a ground for holding that there can be no rational preferences among them only if the 'grueish' projections which Goodman describes are supposed to be as rational as the kind of inductions that we actually favour. Yet even if Goodman's Sceptic can be defeated and a more refined methodology exposed which rationally bars the Goodmanian inductions—however exactly they are to be characterised—, it remains that knowledge of one's own meanings seems to owe nothing to the application of such a methodology but may be consistent with e.g. largescale amnesia concerning one's previous linguistic behaviour.

There is a connection here with what is, I believe, the proper reading of *Investigations* § 201. Wittgenstein writes:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here.
> 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 one contented us at least 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* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.

In allowing, apparently, that there is indeed an ineliminable multiplicity of conflicting possible interpretations, I read Wittgenstein as granting that, if I may so put it, a *merely rational* methodology can indeed yield no determinate conclusions when it comes to disclosing what meaning is incorporated in someone's practice with an expression hitherto. His conclusion, however, is explicitly not the Sceptic's, that there is no fact of the matter concerning the character of the subject's understanding, but rather "that there is a way of grasping a rule which is *not* an *interpretation*"—that is, I take it, that something other than the exercise of interpretative (rational) faculties enters into the capacity to 'read' another's linguistic behaviour. And the additional something is, crudely, human nature: certain sub-rational propensities towards conformity of response, towards 'going on in the same way', which alone make possible the formation of the common conceptual scheme within which our rational capacities can be exercised. A quite common thought in recent moral philosophy in Britain has been that a sharing of certain basic human concerns is prerequisite for the formation of certain of our moral concepts.[6] The thought which Wittgenstein is putting in opposition to the, by comparison with Kripke's, rather straightforward form of scepticism set up in *Investigations* § 185 and following[7] is a generalisation of that: it is not merely concerns which must be shared but a whole plethora of natural classificatory dispositions if we are to find each other's linguistic behaviour intelligible.[8]

This simple point is of great importance, and armed with it, we can perhaps explain how my

former understanding of E could be salient in a sufficient sample of my behaviour, even if no fully rational ground could be given for discounting various unwelcome alternative accounts of it. But, again, that is a point to use against Kripke's Sceptic only from an assumed *third-personal* point of view, as it were, with respect to one's own linguistic past. It would do nothing to legitimate the idea of one's present and former meanings as apt for *avowal* in the germane sense. And it is in the legitimation of that idea that a satisfying response to the Sceptic must ultimately be found.

III. *The (W)right response*

Kripke does consider the response that meaning so-and-so by E might be a *sui generis* state,[9] like the experience of yellow or a toothache, and rightly emphasises the point I built into the summary of the argument above, that no such state can have the general, directive content essential to meaning. Later[10] he briefly considers the idea that meaning so-and-so by E might, as it were, be more *sui generis* yet, beyond illustration by comparison with states of any other familiar kind. But this he presents as merely obscurantist. How could there be such a state, available immediately to the subject, apt for authoritative avowal and noninferential recall, yet possessing determinate, potentially infinite content?

In my previous discussion I set in opposition to this the reflection that, so far from finding any mystery in the matter, we habitually assign just these characteristics to the ordinary notion of *intention*. To summarise: the intuitive notion of intention has it that it is a state of mind, alongside mood, thought, desire, sensation, etc., for which, in at least a very large class of cases, subjects have a special authority, and whose epistemology is first/third person asymmetric. I can be presumed to know, at least in a very large class of cases, what my intentions are, and this knowledge does not proceed, or does not have to proceed, by reflection on what I say and do,—the only basis that you can have for an opinion about what my intentions may be. As far as the relation between intention and thought is concerned, it can come to be true of me that I have a certain intention without my engaging in any process of conscious deliberation or thinking any thought which specifies that intention's content. Rather I may simply find myself with my mind made up, as it were,—able to give an account of my intentions if asked, but with no story to tell about the when or why of their onset. More important in the context of the sceptical argument, it is a feature of the intuitive concept of intention that, even when there is an association with a content-specifying train of thought, the subject does not know of his intention *via* that train of thought. It is not, for instance, because I visualise a golf ball disappearing down a hole that I know that I am currently attempting to put out; rather it is because I know my current intention that I recognize the visualised sequence of events as germane to specifying its content. Knowledge of one's own intentions, in the cases which interest us, is based on inference neither from one's behaviour nor from other occurrent aspects of one's mental life. Finally, intentions may be general, and so may possess, in the intuitively relevant sense, potentially infinite content. I may have the intention to sue if anybody slanders me, to follow Kipling's advice and always respond alike to victory and defeat, and to use 'green' to express green. How can an intention bear on a potential infinity of cases? Well, just because it may be the intention to respond in a certain way to any case of a certain sort, and because there may be a potential infinity of cases of that sort.[11]

Had the sceptical argument been directed against intention in general, rather than at what it is tempting to regard as the special case of meaning, there is no doubt that the intuitive concept seems to contain the resources for a direct rebuttal. Since I can know of my present intentions non-inferentially, it is not question-begging to respond to the Sceptic's challenge to my knowledge of my past intentions to reply that I may simply remember them. And this is just the reply that Wittgenstein does, in effect, give. At the conclusion to Part I of the *Investigation* he writes:

(692) It is correct for someone to say: "When I gave you this rule, I meant you to . . . in this case"? Even if he did not *think* of this case at all as he gave the rule? [My Italics]. Of course it is correct. For "to mean it" did not mean: to think of it . . .

(693) "When I teach someone the formation of the series . . . I surely mean him to write . . . at the hundredth place".—Quite right; you mean it. And evidently without necessarily even thinking about it. This shows you how different the grammar of the verb "to mean" is from that of "to think". And nothing is more wrong-headed than calling meaning a mental activity!

If we may think of meaning so-and-so by E as either consisting in, or relevantly similar to, possession of an intention or complex of intentions,[12] then the Wittgensteinian response to Kripke's Sceptic would evidently be to confront him with the 'grammar' of the ordinary notion, and in particular its distinction from any kind of episodic state of consciousness. The diagnosis would be that the Sceptic has fallen prey to the kind of misunderstanding from which, in Wittgenstein's view, so many philosophical perplexities arise—a misunderstanding based on superficial similarities in functionally quite different forms of discourse. Seeing a verb—"I *intended* him to write thus-and-such at the so-and-so manyeth place",—we are inclined to cast about for a state, or process, or activity of consciousness for that verb to report, which contains, as it were, an instruction for the so-and-so manyeth place. But there need not and, in the case of an open-ended general intention, cannot always be any such appropriately explicit happening in consciousness. The proper conclusion is that the connection between an intention and the act which implements or frustrates it need not and in general cannot be anticipated by states or processes of consciousness distinct from intending. To suppose otherwise is to cause the downfall of the notion—as the sceptical argument shows.

IV. *The real problem*

We cannot leave matters like that, however, for now it is apt to seem utterly mysterious how the connection—between the prior intention and the performance which implements or goes against it—is effected at all. Wittgenstein writes:

(197) There is no doubt that I now want to play chess, but chess is the game it is in virtue of all 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 connection exists between the act of intending and the thing intended?—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?

This brings us to a central, perhaps the central preoccupation of the philosophy of mind of the *Philosophical Investigations*. How can there *be* a state, or act, or however one wants to regard it, with the features I characterized of the intuitive notion of intention? And what is the basis for the authority allowed to the intending subject? What is striking is that Cartesianism, whatever other difficulties it may encounter, is not even of prima facie service to us here. Cartesianism would view the authority as having the same kind of basis which it finds for a subject's authority concerning his or her occurrent sensations. The subject has privileged access to the state, is immediately aware of it in consciousness. Others, in constrast, can approach it only by an indirect, inferential route. But how, for instance, can my authority for the claim that at the so-and-so manyeth place I intended you to write down thus-and-such, be based on introspection, if, as has been stressed, nothing which went on in me and which has any plausible claim to be regarded as a state of consciousness, explicitly anticipated the case of the so-and-so manyeth place at all? An unthinking reply would be: but many not proper performance at the so-and-so manyeth place have been anticipated *implicitly*? But that gets us absolutely no further. For *what* was introspectibly true of you, at the earlier time, which made

it the case that it was *that* kind of performance at the so-and-so manyeth place, rather than something else, which was implicit in the way you meant the rule to be taken?

The problem is not peculiar to meaning and intention and is illustrated in the *Investigations* by a variety of examples.[13] Here are some relevant passages.

(139) When someone says the word "cube" to me, for example, I know what it means. But can the whole *use* of the word come before my mind when I *understand* it in this way?

Well, but on the other hand isn't the meaning of the word also determined by this use? And can these ways of determining meaning conflict? Can what we grasp *in* a *flash* accord with a use, fit or fail to fit it? And how can what is present to us in an instant, what comes before our mind in an instant, fit a *use*?

---

(151) Let us imagine the following example: A writes a series of numbers down; B watches him and tries to find a law for the sequence of numbers. If he succeeds he exclaims: "Now I can go on!"—So this capacity, this understanding, is something that makes its appearance in a moment. So let us try and see what it is that makes its appearance here.—A has written down the numbers 1, 5, 11, 19, 29; at this point B says he knows how to go on. What happened here? Various things may have happened; for example, while A was slowly putting one number after another, B was occupied with trying various algebraic formulae on the numbers which had been written down. After A had written the number 19 B tried the formula $a_n = n^2 + n - 1$; and the next number confirmed his hypothesis.

Or again, B does not think of formulae. He watches A writing his numbers down with a certain feeling of tension, and all sorts of vague thoughts go through his head. Finally he asks himself: "What is the series of differences?" He finds the series 4, 6, 8, 10 and says: Now I can go on.

Or he watches and says "Yes I know *that* series"—and continue it, just as he would have done if A had written down the series 1, 3, 5, 7, 9.—Or he says nothing at all and simply continues the series. Perhaps he had what may be called the sensation "that's easy!" (Such a sensation is, for example, that of a light quick intake of breath, as when one is slightly startled.)

(152) But are the processes which I have described here *understanding*? "B understands the principle of the series" surely doesn't mean simply: the formula "$a_n =$ . . . " occurs to B. For it is perfectly imaginable that the formula should occur to him and that he should nevertheless not understand. "He understands" must have more in it than: the formula occurs to him. And equally, more than any of those more or less characteristic *accompaniments* or manifestations of understanding.

(153) We are trying to get hold of the mental process of understanding which seems to be hidden behind those coarser and therefore more readily visible accompaniments. But we do not succeed; or, rather, it does not get as far as a real attempt. For even supposing I had found something that happened in all those cases of understanding,—why should *it* be the understanding? And how can the process of understanding have been hidden, when I said "Now I understand" *because* I understood?! And if I say it is hidden—then how do I know what I have to look for? I am in a muddle.

---

(184) I want to remember a tune and it escapes me; suddenly I say "Now I know it" and I sing it. What was it like to suddenly know it? Surely it can't have occurred to me *in its entirety* in that moment!—Perhaps you will say: "It's a particular feeling, as if it were *there*"—but *is* it there? Suppose I now begin to sing it and get stuck?—But may I not have been *certain* at the moment I knew it? So in some sense or other it was *there* after all. But in what sense?

---

(187) "But I already knew, at the time when I gave the order, that he ought to write 1002 after 1000"—Certainly; and you can also say you *meant* it then; only you should not let yourself be misled by the grammar of the words "know" and "mean". For you don't want to say that you thought of the step from 1000 to 1002 at that time—and even if you did think of this step, still you did not think of other ones. When you said "I already knew at the time … " that meant something like: "If I had then been asked what number should be written after

1000, I should have replied '1002'." And that I don't doubt. This assumption is rather of the same kind as: "If he had fallen into the water then, I should have jumped in after him".—Now, what was wrong with your idea?

---

(197) "It's as if we could grasp the whole use of a word in a flash."—And that is just what we say we do. That is to say: we sometimes describe what we do in these words. But there is nothing astonishing, nothing queer, about what happens. It becomes queer when we are led to think that the future development must in some way already be present in the act of grasping the use and yet isn't present—For we say that there isn't any doubt that we understand the word, and on the other hand its meaning lies in its use.

These are highly suggestive passages. The most noticeable similarity running through them is that each is concerned with what appears to be a routinely psychological concept—grasping a meaning, understanding how to continue a series, remembering how a tune goes . . . etc.—which, like a headache or sensation, for instance, may have an abrupt onset, as it were, and may be recognized to do so by the subject. The problem Wittgenstein is concerned with is that, in contrast to a headache or sensation, these concepts contain a load which may—and in certain cases must—go quite unreflected in the subject's consciousness as onset takes place. Thus each of the following is a perfectly coherent claim:

> X suddenly grasped the meaning of E—without having its whole use come before his mind.
>
> X suddenly understood how to continue the series—without having the whole series come before his mind.
>
> X remembered how the tune went—without playing it through 'in his head'.
>
> X meant Y to continue '1002' after '1000'—without consciously thinking of that particular case.
>
> X decided to play chess—without thinking of all the rules.

Now, the application of each of these predicates to X will be defeasible if his subsequent performance is inadequate when measured by a standard implicit in the things he didn't think about—if he goes on to misuse E, miscontinue the series, or cannot whistle the tune, or rejects Y's continuing '1002' after '1000', or shows he doesn't understand the rules of chess. But it is not that developments of this kind are taken to be a symptom that certain other constitutive events or processes, of a kind of which X *would* have been fully conscious, did not take place. The point is rather that the events of X's consciousness, *whatever* in fact they were, do not suffice for the applicability of such predicates. Whatever took place in X's consciousness at the time he decided to play chess, for instance, that very same phenomenology could have accompanied a decision of a quite different sort. And as the quoted passages bring out, there is a case to be made that any actual phenomenology is also not necessary. A formula, for instance, may occur to X when he understands how to continue the series, but need not.

These concepts, then, seem to incorporate a tension. You could say: some aspects of their employment seem proper only if they characterize states of consciousness while others suggest something quite different. How, if nothing happening in the subject's consciousness is uniquely distinctive of the concept that comes to apply to him, can he be in *position* to apply it, usually with complete confidence? The matter is apt to seem all the more puzzling when the subject's awareness is retrospective—as Wittgenstein brings out beautifully for the case of intention:

(633) "You were interrupted a while ago; do you still know what you were going to say?—If I do know now, and say it—does that mean that I had already thought it before, only not said it? No. Unless you take the certainty with which I continue

the interrupted sentence as a criterion of the thought's already having been completed at that time.—But, of course, the situation and the thoughts which I had contained all sorts of things to help the continuation of the sentence.

(634) When I continue the interrupted sentence and say that *this* was how I had been going to continue it, this is like following out a line of thought from brief notes.

Then don't I *interpret* the notes? Was only one continuation possible in these circumstances? Of course not. But I did not *choose* between interpretations. I *remembered* that I was going to say this.

(635) "I was going to say . . . "—You remember various details. But not even all of them together show your intention. It is as if a snapshot of a scene had been taken, but only a few scattered details of it were to be seen: here a hand, there a bit of a face, or a hat—the rest is dark. And now it is as if we knew quite certainly what the whole picture represented. As if I could read the darkness.

(636) These 'details' are not irrelevant in the sense in which other circumstances which I can remember equally well are irrelevant. But if I tell someone "For a moment I was going to say . . . " he does not learn those details from this, nor need he guess them. He need not know, for instance, that I had already opened my mouth to speak. But he *can* 'fill out the picture' in this way. (And this capacity is part of understanding what I tell him.)

(637) "I know exactly what I was going to say!" And yet I did not say it.—And yet I don't read it off from some other process which took place then and which I remember.

Nor am I *interpreting* that situation and its antecedents. For I don't consider them and don't judge them.

This is a graphic expression of the non-inferential character of knowledge of one's own intentions. But Wittgenstein's point is not just that the recollection of events in consciousness of other *sorts* may provide an insufficient basis for a judgement about the former intention. It is that recollection of anything properly viewed as an event, state or process of consciousness is so insufficient—that the intention, construed as such an event in its own right, is evanescent.[14]

V. *Towards a solution*

The general problem being posed, it will be evident, has two distinct aspects. First, there is the question of with what to replace the Cartesian conception of these concepts. When such a concept comes to apply to a subject—perhaps quite abruptly—what *kind* of change in him does it mark, if not one in his conscious mental state? The kind of answer which Wittgenstein seems to want to accept is, very crudely, that the change marked is the onset of some kind of capacity or complex of capacities, or of a range of dispositions. Thus in 187, quoted above, he offered

> 'When you said "I already knew at the time . . . ' that meant something like: "if I had then been asked what number should be written after 1000, I should have replied '1002'" and that I don't doubt. This assumption is rather of the same kind as "If he had fallen into the water then, I should have jumped in after him". (Compare 682—4, quoted below)

If X comes to grasp the meaning of a particular expression, or to understand how to continue a particular series, or remembers how a tune goes, he does so only if in each case he acquires certain constitutive abilities which he did not possess just beforehand. On the other hand what has to be true of him, if it is to be said that he, at some specific earlier time, meant Y to continue subsequently in a particular way, is that he had certain dispositions at the earlier time—roughly, to allow or reject certain responses by Y; while the decision to play chess is answerable to the disposition to go on to do so, provided no interfering factor intrudes. This

all seems unexceptionable as far as it goes, and there is a temptation to see the purpose of Wittgenstein's discussion as being no more than to effect a certain kind of overall adjustment in the way we think of the kind of concept in question—an adjustment, summarised in the slogan that "Understanding is not a mental state", which takes us away from the Cartesian picture of intention and its kin as movements within an inner theatre and in the direction of something more functional.

That is, no doubt, an adjustment which needs to be learned, even if to take the point is hardly to swim against the tide. But matters will not rest there. We want more detail, of course.[15] It is not however lack of detail but the second aspect of the overall problem which now presses us. In recognizing that the facts marked by the application of these concepts are persistently fugitive if sought within the sphere of consciousness, we seem to have made a mystery out of a phenomenon of first-person avowal of the application of such concepts. This was the point with which we earlier faced the dispositional and other responses to Kripke's Sceptic. How is the *subject* to know of the application of such a concept to himself if nothing of which he is conscious intimates the fact to him? And how, moreover, can it be reasonable to credit a subject—in the cases where we do—with a *special* authory concerning such states of himself if the Cartesian picture quite misrepresents their epistemology? The authority is easily understood if the subject is conceived as the sole witness, as it were. But when that is no longer the picture, what sense can be made of this aspect of our linguistic practice? There are plenty of dispositions and capacities for which there is no institution of authoritative self-ascription—the confident self-ascription, without behavioural grounds, of intelligence, courage, patience, or endurance, for instance, is, so far from being authoritative, a mere conceit. Why is not a conceit throughout the range of concepts with which we are concerned?

The question is difficult and probably admits of no uniform answer. If a subject's knowledge or—in cases where he has it—his special authority is not owed to the introspectibility of the relevant kind of state of affairs, only two possibilities remain. The first is that it is based on introspection of something else; and the second is that it is not based on introspection at all. The former is at least the natural place to start in the case of the first three concepts illustrated—grasping the meaning of an expression 'in a flash', suddenly understanding how to continue a series, and remembering how a tune goes. In these cases there is no institution of avowal in the strict sense—the subject's word carries no *special* authority. What needs explanation is rather how a subject can be so much as in position to be *reliable*. If the proof of the pudding is in subsequent performance, what basis can there be for an opinion at a stage at which a third party can have none? The beginnings of an answer would be that there *are* certain kinds of phenomenological episodes—flashes of imagery, surges of confidence, and so on—which, experience teaches, correlate often enough with the capacity to go on to deliver the right kind of performance, whatever it may be, and which accordingly provide an experimental basis for claiming the capacity. But this cannot be the right place to look with the other two examples—meaning Y to continue 1002 after 1000, and deciding to play chess—or with self-ascriptions of intention generally. What I actually feel, e.g. when I exclaim "Now I can go on" may be the same in different cases involving different series. Not only may this be so—it must be so if I am to have an *inductive* basis for expecting my confidence to be fulfilled. For the series in question may be one which I have never encountered before. But the striking feature about the self-ascription of an intention is that one *identifies* it at the time: there is no such thing as knowing that one has an intention of some sort but not knowing *what* is intended. Since we are able confidently and authoritatively to self-ascribe intentions which we have never had before, it simply cannot be that we identify our intentions in general by inductive association with other, genuine states of consciousness. (Because this point is decisive, we may prescind from consideration of the additional difficulties associated with the question, how such an inductive association might anyway be established.)

For these cases, then, we require a different explanation, dissociated from introspection. So far as I can see, there is only one possible broad direction for such an explanation to take. The authority which our self-ascriptions of meaning, intention, and decision assume is not

based on any kind of cognitive advantage, expertise or achievement. Rather it is, as it were, a *concession*, unofficially granted to anyone whom one takes seriously as a rational subject. It is, so to speak, such a subject's right to declare what he intends, what he intended, and what satisfies his intentions; and his possession of this right consists in the conferral upon such declarations, other things being equal, of a *constitutive* rather than descriptive role.

Authoritative utterance of this constitutive sort is not without analogues. Someone umpiring a tennis match, for instance, may declare "The ball was good", intending only to report the physical fact of where the ball bounced. But if he speaks as *umpire*, then what he says will determine another, institutional fact, which is constituted by his decision, and which—to the possible chagrin of the players—contributes towards a resolution of the outcome of the game in a way in which the physical fact may not. If the analogy is to go further, of course, we shall require that the umpire's decision be not final but, at least exceptionally, defeasible, (in the light, perhaps, of the judgement of the tournament referee,)—just as avowals of intention may on occasion reasonably be discounted by a third party. Even so, the analogy limps in this respect: the umpire's pronouncements are made in the light of physical facts which are available to anyone, and the sole ground for criticism of them is the appraisal of these facts. By contrast, an avowal of intention need have no basis, and insofar as it may be criticisable, the basis for the criticism may only be constituted by states of affairs that were not salient, or even did not exist, at the time of the avowal—par excellence, the subject's subsequent behaviour.

Wittgenstein writes:

> (682) "You said, 'It'll stop soon'.—Were you thinking of the noise or of the pain?" If he answers "I was thinking of the piano-tuning"—is he observing that the connexion existed, or is he making it by means of these words?—Can't I say *both*? If what he said was true, didn't the connexion exist—and is he not for all that making one which did not exist?
>
> (683) I draw a head. You ask "Whom is that supposed to represent?—I: "It's supposed to be N."—You: "But it doesn't look like him; if anything, it's rather like M."—When I said it represented N.—was I establishing a connexion or reporting one? And what connexion did exist?
>
> (684) What is there in favour of saying that my words describe an existing connexion? Well, they relate to various things which didn't simply make their appearance with the words. They say, for example, that I *should have* given a particular answer then, if I had been asked. And even if this is only conditional, still it does say something about the past.

This is encouragement that the present suggestion is, for the relevant class of concepts, on the right exegetical tracks. But what more definite cast can we give it? One strategy is to proceed by asking under what circumstances we would discount a subject's avowal of an intention. Well, supposing that we are concerned with someone who is in general deemed *able* to be reliable on such matters,—so not some form of delusional psychotic, for instance—we would do so only in fairly special circumstances: when there was independent evidence of insincerity, or of some sort of muddle about the content of the intention claimed, or of some relevant self-deception, or—in the case of a retrospective avowal—of some major disorientation of memory. Suppose that the range of grounds on which such an avowal may be legitimately discounted is both finite and circumscribable.[16] Then we may expect that some such biconditional as this will hold a priori:

> X intends that P if and only if X is disposed to avow the intention that P, and would be sincere in so doing, and fully grasps the content of that intention, and is prey to no material self-deception, and . . . and so on.

Now, such a biconditional may be read in two contrasting ways.[17] One reading—the *detective*—would hold that the left-hand side serves to describe a determinate state of affairs which, if all the provisos on the right-hand side are met, the subject is able to *apprehend*. On this view, the provisos collectively determine the conditions for a cognitive success, which an avowal

may then serve to report. The alternative, however, is to accord priority to the right-hand side. The resulting view would see the disposition to make the avowal as *constituting* the state of affairs reported by the left-hand side when the provisos are met. So the subject's cognition of an independent state of affairs does not come into the picture. Rather, he is *moved* to make the avowal and, subject to the provisos, it stands. Accordingly, any ground for discounting the avowal has to be cast in the form of reason to say that some one or more of the provisos was not really satisfied.

It will be similar with the authority for retrospective avowals of intention, etc. Rather than as based on recollection of an antecedent fact, the non-detectivist will regard it as owing to the operation of a similar right-to-left priority, only with the right-hand side of the biconditional modified so as to include at least the proviso that there is no major disorientation of memory at work. Cartesianism would, of course, enjoin a detective reading of such biconditionals. But its failure to deliver any credible conception of the constitution or epistemology of the states of affairs in question may be taken to suggest not that we should dismiss them as fiction, but that the relevant biconditionals should be read the other way round.

I do not expect this proposal to be satisfying at this stage of development, but I cannot take it further on this occasion. The view at which we have arrived—I would like to think that it might be close to Wittgenstein's view—might be summarised as follows. It is part of regarding human beings as persons, rational reflective agents, that we are prepared to ascribe intentional states to them, to try to explain and anticipate their behaviour in terms of the concepts of desire, belief, decision and intention. And it is fundamental anthropological fact about us that our initiation into the language in which these concepts feature results in the capacity to be moved, who knows exactly how, to self-ascribe states of the relevant sorts—and to do so in ways which not merely tend to accord with the appraisals which others, similarly trained, can make of what we do but which provide in general a far richer and more satisfying framework for the interpretation and anticipation of our behaviour than any at which they could arrive if all such self-ascriptions were discounted. The roots of first-person authority for the self-ascription of these states reside not in cognitive achievement, based on cognitive privilege, but in the success of the practices informed by this cooperative interpretational scheme.

How do matters now stand with the recommended response to Kripke's Sceptic? The response was, in effect, that your former understanding of E either resided in, or may appropriately be compared to, your possession of certain former intentions, which you may now non-inferentially recall. This response is intact but is cast in a totally different perspective. The conception of meaning as a '*sui generis*' mental state involved regarding your authority for your intentions, present and former, as based on privileged access to a condition of consciousness, of such content as to determine of a potential infinity of doings/sayings whether they fit or fail to fit it. Such fit, or failure of fit is conceived as settled just by the character of an intention, independently of any judgement of yours, though—except where you misapprehend *what* is said or done—it is supposed that it is only in exceptional circumstances that your judgment about such fit or failure of fit would go astray.[18] In short: the *sui generis* mental state conception goes for a detective reading of the biconditionals whose left-hand sides describe your former meanings or intentions. The new perspective reverses the reading. Avowals of present or past intention are disbelievable only in special circumstances —those embraced by the provisos in the (complete) biconditional—and elsewhere are authoritative not because the subject is uniquely well placed to know but just because he or she is the subject.

A corollary. Since to identify a *former* intention is to identify what would fit or fail to fit it, taking a non-detective view of a subject's avowal of a former intention involves taking the same view of his judgment about what, if anything, now implements or frustrates that intention. Accordingly, what, if any pattern of performance is imposed on a subject by the constraint of compliance with a former intention *is not settled independently of his judgement of the matter*. His judgements may be discounted—in the special circumstances when some one

or more of the provisos is not met—but otherwise they serve to determine, rather than objectively accord with or violate the content of his anterior intention. Elsewhere[19] I have urged that one central moral which Wittgenstein wished to draw from his discussion of rule-following was that the notion we tend to have of the objectivity of meaning is untenable: the idea that meanings can somehow be constituted, once and for all, either within a community or by a single subject, by finitely many events—explanations, uses, episodes in consciousness—so that thereafter there is only the objective question of *fit* between new uses of the relevant expression and the meanings thereby laid down. There are a variety of grounds for disquiet with this notion, and I am now suggesting that the real message of Kripke's dialectic is to teach us one more. For intention, when non-detectively construed, *offers no resources for the construction of such objective meanings*. And the sceptical argument succeeds only if it is allowed to restrict the search for the fact constitutive of what I meant by E at t in such a way to exclude my *subsequent* determination of it. It is a *reductio ad absurdum* of meaning, understanding and intention when, but only when, they are detectively construed.

NOTES

1 Kripke, Saul, *Wittgenstein on Rules and Private Language* (Oxford 1981).
Wright, Crispin, "Kripke's Account of the Argument against Private Language", *Journal of Philosophy* LXXXI (1984), pp. 759—778.

2 A remark on the relation between the 'paradox' of which Wittgenstein speaks in 201 and Kripke's. The former is indeed to the effect that there is no such thing as accord or conflict with a rule. But it is explicitly conditional on the assumption that what complies with a rule is determined via *interpretation* of the rule, and seems to consist essentially just in the reflection that, with ingenuity, any particular rule can always be so interpreted as to sanction any particular piece of behaviour. By contrast, the second phase of Kripke's paradox explicitly concerns the fugitive character of *non-interpretative* knowledge of one's own (former) rules and meanings. It will be clear as this paper proceeds that I believe that Kripke's second phase does indeed pose a problem which Wittgenstein attempted to solve and which is prominent elsewhere in the *Investigations*. But the 'paradox' of 201 is only part of the paradoxical train of thought which Kripke expounds and exegetes will search 184—202 in vain for an unmistakable anticipation of the latter.

3 In case someone doesn't know: an object is grue_{2000} just in case it is green at all times in its history up to midnight on December 31st, 1999, and blue at all times in its history thereafter.

4 See Wright, op. cit. n. 1, p. 771 n. 3.

5 A useful discussion is Peter Hacker's in his *Insight and Illusion* (Oxford 1972), ch. 9. See also Andrew J. Hamilton, *The Self and Self-Consciousness*, St. Andrews Ph. D. thesis 1987.

6 See, for instance, John McDowell's "Non-cognitivism and Rule-Following" in: S. Holtzman and C. Leich (eds.), *Wittgenstein: To Follow a Rule* (London 1981), pp. 141—62, esp. pp. 144—5.

7 Cf. note 2 above.

8 Cf. section III of my "Skolem and the Skeptic", *Aristotelian Society Supp.* Vol. LIX (1985), pp. 117—137, and the Introduction to my *Realism, Meaning and Truth* (Oxford 1986), pp. 27—9.

9 Kripke, op. cit. n. 1, pp. 41—50.

10 Kripke, op. cit. n. 1, pp. 50—53.

11 Cf. Wright, op. cit. n. 1, pp. 775—7.

12 Cf. *Investigations* pt. II xi, p. 214: "In a law-court, for instance, the question might be raised how someone meant a word. And this can be inferred from certain facts.—It is a question of *intention*."

13 I draw here on Malcolm Budd's and Peter Carruthers' excellent respective discussions in "Wittgenstein on Meaning, Interpretation and Rules", *Synthese* 58 (1984), pp. 303—23, and "Ruling out Realism", *Philosophia* 15 (1985), pp. 61—78.

14 Compare *Investigations* II xi, pp. 216—7:

> Someone tells me: "Wait for me by the bank". Question: Did you, *as you were saying the word*, mean this bank?—This question is of the same kind as "Did you intend to say such-and-such to him on your way to meet him?" It refers to a definite time (the time of walking, as the former question refers to the time of speaking)—but not to an experience during that time. Meaning is as little an experience as intending.
>
> But what distinguishes them from experience?—They have no experience-content. For the contents (images for instance) which accompany and illustrate them are not the meaning or intending.

The intention *with which* one acts does not 'accompany' the action any more than the thought 'accompanies' speech. Thought and intention are neither 'articulated' nor 'non-articulated'; to be compared neither with a single note which sounds during the acting or speaking, nor with a tune.

[15] Though more refined accounts of the conditions of application of concepts of the kind in question are unlikely to throw up anything of comparable philosophical importance to that of grasping in general, in advance of the detail, such concepts' role in marking the presence of capacities and dispositions.

[16] Actually, it is not important to the suggestions which follow that this should be so.

[17] Compare the concluding pages of my "Realism, Anti-Realism, Irrealism, Quasi-Realism" in the *Midwest Studies in Philosophy* number on Realism, ed. H. Wettstein, forthcoming 1987. I owe this form of contrast to Mark Johnston.

[18] Quite what is the *ground* for this belief, according to the detective view, is of course dark.

[19] See in particular my *Wittgenstein on the Foundations of Mathematics* (London 1980), passim, and "Rule-Following, Meaning and Constructivism" in: C. Travis (ed.), *Meaning and Interpretation* (Oxford 1986), pp. 271–97.

* * *

SGEGENSTÄNDEN IM ANSCHLUSS TEINS *TRACTATUS*

Albert Menne
Universität Bochum, BRD

„Die Welt ist alles was der Fall ist" (*TLP* 1.1). „Die Gesamtheit der bestehenden Sachverhalte ist die Welt" (*TLP* 2.04). „Die Logik erfüllt die Welt; die Grenzen der Welt sind auch ihre Grenzen" (*TLP* 5.61). „Die Tatsachen im logischen Raum sind die Welt" (*TLP* 1.13). „Daß die Sätze der Logik Tautologien sind, das *zeigt* die formalen – logischen – Eigenschaften der Sprache, der Welt" (*TLP* 6.12). Nach Wittgenstein gibt es demnach in der Welt Tatsachen, es gibt Sprache, es gibt Logik.

„Was der Fall ist, die Tatsache, ist das Bestehen von Sachverhalten" (*TLP* 2). Tatsachen sind also bestehende Sachverhalte. Aber es gibt daneben nach Wittgenstein auch nicht bestehende Sachverhalte. In der Sprache können beide mit Hilfe von logischen Strukturen dargestellt werden.

Die Sprache selbst als vollzogener Sprechakt ist eine Tatsache. Verlassen wir nun die Wittgenstein'sche Art der Darstellung und wenden uns unter erkenntnistheoretischem Aspekt dem menschlichen Bewußtsein zu. Sprache ist nach Gerhard Frey „Ausdruck des Bewußtseins".[1] In unserem Bewußtsein treffen wir Vorstellungen an, die sich offensichtlich auf etwas beziehen, was auch außerhalb des Bewußtseins Bestand hat: Das sind die Vorstellungen von Tatsachen. Ein einfaches Beispiel: Wenn ich mir diesen Tisch vor mir vorstelle, dann setzt dieser Tisch, der Gegenstand meiner Vorstellung, einem Faustschlag von mir Widerstand entgegen. Das Vorhandensein dieses Tisches ist also ein bestehender Sachverhalt, eine Tatsache. Genauer noch: Eine materielle Tatsache. Und solche materiellen Gegenstände sind uns in Raum und Zeit gegeben. Sie sind nicht unveränderlich, sondern können ihre räumliche Position ändern, können im Laufe der Zeit ihre Eigenschaften ändern.

Daneben treffe ich in meinem Bewußtsein eine andere Art von Vorstellungen an, die sich auf nicht bestehende Sachverhalte beziehen, wie z.B. den Pegasus oder den Zentauren. Während ich in meinem Bewußtsein die Tatsachen, die Vorstellungen vom Wirklichen so hinnehmen muß, wie sie mir gegeben sind, wie sie unabhängig von mir ihre Wirkungen ausüben, kann ich über Vorstellungen von nicht bestehenden Sachverhalten relativ frei verfügen: Ich kann dem Zentauren ein weißes oder ein braunes Fell zusprechen, ich kann ihn mir mit krausem oder mit glattem Haar, ja ich kann ihn mir auch glatzköpfig vorstellen. Ja ich kann mir sogar widerspruchsvolle Vorstellungen von nicht bestehenden Sachverhalten machen, wie z.B. Städte, die nördlich von sich selber liegen, viereckige Kreise und so fort.

Doch diese beiden Arten von Erkenntnisgegenständen, die Vorstellungen von Tatsachen, von Wirklichem, und die Vorstellungen von nicht bestehenden Sachverhalten, von Phantasieprodukten, das Pur Mentale, erschöpfen nicht die vorhandenen Arten von Erkenntnis-Gegenständen: In keine unserer bisherigen beiden Gegenstandskategorien passen die strukturellen Gebilde der Logik und der Mathematik. Als Beispiel sei der Kreis genommen, der definiert sei als „der geometrische Ort derjenigen Punkte, die von einem gegebenen Punkt einen gegebenen Abstand haben." Setze ich das als Bewußtseinsinhalt, so entspricht dem nichts im Bereich des Wirklichen: es gibt in der Wirklichkeit keinen exakten Kreis! Es gibt dort höchstens Zeichen für Kreise oder es gibt dort Gegenstände mit Eigenschaften, die in mehr oder weniger grober Annäherung dem geometrischen Kreis ähnlich sind. Nun kann man oft hören, der Kreis sei dann eben durch Abstraktion aus diesen kreisähnlichen Gebilden der Wirklichkeit entstanden.

Darauf ist dreierlei zu antworten: