CRISPIN WRIGHT

RAILS TO INFINITY

Essays on Themes from Wittgenstein's *Philosophical Investigations*

HARVARD UNIVERSITY PRESS · Cambridge, Massachusetts · London, England
2001

7 Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics

I

The Central Project, and the creative aspect of language.

The Central Project of theoretical linguistics, for present purposes, is to achieve an understanding of one component in what Chomsky has termed the 'creative aspect'[1] of our language use—our ability, after exposure to the use of no more than a small part of our language, to recognise of indefinitely many unencountered strings both whether they constitute well-formed sentences and what, if anything, they could be used in a particular context to say. No doubt this ability admits of *some* kind of scientific theoretical explanation. What is characteristic of the Project is the thought that, in the first place anyway, it is appropriate to seek an explanation in, broadly, *cognitive-psychological* terms: that it is an ability which we have because we are appropriately related to a finite body of *information* which may be

The other component: language use is stimulus-free.

1. The other component, as is familiar, is that uses of language are characteristically ***stimulus-free;*** language is 'undetermined by any fixed association of utterances to external stimuli or physiological states (identifiable in any non-circular fashion)' (Chomsky (1966), pp. 4–5). Chomsky speaks (1964), p. 7, of the first component as 'the central fact to which any significant linguistic theory must address itself'.

inferentially manipulated in such a way as to entail, for each novel string on which we can successfully exercise our 'linguistic-creative' power, appropriate theorems concerning its grammaticalness and content.[2]

Wright's question: is the Project at odds with Wittgenstein?

It is a basic and familiarly thorny question how the 'appropriate relation' which is to mediate between the contents of a suitable grammar, or theory of meaning, and the abilities of actual speakers may be best conceived—whether, for instance, it is happy to involve a notion of *knowledge*, however attenuated. But that is not now my immediate concern.[3] My question, perhaps even more basic, is whether, even in this inchoate and highly general formulation, the Central Project is not somehow already at odds with lessons to be learned from Wittgenstein's discussions of rule-following in *Philosophical Investigations* and elsewhere.[4]

2. In Chomsky (1967), pp. 4–5, for instance, he writes: "A person's competence can be represented by a *grammar*, which is a system of rules for pairing semantic and phonetic interpretations. Evidently, these rules operate over an infinite range. Once a person has mastered the rules (unconsciously, of course), he is capable, in principle, of using them to assign semantic interpretations to signals quite independently of whether he has been exposed to them or their parts, as long as they consist of elementary units that he knows and are composed by the rules he has internalized." Compare Dummett (1976), p. 70: "A theory of meaning will, then, represent the practical ability possessed by a speaker as consisting in his grasp of a set of propositions; since the speaker derives his understanding of a sentence from the meaning of its component words, these propositions will most naturally form a deductively connected system." A useful assembly of source material for this basic thought is provided, their deflationary purposes notwithstanding, by Baker and Hacker (1984a), chaps. 8–9.

3. I have considered some of the issues involved in Wright (1986c), chap. 6, 'Theories of Meaning and Speakers' Knowledge'.

4. Both the present chapter and my 'Theories of Meaning and Speakers' Knowledge' are attempts to pursue issues raised in the symposium I had with Gareth Evans (Evans (1981); Wright (1981) (this volume Essay 2)). The principal purpose of my contribution (cf. Wright (1980), chap. 15), was to raise the question to which the present paper is addressed. But Evans's reply focused on the more specific questions to do with the notion of implicit knowledge which 'Theories of Meaning and Speakers' Knowledge' tried to pursue further. Somehow or other, with the obvious exceptions of Kripke's brief remarks and Chomsky's response (1986), there has been virtually no discussion of the present question since.

Kripke's Wittgenstein: meaning-talk states nothing true or false.

The tension is very immediate on Kripke's widely discussed interpretation (1982) of Wittgenstein. According to Kripke's Wittgenstein, all our discourse concerning meaning, understanding, content and cognate notions fails of strict factuality—says nothing literally true or false—and is saved from vacuity only by a 'Sceptical Solution', a set of proposals for rehabilitating meaning-talk in ways that prescind from the assignment to it of any fact-stating role. If that claim were correct, there would be nothing strictly and literally true to say about the content of a novel utterance nor, therefore, anything to be strictly known or recognised concerning its content. So linguistic creativity, conceived as the ability to *know* the syntactic and semantic status of encountered strings, would be a myth: there would be, simply, no such knowledge-forming power. The apparently remarkable, unbounded character of our linguistic competence would no more demand a deep, *cognitive* explanation than would a disposition to consensus about the *comic* qualities of an indefinite variety of situations (if indeed we were so disposed, and assuming, for the sake of argument, the non-factuality of claims concerning comedy). It would still, no doubt, be possible to raise the question what *does* explain our ability to communicate using novel linguistic constructions, if it is not that we deploy essentially the same store of syntactic and semantic information to arrive at the same answers to the same problems. Maybe the question could be so cast as to admit of some sort of natural scientific treatment. But we would have surrendered the vision which inspired the Project: the vision of basic English, say, as an articulate system of syntactic and semantic *fact*, whose mastery consists in the internalisation of a store of principles out of which all aspects of the system are generated. That vision calls for the existence of *truths* concerning the content of novel utterances, truths which Kripke's Wittgenstein repudiates. There would still be something, our 'agreement in semantic reactions', which might seem remarkable. But since it would be, precisely, no longer an agreement in *judgements*, properly so described, there would seem to be no legitimate project of enquiry into its cognitive provenance.

Then creativity is a myth, and the comedy comparison.

Without semantic facts, no enquiry into agreement's cognitive source.

There is an additional, more specific point of tension between Kripke's Wittgenstein and the Central Project. The latter is a project

in *psychology*, a project of explaining manifest abilities of the individual in terms of his cognitive-psychological (and ultimately, neurophysiological) equipment. Explanans and explanandum are thus both exhausted by characteristics of individuals 'considered in isolation'. By contrast, so Kripke's Wittgenstein argues, we can refashion the notion of meaning in the wake of the Sceptical Argument only by recognising that a proper account of the (non-fact-stating) use of statements essentially involving it and other cognate notions must make essential reference, one way or another, to the purposes and practices of linguistic *communities*. Meaning—what the Sceptical Solution saves of it—cannot lie within the province of individual psychology, so the methodological individualism which characterises Chomsky's work[5] would apparently be guilty of a compound error. There are no facts of the kind which it seeks to explain by reference to (non-existent) states of individuals; and the concept of meaning, as entrenched in our (non-fact-stating) talk of meaning, has an essentially *social* character which the methodologically individualistic perspective imposed by the spurious explanatory project makes it additionally difficult to recognise.

Kripke's own remarks on the relations between the ideas of his Wittgenstein and the Central Project are somewhat low key. He writes, for instance, that

> *if* statements attributing rule-following are neither to be regarded as stating facts, nor to be thought of as *explaining* our behaviour . . . it would seem that the *use* of the ideas of rules of competence [in Chomsky's technical sense] in linguistics needs serious reconsideration, even if these notions are not rendered 'meaningless'.[6]

Maybe such a 'serious reconsideration' might in the end succeed in conserving a use for the ideas of 'rules' and 'competence' in some kind of theoretical description of our linguistic practices. But Chomsky has

5. Cf. d'Agostino (1986), chap. 1.
6. Kripke (1982), pp. 30–1, n. 22.

conspicuously declined to take up the invitation, preferring a frontal attack on the Sceptical Argument and its alleged Sceptical Solution.[7] Concerning the first, the essence of his response, if I may simplify somewhat and presuppose some familiarity with Kripke's dialectic, is that the search for facts constitutive of what you formerly meant by a particular expression, or of its being a particular rule that it was your former practice to follow, is only improperly restricted by Kripke's sceptic to the domains of your former linguistic behaviour and former conscious mental life. Rather, the claim, for example, that you formerly followed a certain specified rule is a *theoretical* claim, and answerable, therefore, in the context of an appropriate embedding theory, to an indefinite pool of past, present *and* future evidence.[8] Any fact may seem fugitive if we do not look in the right place.

Chomsky's answer: which rule you followed is a theoretical question.

If that reflection rebuts the Sceptical Argument, there is of course no need for a Sceptical Solution. But Chomsky finds the Sceptical Solution 'far from descriptively adequate'[9] in any case, demurring at what he takes to be the suggestion of Kripke's Wittgenstein that the assertibility, by the members of a given speech community, that some third party is following rules, requires that 'his responses coincide with theirs'.[10] For this would frustrate the often unproblematic identification of practices governed by rules different to any pursued by the community in question. It is open to question whether this complaint does justice to the resources of the Sceptical Solution.[11] But whether it does or not, it would anyway now be widely accepted that no refinement of the Sceptical Solution can *be* a solution.[12] The crucial ques-

7. Chomsky (1986), pp. 223–43.
8. Which need not be restricted to aspects of behaviour and conscious mental life. See ibid., pp. 236 ff.
9. Ibid., pp. 227 ff.
10. Ibid., p. 266.
11. See George (1987), pp. 158–60. Chomsky replies to George's criticism in the same number of the journal (1987, pp. 189–91).
12. For a variety of points of dissatisfaction, besides Chomsky's, see, e.g., Baker and Hacker (1984b), pp. 429–32; Goldfarb (1985), pp. 481–4; McDowell (1984), pp. 329–30; McGinn (1984), pp. 180–91; Wright (1984), pp. 768–70 (this volume Essay 4).

tion concerns what is wrong with the Sceptical Argument—for it can only have the status of a paradox—and what relation it bears to (the real) Wittgenstein's discussion.

Against Chomsky: it makes first-person knowledge a mystery.

Chomsky's response to the Sceptical Argument is unsatisfying, it seems to me, for reasons which also apply to the dispositional response[13] effectively criticised by Kripke himself.[14] Kripke himself objects to the dispositional response that it cannot account for the normativity of understanding an expression in a particular way, intending to follow a particular rule, and so on. The reason for dissatisfaction which I have in mind, however, is not this. It is rather that Chomsky's suggestion, that the identity of followed rules is a strictly theoretical question, threatens, like the dispositional account, to make a total mystery of the phenomenon of non-inferential, first-personal knowledge of past and present meanings, rules and intentions. The whole temptation to think of understanding, intention, etc., as mental

13. Kripke (1982), pp. 22–3, characterises it like this:

> According to this response, the fallacy in the argument that no fact about me constitutes my meaning plus lies in the assumption that such a fact must consist in an *occurrent* mental state. Indeed the sceptical argument shows that my entire occurrent past mental history might have been the same whether I meant plus or quus, but all this shows is that the fact that I meant plus (rather than quus) is to be analyzed *dispositionally,* rather than in terms of occurrent mental states. Since Ryle's *The Concept of Mind,* dispositional analyses have been influential; Wittgenstein's own later work is of course one of the inspirations for such analyses, and some may think that he himself wishes to suggest a dispositional solution to his paradox.
>
> The dispositional analysis I have heard proposed is simple. To mean addition by '+' is to be disposed, when asked for any sum 'x + y' to give the sum of x and y as the answer (in particular, to say '125' when queried about '68 + 57'); to mean quus is to be disposed when queried about any arguments, to respond with their quum (in particular to answer '5' when queried about '68 + 57'). True, my actual thoughts and responses in the past do not differentiate between the plus and the quus hypotheses; but, even in the past, there were dispositional facts about me that did make such a differentiation. To say that in fact I meant plus in the past is to say—as surely was the case!—that had I been queried about '68 + 57', I *would* have answered '125'. By hypothesis I was not in fact asked, but the disposition was present none the less.

14. Ibid., pp. 24–32.

states, against which Wittgenstein repeatedly inveighs in the *Investigations*, flows from this point of analogy with things which *are* genuine mental states and processes—sensations, after-images, having a tune go through one's head. Each of us is, for the most part, effortlessly authoritative, without inference, about our past and present intentions, about the rules we 'have in mind', about how we understand or have understood particular expressions. An account of the truth conditions of such claims which, like Chomsky's or the dispositional account, makes a puzzle of this aspect of their first-personal epistemology should be rejected.[15]

We're effortlessly authoritative about what we mean and intend.

The sceptic won't accept that I recall meaning addition itself.

The Sceptical Argument is flawed in any case and needs no response of the sort which Chomsky proposes. For there is an explicit and unacceptable reductionism involved at the stage at which the sceptic challenges his interlocutor to recall some aspect of his former mental life which might constitute his, for example, having meant addition by 'plus'. It is not acceptable, apparently, if the interlocutor claims to recall precisely that. Rather the challenge is to recall some *independently characterised* fact, in a way which does not simply beg the question of the existence of facts of the disputed species, of which it is then to *emerge*—rather than simply be claimed—that it has the requisite properties (principally, normative content across a potential infinity of situations). The search is thus restricted to phenomena of consciousness which are not—for the purposes of the dialectic—permissibly assumed 'up front' to have a recollectable *content*. (For to suppose that one could recall the content of a mental state would just be, the sceptic will claim, to assume a positive answer to the question at issue—namely, whether there is such a thing as knowing one's former meanings.) If the sceptic is allowed to put the challenge in this way, then it is no doubt unanswerable. But so put, it is merely an

Unanswerable so put, but only by prejudging meaning.

15. Cf. Wright (1987a), pp. 393–5 (this volume Essay 5).

> Of course, one will be inclined to play down, or query, the claimed feature of the first-personal epistemology of intentional states if one is independently convinced of the 'theoretical' view. And this, unsurprisingly, is what Chomsky does. See, for instance, the rebuttal of Dummett's claim that knowledge of meaning is conscious knowledge in Chomsky (1986), p. 271.

implicit prejudice against the ordinary notions of meaning and intention, according to which we may and usually do non-inferentially know of our present meanings and intentions, and may later non-inferentially recall them.[16]

The flat-footed reply, and where the real problem lies.

This somewhat flat-footed response to Kripke's sceptic may seem to some to provide a good example of 'loss of problems'. (The phrase is Wittgenstein's own, of course, from *Zettel* §456.) In fact, though, and on the contrary, I think the real problem posed by the Sceptical Argument is acute, and *is* one of Wittgenstein's fundamental concerns. But the problem is not that of *answering* the Argument. The problem is that of seeing how and why the correct answer just given can *be* correct.

Meaning hovers between two paradigms: headaches first.

A central preoccupation of the *Investigations*, and Wittgenstein's other late writings on the philosophy of psychology, is with concepts which—like meaning, understanding, intending, expecting, wishing, fearing, hoping—seem to hover, puzzingly and unstably, between two paradigms. To the left, as it were, stand genuine episodes and processes in consciousness: items which, like headaches, ringing in the ears, and the experience of a patch of blue, may have a determinate onset and departure, and whose occurrence makes no demands upon the conceptual resources of the sufferer.[17] To the right, by contrast, stand qualities of character—like patience, courage, and conceit—which are naturally viewed as constituted in the (broadly) behavioural dispositions of a subject, are fully manifest in things he is inclined to say and do, and advert to no inner phenomenological causes of these inclinations. Descartes's conception of the mental tended to draw everything towards the left-hand pole. The Rylean reaction, by contrast, attempted to colonise as widely as possible on behalf of the right. And the difficulty raised by the concepts with which Wittgenstein was preoccupied is that we are pulled in both

The dispositional pole, and Descartes against Ryle.

16. Cf. Wright (1984), pp. 771–8 (this volume Essay 4); Wright (1987a), pp. 395–403 (this volume Essay 5).

17. There are also genuinely episodic states whose occurrence is conditional on the conceptual resources of the sufferer. Having a tune run through one's head is arguably of this character.

directions simultaneously. Their first-personal epistemology pulls us to the left—since explaining it seems to call for construal of such states as objects of consciousness. But, as Wittgenstein and his Kripkean ersatz both effectively argue, nothing strictly introspectible has, in the case of these concepts, the right kind of characteristics. We cannot, honestly, find anything to *be* the intention, etc., when we turn our gaze inward; and anything we might find would have no connection or, at best, the wrong kind of connection with those subsequent events—what we go on to count as fulfilment of an intention or expectation, etc.—on which the correctness of an earlier ascription to us of the intention, etc., depends. But if we accordingly allow ourselves to be pulled towards the right-hand pole—interestingly, Chomsky's response—we encounter the difficulties with normativity (at least in some cases), and with saving ordinary first-personal epistemology, which have already been noted. We need, accordingly, some account or perspective on these concepts which is conservative with respect to the features which give rise to the difficulty—their combination of first-personal avowability and 'theoreticity'—yet dissolves the temptation to assimilate them to cases which they do not fundamentally resemble.

Pulled both ways: nothing introspectible fits.

Keep avowability and theoreticity, assimilate to neither.

Supposing that it is via provision of such an account that the Sceptical Argument can and must ultimately be defused, does any point of contact, or abrasion, remain between Wittgenstein's thought about rule-following and the Central Project? If there is no compelling argument to the conclusion that meaning, etc., are non-factual, is not the Project simply in the clear? Well, whether Wittgenstein's later discussions should be seen as issuing a challenge to the factuality of meaning and rule-following, they undeniably feature a recurrent, explicit discomfort with a certain conception of the *autonomy* of rules, the image of a rule as a rail laid to infinity, tracing out a proper course for a practice quite independently of any judgement of the practitioners (*Investigations* §218). And it is hard to avoid finding this conception of autonomy in the picture of language and linguistic competence which drives the Central Project: the picture of language as a kind of

Is the Project in the clear? The rail laid to infinity.

syntactico-semantic mechanism, our largely unconscious knowledge of which enables us to compute the content which, independently and in advance of any response of ours, it bestows on each ingredient sentence. Chomsky wrote, over forty years ago:

> It is not easy to accept the view that a child is capable of constructing an extremely complex mechanism for generating a set of sentences, some of which he has heard, or that an adult can instantaneously determine whether (and if so, how) a particular item is generated by this mechanism, which has many of the properties of an abstract deductive theory. Yet this appears to be a fair description of the performance of the speaker, listener, and learner.[18]

The mechanism generates, the adult keeps track: RFM kinship.

Note that the *mechanism* does the generating, and the competent adult merely *keeps track* of what (and how) it generates. The conception of language manifested in those remarks—a continuing fundamental theme of Chomsky's thought—is all one, it seems, in its implicit conception of the character of rule-following, with what Wittgenstein has in view in his remarks about mathematical rules:

> But might it not be said that the *rules* lead this way, even if no one went it? For that is what one would like to say—and here we see the mathematical machine, which, driven by the rules themselves, obeys only mathematical laws and not physical ones.[19]

Wittgenstein's remark rewritten with grammar for arithmetic.

What motive could there be for such a view of, for example, the rules of multiplication which would not also motivate:

> But might it not be said that this sentence is grammatical, and has the meaning it does, even if no one ever considers it? For that is what one would like to say—and here we see the language-machine which, driven by the rules of the language themselves, obeys only linguistic laws and not physical ones.

18. Chomsky (1959), p. 57.
19. *Remarks on the Foundations of Mathematics* IV, 48.

Why reject the engine, and not Kripke's way? The shadow case.

Unquestionably, Wittgenstein came to regard this conception as deeply misconceived.[20] But why? And is there any good reason for that verdict which does not in the end lead to the same place as the argumentation of Kripke's sceptic—the conclusion that there are, in truth, no such things as the requirements of rules? Anything which is genuinely an instance of rule-following, it is natural to think—rather than mere charade—has to consist in keeping abreast of certain objective requirements. If it is not that, if the requirements are not settled independently of the way I find myself inclined to go, then I do not *follow* anything—any more than, when walking in bright sunlight, I follow my shadow.[21] Is not the image of the syntactic-semantic engine simply a graphic expression of the rule-governed nature of linguistic practice? And if we reject it, are we not, in effect, rejecting the idea that proper linguistic practice *is* rule-governed—and thereby rejecting the notion of strictly *correct* linguistic practice, and with it the notion of meaning?

II

A reminder of §§185–219 and RFM VI.

A reminder is desirable of the general gist of Wittgenstein's discussions of rule-following. I shall try, briefly, to provide one by concentrating (largely) on §§185–219 of the *Investigations,* and 23–47 of part VI of *Remarks on the Foundations of Mathematics.*

20. Or better, the conception which someone who expressed themselves in this way would be likely to intend to convey. Wittgenstein's conservatism almost always leads him to fight shy of direct criticism of a problematical form of expression. Cf. *Philosophical Investigations* §195:

> 'But I don't mean what I do now (in grasping a sense) determines the future use *causally* and as a matter of experience, but that in a *queer* way, the use itself is in some sense present'—But of course it is, 'in *some* sense'! Really the only thing wrong with what you say is the expression 'in a queer way'. The rest is all right . . .

The problems reside in our *interpretations* of what we say.

21. The comparison is James Hopkins'.

Epistemology, not the reality of rules: we overdignify.

It's notable that the characteristic concerns of these passages seem to have nothing at all to do with the *reality* of rules. Rather, they are epistemological: Wittgenstein is preoccupied with the sense, if any, in which a rule can be genuinely an object of intellection, something whose requirements we *track*—to use a more recent piece of jargon—by grace of some intuitive or interpretative power. Undoubtedly the intention—or one might better say, the suspicion—of these passages is negative. Wittgenstein thinks we tend to badly misunderstand what the accomplishment is of someone who is a competent rule-follower, and that our misunderstandings lead us to overdignify the nature of the constraint imposed by a rule, and the character of our accomplishment in following it. But there is, emphatically, no explicit denial of the *existence* of the constraint, and no consequential rejection of the very notion of accomplishment. It is not *impossible* that their absence is to be accounted for by Wittgenstein's failure to follow through all the way on his own thought, but I shall try to suggest a better explanation.

Four interlocking themes are prominent. First:

> *One's own understanding of a rule does not exceed what one can explain.* (*Investigations* §§209–10; *RFM* VI, 23)

The temptation: more gets across than the explanation says.

The temptation to suppose otherwise arises from the reflection that any explanatory process is finite and, hence open in principle to an indefinite variety of interpretations. Yet explanation does usually—or so we suppose—secure mutual understanding. So, somehow, more is got across—the thought continues—than an explanation makes explicit. Correct uptake of an explanation is having the *right* 'something' come into one's mind as the result of the explanation; and the resulting informational state is then expressed in, but essentially transcends, one's subsequent practice with the concept concerned.

It is important to see that this picture of what is involved in successfully giving and receiving explanations is a cornerstone for the thought:

> 'Once you have got hold of the rule, you have the route traced for you'. (*RFM* VI, 31; Wittgenstein's own scare quotes)

Cornerstone of rules-as-rails: the nth place and the 'something'.

which is elaborated in the 'rules-as-rails' imagery of *Investigations* §§218–9. For if the rule governing a particular arithmetical series, say, really does determine its every *n*th place quite independently of any judgement of ours, then—since any feasible explanatory performance will allow of alternative interpretations generating disputes about the identity of *n*th places which were not explicitly covered—the every-*n*th-place-determining 'something' which someone who follows the explanation correctly comes to have in mind—the 'essential thing' which we have to 'get him to *guess*' (*Investigations* §210)—must somehow have been at best clumsily and imperfectly conveyed. The idea which Wittgenstein is opposing when he writes:

> If you use a rule to give a description, you yourself do not know more than you say. . . . If you say 'and so on', you yourself do not know more than 'and so on'. (*RFM* IV, 8)

What Wittgenstein opposes follows from rails plus shared uptake.

is hence a direct consequence of the rules-as-rails picture, joined with the supposition that, by dint of ordinary training and explanations, we do indeed succeed in perpetuating a community of mutual understanding of arithmetical and other rules.

The second theme is:

> *It might be preferable, in describing one's most basic rule-governed responses, to think of them as informed not by an* intuition *(of the requirements of the rule) but a kind of* decision. (*Investigations* §§186, 231; *RFM* VI, 24; *Brown Book* §5)

The point of contrast between the terms 'intuition' and 'decision'[22] is that the former implies, and the latter repudiates, the suggestion that following a rule is, even in the most basic cases when one can say nothing by way of justification for one's particular way of proceeding, a matter of *cognition* of an independent requirement. 'Intuition'

22. *Intuition* and *Entscheidung*.

Intuition, unlike decision, would track a requirement with no story.

suggests a primitive, unarticulated apprehension, a form of receptive knowledge too basic and immediate to admit of any further account. Such an intuitive faculty would ensure that, once we have mastered, for example, an arithmetical rule, our successive applications of it are disposed to be appropriately sensitive to its requirements—either already in line or able to be brought into line. But it seems there could be no further story to be told about how the intuitive faculty accomplishes this harmony. For precisely that reason, however, we have no means to resist a sceptical suggestion:

> If intuition is an inner voice—how do I know *how* I am to obey it? And how do I know that it doesn't mislead me? For if it can guide me right, it can also guide me wrong. (Intuition an unnecessary shuffle.) (*Investigations* §213)

No answer, yet the doubt is barely intelligible (§192).

The reason why there is no response to this is because we have done, and can do, nothing to *fill out* the thought that, in the most primitive case of rule-following, we track a set of independent requirements. But the sceptical thought is barely intelligible. For we have no accountable idea of what would constitute the direction taken by the rule by itself if the deliverances of our intuitive faculties were to take us collectively off track—'no model of this superlative fact' (*Investigations* §192). That the sceptical thought makes no even apparent sense[23] but arises inevitably out of the intuitional epistemology of rule-following, brings out the cognitive pretentiousness of the latter, and the purpose of:

> It would almost be more correct to say, not that an intuition was needed at every stage, but that a new decision was needed at every stage. (*Investigations* §186)

23. Contrast, say, Cartesian scepticism about other minds, where we precisely do have an apparent conception of the truth-conferrers for descriptions of others' mental states, based on a seemingly plausible projection from one's own case, which sets up an epistemological problem.

The third theme: having the rule in mind.

The third theme is complementary to the second and elaborates Wittgenstein's doubts about the possibility of a 'tracking' epistemology of rule-following:

> *Supposing that grasping a rule is a matter of coming to have something 'in mind', how am I supposed to recognise what its requirements are?* (*Investigations* §§198, 209–13; *RFM* VI, 38, 47)

It is in the context of this theme that the 'paradox' is presented on which Kripke's exegesis concentrates:

> '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 along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning. (*Investigations* §198)

The series example: how does having the rule in mind help?

I undergo some process of explanation—somebody writes down for me a substantial initial segment of some arithmetical series, say—and as a result, let us suppose, I come to have the right rule 'in mind'. But now consider any *n*th place in the series which lies beyond the demonstrated initial segment and which, in coming to a correct understanding of the intended rule, I nevertheless did not think about at all. How, when it comes to the crunch—at the *n*th place—does having the rule 'in mind' help? A natural response—because one tends to think of 'having the rule in mind' on the model of imagining a formula or something of the sort—is to concede that, strictly, merely having the rule in mind is no help. For it is possible to have a formula in mind without knowing what it means. So—the response continues—it is necessary in addition to *interpret* the rule. But then we immediately get the 'paradox' which Wittgenstein's interlocutor[24] blunders into in *Investigations* §198. *Any* selection for the *n*th place can be reconciled, on *some* interpretation, with the rule. An interpretation is of help to me, therefore, only if it is *correct*. But the fact that the idea of correctness has to be invoked at this point makes the play with interpretation idle. 'Knowing the correct

Any choice fits some interpretation, so interpreting is idle.

24. For note the quotation marks.

interpretation of the rule' becomes just a piece of patter equivalent to 'knowing the rule'. Or if it is not—if we try to think of it as the having of something further in mind, some expanded account, for instance, of the content of the rule—then the problem will simply recur.

An expanded account in mind just raises the problem again.

That is, if having the expanded account in mind is having an uninterpreted object in mind, then I get no guidance; but if its guidance is conditional on an interpretation, then I need to arrive at *its* correct interpretation, and an account is still owed of how I am to recognise *that*.

The dilemma for a rule that transcends its explanation.

The relations between the three themes should, I suggest, be viewed like this. Suppose that what I take up from an episode of explanation, where it is successful, does indeed essentially transcend that explanation and any other that I might give in turn. I come to have the right rule in mind, but might, save for a kind of felicity, equally well have arrived at a wrong one despite having missed no overt feature of the explanation. This idea, challenged in the passages cited in connection with the first theme, connects with the second and third themes in that they jointly confront it with a dilemma. How does the explanation-transcendent rule which I supposedly have 'in mind' tell me what to do in novel cases? If I have to interpret it, that could be done in lots of ways. So how do I tell which interpretation is correct? Does that call for a further rule—a rule for determining correct interpretation of the original—and, if so, why does it not raise the same problem again, and so generate a regress? If, on the other hand, I don't have to interpret the original rule, then the only possible answer appears to be that I have some unmediated, intuitional contact with its requirements—the thought challenged by the second theme.[25] The upshot is, accordingly,

25. There is some question, I think, whether the third theme really strikes home at what someone is likely to be thinking who is working with the conception of the rule as explanation-transcendent but manifest to the subject. One is—I suppose this is a point of autobiography—much more likely to be thinking in terms of a kind of *inner illumination* than in terms of having the rule as, literally, an object of consciousness. But the connection between the inner illumination and the step-by-step judgements involved in applying the rule remains irremediably intuitional, in the sense that opens it to the sceptical attack of the second theme. For an elaboration of the attack, see Wright (1980), pp. 36–8 (this volume Essay 1, pp. 28–31).

that if we attempt to construe grasp of a rule as the presence in mind of an explanation-transcendent item, as the conception of the autonomy of rules expressed in the rules-as-rails imagery dictates, we are beggared for any satisfactory epistemology of step-by-step rule-following.

Kripke's paradox isn't Wittgenstein's: §201 and §198.

How do these ideas relate to those developed by Kripke? Several commentators[26] have challenged Kripke's interpretation on the ground that the second paragraph of *Investigations* §201 makes it plain that Wittgenstein did not accept the 'paradox' which Kripke makes pivotal. These commentators might also have emphasised the point, noted above, that the paradox is first presented, in §198, in 'interlocutor-quotation', as a suggestion for correction. And a third point of disanalogy, not generally noted, is that the focus of the regress-of-interpretations paradox of §§198–201 does not coincide with that of Kripke's sceptic. Kripke's sceptic challenges his interlocutor to substantiate a claim to know what rule the interlocutor formerly followed—he is to call attention to facts concerning his former behaviour and/or mental life which nail the rule down. So the problem is to describe a cognitive pathway *to* the (former) rule. By contrast, the regress-of-interpretations paradox focuses on a particular conception of the path *from* the rule to a judgement about its proper application in a new case. The rule is assumed to be in place—'in mind'; the question is, how does it help me to have it there?[27] And the problem arises only on a certain conception of what rules are, one which conceives of the relation between receipt of an explanation and successful ongoing practice as essentially mediated by cognition of the requirements of something interiorised. So not merely do the

Kripke asks which rule I followed, the regress how it guides me.

26. For instance McDowell (1984); McGinn (1984); Baker and Hacker (1984b). See also Malcolm (1986), p. 154.

27. I am not, in presenting this contrast, forgetting that Kripke's sceptic is a device for focusing on an *ontological* issue. It remains that the device works by raising, at the relevant stage of the dialectic, an epistemological worry. The transition to an ontological conclusion is mediated by the further assumption, disputed for example by the dispositionalist, that facts about (former) rules—if they existed—would have to be accessible, at least in principle, via reflection on (former) behaviour and conscious mental episodes.

two paradoxes focus on ostensibly different—though of course connected—kinds of question concerning rules, namely:

How can I tell which rule I (used to) follow?

and,

How can I tell what the rule I grasp requires of me here?

Kripke aims at rules' existence, Wittgenstein at a misunderstanding.

In addition, while Kripke's paradox is directed at the very existence of rules and rule-following, Wittgenstein's paradox is directed, in intention anyway, at a misunderstanding of the epistemology of rule-following competences, which—he apparently believes—can be corrected without calling into question their reality.

Unless autonomy is disturbed, the themes end where Kripke does.

If that was Wittgenstein's intent, however, the question remains whether he succeeded in carrying it through. If the interiorised, explanation-transcendent rule, with its attendant hopeless epistemological difficulties is, as I suggested, the inevitable upshot of a certain conception of the autonomy of rules, then that has to be a casualty too. So unless the thought that rules are nothing if *not* autonomous in that way is somehow disturbed, the upshot of the three themes which I have been describing is going to be the same as that of Kripke's Sceptical Argument. Does Wittgenstein manage to disturb that thought?

What an alternative needs: normativity without autonomy.

Disturbing it would seem to require indicating an alternative: a conception of rules, and rule-governed practices, which allows sufficient distance between the requirements of a rule and the subject's reaction in a particular case to make space for something worth regarding as normativity, yet abrogates the spurious autonomy which generates the difficulties. It is clear enough what Wittgenstein regards as the *kind* of considerations that point us towards the right perspective on the matter. They are the considerations which constitute his fourth principal theme:

> *The foundations of language, and of all rule-governed institutions, reside not in the circumstance that we have internalised the same strongly autonomous, explanation-transcendent rules, whose requirements we*

> *then concur—or concur enough—about, but in primitive dispositions of agreement in judgement and action.* (*Investigations* §§211, 217, 242; *RFM* VI, 39)

The other way about: agreement sustains the rules.

There is no essential *inner* epistemology of rule-following. The connection between the training and explanations which we receive and our subsequent practices, although effected in ways which, no doubt, could be sustained only by conscious, thinking beings, is not mediated by the internalisation of explanation-transcendent rules that have to be guessed at. It is a basic fact about us that ordinary forms of explanation and training do succeed in perpetuating practices of various kinds—that there is shared uptake, a disposition to concur in novel judgements involving the concepts in question. The rules-as-rails mythology attempts an explanation of this fact. But the reality is the other way about: it is the agreement which sustains all rules and rule-governed institutions. The requirements which our rules impose upon us owe their existence to it.

The fatal simplification, and §241's warning against it.

This aspect of Wittgenstein's thought is familiar enough to need no further elaboration here—familiar, and ill understood. The great difficulty is to stabilise it against its natural drift to a fatal simplification: the idea that the requirements of a rule, in a particular case, are simply *whatever we take them to be.* That idea would, in effect, surrender the notion of a requirement. And Wittgenstein, it seems, explicitly cautions against it. (See *Investigations* §241; *RFM*, VII, 40.)[28] But then what is the upshot? We are told that the requirements of rules exist only within the framework of institutional activities which depend on basic human propensities to agree in judgement, but reminded that such requirements are also, in any particular case, independent of our judgements, supplying genuine norms in terms of which those judgements, even consensual judgements, may be evaluated. We have accordingly been told what does *not* constitute the requirement of a rule in a particular case: it is *not* constituted by our agreement about the particular case, and it is *not* constituted

Told what doesn't constitute a requirement, not what does.

28. But a distinction is necessary here. See p. 210 below.

autonomously, by a 'rule-as-rail', our cognitive connection with which there would be no accounting for. But we have not been told what *does* constitute it; all we have been told is that there would be no such requirements in the first place but for the phenomenon of human agreement in judgement.

Mistrust of constitutive questions: expose, don't replace.

I believe it is vain to search Wittgenstein's texts for any more concrete positive suggestion about the constitutive question. His later conception of philosophy is, indeed, conditioned by a mistrust of such constitutive questions. Thus, consensus cannot constitute the requirements of a rule because we do, on occasion, make use of the notion of a consensus based on ignorance or mistake. But we must guard against the tendency to erect out of a practice in which that notion is given content the mythological picture of rule-following challenged by the first three themes. The mythological picture is at work in the Platonist philosophy of mathematics, and in the ease with which it comes to us to think that a private linguist could establish objective standards of correctness and incorrectness for himself. So it is important to expose it. But, once exposed, it does not need to be supplanted. Any further urge for clarity can be gratified only by a kind of natural-historical *übersicht* of rule-governed institutions and practices.

The official line is hardly friendlier to the Project.

I mean the foregoing to be recognisable as an 'official' Wittgensteinian line.[29] The most striking thing about it for present purposes is that it is hardly more hospitable to the Central Project than Kripke's Sceptical Argument. The Central Project is fuelled by the demand for a *deep* explanation of our ability to recognise the meanings of novel utterances. The Sceptical Argument entails that there is no such ability, since there is nothing to recognise. So the demand for deep explanation is misplaced; and insofar as candidate deep explanations advert to covert structures of tacitly known rules, they advert to nothing real. The 'official' response, by contrast, allows us to think, as we ordinarily do, that there is some fact about what a particular novel

29. I am thinking especially of Baker and Hacker's exegesis (1983a, 1983b) of the *Investigations*. But it is a nice question whether, if at all, John McDowell (1984), in his conception of semantic facts as kind of bedrock, departs from the official line.

utterance, used in a particular context, means: a fact which it is possible to mistake, and which is normative with respect to appraisals of and responses to that particular utterance in that context. So facts about meaning are safe. But it does not follow that we may immediately raise the question which drives the Central Project. It may be granted that, in any ordinary sense, our agreement about the meaning of a novel utterance, like our agreement about a hitherto unconsidered *n*th place in some arithmetical series, is a rule-guided agreement. But it is quite another thing to accept that it calls for an account in terms of covert cognitive operations. The whole lesson of the negative parts of Wittgenstein's discussion was—wasn't it?—that not *all* human agreement in judgements can coherently be thought of as a product of covert cognitive operations. Phenomenologically, at any rate, construal of a novel utterance is often immediate and spontaneous; what obstacle is there to the suggestion that agreement about content is, in such cases, in no way different to agreement about hitherto unconsidered places in simple mathematical series? If some of our judgemental agreement neither calls for nor admits of any further cognitive account, why should not our agreement about the content of novel utterances come into that case?

Facts about meaning are safe, but no call for covert operations.

Construal is often immediate, like the next place in a series.

To stress: the transition from linguistic training to linguistic competence may be mediated in ways which are possible only for certain sorts of conscious, thinking beings. But that is not to say that it is informed by rules—whether put there by the training or, in part at least, in place already—of which linguistic competence works by repeated but unconscious consultations. If we reject the constitutive question—what makes it the case that this is a correct application of the rule at this point? what makes it the case that this novel utterance means just that?—the phenonenon to be explained, it seems, can only be our agreement in judgements on the relevant point, and on similar matters. And if one message of Wittgenstein's discussion is that it is philosophically misguided to ask for an account of that in general, what is special about the particular case of agreement about the content of novel utterances? Can the sort of story which the generative grammarian tells be given without commitment to the mythology at

If the constitutive question goes, what's special about content?

which the negative part of Wittgenstein's discussion is directed? The 'official' Wittgensteinian view is apt to encourage a dismissal of the Central Project altogether.[30]

III

A third way, neither Kripkean nor official: don't despair too soon.

I want to canvass a third possibility: a response to the central insight of Wittgenstein's discussion of rule-following which is neither Kripkean nor 'official'. It may be that the 'official' view is exegetically correct, and that I here part company with the intentions of the actual, historical Wittgenstein. But it seems to me that it is an important methodological precept that we do not despair of giving answers to constitutive questions too soon; if the accomplishments of analysis in philosophy often seem meagre, that may be because it is difficult, not impossible.

The question: objective without a tracking epistemology?

The rule-following considerations attack the idea that judgements about the requirements of a rule on a particular occasion have a 'tracking' epistemology, answer to states of affairs constituted altogether independently of our inclination to make those judgments. How can judgements lack a substantial epistemology in this way, and yet still be *objective*—still have to answer to something distinct from our actual dispositions of judgement?

The parallel: secondary qualities, and irrealism as overreaction.

A good example of a broadly parallel problem is provided by secondary qualities of material objects—qualities of colour, taste, smell, palpable texture, audible sound and so on. It is an old idea that, in our judgements concerning such qualities, we respond more to aspects of our own affective phenomenology than to anything real in nature, and there is a corresponding perennial temptation towards an irrealist construal of such judgements.[31] But the irrealist response is, in this case, an overreaction. What may be true, I believe, is that (a

30. For elaboration, see Baker and Hacker (1984a), pp. 345–56.

31. Famously succumbed to by Locke, of course, and more recently by J. L. Mackie (1976), pp. 17–20.

The order-of-determination test: best judgement and truth.

large class of) judgements of colour, for instance, fail what I have elsewhere called the *order-of-determination* test.[32] Judgements of shape, by contrast, to take the most often discussed example of a Lockean primary quality, arguably pass the test. The order-of-determination test concerns the relation between *best* judgements—judgements made in what are, with respect to their particular subject manner, *cognitively ideal* conditions of both judge and circumstance—and truth. Passing the test requires that there be some content to the idea of best judgements *tracking* the truth—the determinants of a judgement's being true and of its being best have to be somehow independent. Truth, for judgements which pass the test, is a standard constituted independently of any considerations concerning cognitive pedigree. For judgements which fail the test, by contrast, there is no distance between being true and being best; truth, for such judgements, is constitutively what we judge to be true when we operate under cognitively ideal conditions.

Extension-determining against extension-reflecting.

The contrast, then, is between judgements among which our best opinions *determine* the extension of the truth predicate, and those among which they at most reflect an extension determined independently—henceforward *extension-determining* and *extension-reflecting* judgements respectively. So expressed, it is an intuitive and inchoate contrast, which can doubtless be elaborated and refined in a variety of ways. To fix ideas, let us look a bit more closely at the way matters might proceed first in the cases of (primary) colour and (visually appraisable, three-dimensional) shape, and then in the case of psychological characteristics.

The royal blue surface, and the C-conditions for colour.

Consider a plane surface one foot square. What conditions on the circumstances of judgement should we impose in order to ensure that a thinker's (visual) judgement will be that the surface is, say, uniformly royal blue only if it is? Well, the surface must be in full view and in good light, relatively stationary, and not too far away; and the thinker must know which object is in question, must observe it attentively, must be possessed of normal visual equipment and be other-

32. In Wright (1993).

wise cognitively lucid, and must be competent with the concept *blue*. In addition, the thinker must be free of doubt about the satisfaction of any of these conditions—for a doubt might lead to an unwillingness to make any judgement, or even to the making of some bizarre, compensatory judgement, in circumstances which were otherwise ideal for the appraisal of colour.

It's a priori that colour and C-conditioned judgement covary.

Now, it is presumably necessary, in order for our judgements, when appropriately constrained, (partially)[33] to *determine* the extension of some concept, that it be a priori true that the concept applies when, so constrained, we judge that it does. And it is, I suggest, a priori true that when all the foregoing conditions are met, the fact of the object's colour—at least at the level of refinement captured by a predicate like 'blue'—and the thinker's judgment of the fact will, as it were, covary.[34]

33. 'Partially' because we are, in effect, considering something of the form:

If S judges under conditions C, then (P if and only if S believes P),

which says nothing at all about the truth conditions of P-type propositions under non-C-conditions. But we cannot plausibly consider, for present purposes, the stronger *basic equation* (to use Mark Johnston's term), namely:

P if and only if (if S judges under conditions C, S believes P),

unless we can foreclose on the possibility that bringing about conditions C might materially affect the truth-status of P. And that cannot be done with colour, or any characteristics sustained by a causally active and acted-upon base. Cf. note 26 in Wright (1988). For fuller discussion, see the Appendix to chap. 3 in Wright (1992).

34. Two possible doubts about the sufficiency of the listed conditions would need to be addressed before this claim could be finally sustained.

(I) Might a subject not be possessed of eccentric background beliefs—for instance, that there are no blue things at all—which would prevent his formation of the appropriate belief about the object's colour, even though he met the conditions as stated? (I am here indebted to Paul Boghossian.) It is not clear. If the object is blue, it will look blue to him under the stated conditions; and then, if he believes that

Blue things look blue to *normally visually equipped* subjects in *good light*,

he is going to be constrained by the eccentric belief—if cognitively lucid—to doubt whether both those conditions are met. So he will violate the extremal condition, of being free of doubt, etc. But if he doesn't believe the principle connecting blueness and

Three more things needed for extension-determining status.

It is another question whether such a priori covariance is sufficient for the judgements to enjoy extension-determining status. But there are, in this case, three supplementary considerations which, if correct, arguably confer such sufficiency.

First, the C-conditions mustn't be whatever it takes.

First, apriority in such a claim—a claim that, under certain conditions, C, a thinker will hold a certain belief if and only if it is true (henceforward, a *provisional equation*)—may be the product of a certain triviality, consequent on C's receiving no substantial specification but being described purely in terms of 'suitability', 'conduciveness', or, generally, as whatever it takes to appraise judgements of the relevant sort correctly. Clearly, we would have made no case for regarding best opinion as *determining* the extension of the truth predicate among a given class of judgements if, although we had constructed an appropriate kind of provisional equation whose instances held a priori true, their apriority was owing to this kind of trivial specification of the C-conditions. For it is an a priori truth of *any* kind of judgement whatever that, if I operate under conditions which have everything it takes to ensure the correctness of my opinion, then it will be the case that P if and only if I take it to be so. But the conditions listed above for the appraisal of colour allow, it is plausible, substantial, non-trivial elaboration in a manner conservative of the a priori connection between their satisfaction and the correctness of the thinker's

A whatever-it-takes equation is a priori for any judgement.

blue appearance, does he count as appropriately competent with the concept *blue*? Still, that is just one kind of eccentric belief.

(II) Background colour affects colour appearance; a cream circle, for instance, may look pink, or eau-de-Nil, even in normal light, depending on the colour of the surface behind it. May we not need to strengthen the C-conditions to contain some stipulation of an appropriately coloured background? There would be great difficulties with the important independence condition (see p. 195) if we do. But does the phenomenon affect *royal blue, scarlet, emerald green, lemon yellow* to any extent which might result in a subject's erroneously judging (or withholding the judgement) that an object meets one of those descriptions? And would it anyway be absurd to think of colour as involving an element of situation-relativity—so that the judgements about, for example, the cream circle are all correct?

opinion. We can, for instance, specify normal visual equipment on the part of the thinker as: equipment which is actually statistically usual among human beings. Likewise, good lighting conditions can be specified as: conditions like those which actually typically obtain out-of-doors and out-of-shadow at noon on a cloudy summer's day.[35]

Second, the C-conditions mustn't presuppose colour facts.

The second supplementary consideration is that the question whether the C-conditions, so substantially specified, are satisfied in a particular case is logically independent of any truths concerning the details of the extension of colour concepts. If this were not so, it would be open to question whether thinkers' opinions, formed under these C-conditions, could be extension-determining; for satisfaction of the C-conditions would always presuppose some anterior constitution of colour facts. It is here that one disanalogy opens up with the case of shape, as we shall see in a moment.

Third, no rival account from which the rest would follow.

The final supplementary consideration is that there is to hand no *other* account of what does determine the extension of the truth predicate among simple judgements of colour, of which the apriority of provisional equations of the kind in question, whose C-conditions are substantially specified and, in the requisite way, free of logical presupposition about the extension of colour concepts, would be a derivable consequence. So there is, to put the matter another way, no *explaining away* the case which the other considerations supply for saying the judgements formed under the conditions in question are extension-determining rather than extension-reflecting.

The suggestion, in summary, is that there is at least a strong prima facie case for regarding a base class[36] of our best judgements about

35. The occurrences of 'actually' in these two specifications are to be understood as securing rigidity of reference to the status quo. So counterfactual situations in which other states of affairs are statistically usual are not C-conditions as specified.

36. A base class, rather than merely a proper subclass, because our beliefs about objects' colours under non-C-conditions are variously constrained by the characteristics of colours as determined by C-conditioned judgements. It is, for instance, conclusive justification for the belief that something is blue in the dark to justify the claims (i) that we would judge it to be blue if we saw it in good light and under the other C-conditions; and (ii) that bringing these conditions about would effect no changes in

The four conditions summed up, (i) to (iv).

colour as extension-determining. The case consists in the circumstances (i) that we can construct a priori true provisional equations for such judgements; (ii) that the C-conditions in these equations can be substantially specified, in a manner free of the triviality associated with whatever-it-takes formulations; (iii) that the satisfaction of the C-conditions is, in any particular case, conceptually independent of the details of the extensions of colour concepts; (iv) that no other account is available of what else might determine the extension of the truth-predicate among judgements of colour, of which the satisfaction by the relevant provisional equations of conditions (i)–(iii) would be a consequence.[37]

any determinable (contrast: determinate) aspect of the object which would need to be mentioned in an explanation of the form which our C-conditioned response would assume.

A connected point, which may help to forestall possible confusion, is that the relationship between the characteristic being-judged-to-be-blue-under-C-conditions and being blue is not to be compared to that between the characteristic marks of the instance of a natural kind and being an instance of that kind. It is true that the marks of gold, for example—its colour, lustre, heaviness, resistance to corrosion, etc.—are in some sense a priori; we do not learn what gold is and then discover that it has these characteristics. But in the case of genuine natural kinds we are (a) open to the discovery that some things which have the marks whereby we succeeded in identifying the kind are not actually instances of the kind; and (b) open to the discovery that actually no genuine kind of thing *is* individuated by those marks. By contrast, if what I have been suggesting about colour is correct, we are not open to the discovery that some objects judged to be blue under C-conditions are not blue; nor open to the discovery—should the microphysics of blue things prove bizarrely heterogeneous—that there is no such thing as an object's being blue.

37. It is not inconsistent with thinking of best judgements about colour as being extension-determining simultaneously to hold that the extensions of colour concepts are determined by microphysical characteristics of objects—supposing that the physics *doesn't* prove bizarrely heterogeneous. For it will be best judgements about colour which determine *which* microphysical characteristics are fit to play a (supplementary) extension-determining role. And no obstacle to colour's satisfying condition (iv) is posed by this view of the relationship between colour and the microphysical, since the microphysical account of the determinants of the extension of 'blue', for example, will not, presumably, entail that appropriate provisional equations can be formulated satisfying conditions (i)–(iii).

Now shape, pear-shaped x: several observations needed.

Contrast now the situation of shape. Suppose x is some nearby middle-sized object, and consider the judgement, 'x is pear-shaped'. We will want to characterise the conditions which are cognitively ideal for the visual appraisal of x's shape in terms very similar to those suggested for the case of colour. But there are differences. One, relatively unimportant, is that the lighting conditions do not have to be as good; sodium street-lighting, for instance, is suitable enough for recognising shapes. But a second, much more important, consideration is that a single thinker's *best* opinion about three-dimensional shape, if visually grounded, must needs be the product of *several* observations, from a suitable variety of spatial positions. And in order for it to be a priori true that, subject to whatever other conditions we wish to impose, such a subject's opinion about the shape of x will be correct, we must ensure that no *change* in x's shape takes place through the period of these several observations. But that calls for some ingredient in the C-conditions of which it is an *a priori* consequence that whatever it is true to say of x's shape at any time during the subject's observations is also true at any other time within the relevant period. Some independent determinant is therefore called for of what it *is* true to say about x's shape during that period—independent, that is, of the opinion formed by the thinker. There is accordingly, it seems, no immediate prospect of a provisional equation for 'x is pear-shaped' meeting both conditions (i) and (iii) above.

No change meanwhile: so an independent determinant of shape.

Teamwork filters instability, but operational criteria run deeper.

A natural response is that there is no reason why a particular kind of subject-matter might not dictate that the formation of a best opinion required *teamwork*. What if we discharge the single observer and consider instead the opinion concerning x's shape which would be arrived at cooperatively by a number of strategically positioned subjects who observed x at the same time? That should filter out the problem of instability. But there is a deeper problem which brings out, I think, the real point of disanalogy between shape and colour. The application of shape predicates, even ones as rough and ready as 'pear-shaped', is answerable to a variety of considerations besides visual appearance. For instance, a solid is pear-shaped only if any maximal two-dimensional section of it describes two contiguous circles of substantially different sizes. But it cannot be an a priori truth that

Müller-Lyer, and why shape's reliability is only a posteriori.

conditions of the kind which we regard as optimal for the visual detection of shape are adequate for the reliable visual detection of characteristics which are in this way answerable to such operational considerations. The operational criteria dominate the visual—that is why the Müller-Lyer illusion is a *visual* illusion. And we can well enough imagine starting out again, as if at the dawn of our intellectual history, armed with the concepts which we have now but with no experience of the world, and finding that the cost of maintaining the thesis that reliable visual appraisals of shapes are generally possible would be a disorderly plague of hypotheses about changes in shape, forced on us by the need to reconcile our visual appraisals with operational ones. It is an a posteriori courtesy of experience that the world in which we actually live is not of this awkward kind. And if that is right, then no kind of true provisional equation for visual judgements of shape, even if it invokes teamwork and thereby meets the independence condition (iii), can be both a priori and substantial.[38]

I am not, of course, presenting these remarks as establishing conclusively that (a base class of) judgements of colour fail the order-of-determination test, while judgements of shape pass it.[39] Nor is it clear that the considerations advanced, even as far as they go, will generalise to other instances of the primary/secondary distinction. My pur-

38. This train of thought is elaborated somewhat in Wright (1988), pp. 19–20.

39. Various interesting further questions have to be negotiated which I cannot go into here. For instance, might best *operationally determined* (contrast: visually determined) beliefs about shape arguably play an extension-determining role, or are even such beliefs at most extension-reflecting? One germane consideration would be that very many shape concepts—'pear-shaped' is an example—are, while operationally constrained, subject to no obvious operational, as opposed to tactuo-visual criteria of application. Another would be that we would expect to need stability provisos in the appropriate lists of C-conditions for operationally determined beliefs, since proper execution of the appropriate kinds of operation is essentially subject to invalidation by instability. It would be pleasant if that consideration were decisive, but I do not think that it is. For it may perhaps be mitigated by some play with the *positive-presumptiveness* of assumptions of stability, along the lines illustrated in the next section of the essay for the role of lack of self-deception in the C-conditions for appraisal of one's own intentions.

pose is, rather, to indicate one framework for the discussion and development of the test: to give credence to the thought that there is a distinction here—perhaps a number of distinctions—and an explanatory programme of potential importance.

Colour as a mix: not tracked, yet a whole community can be wrong.

Nevertheless, if the gist of the foregoing is correct, visual judgements of colour will emerge as an interesting mix of subjectivity and objectivity. They should not, at least in a basic class of cases, when appropriate C-conditions are met, be regarded as responsive to states of affairs which are constituted independently—our best opinions about colour do not, in that sense, *track* colour. But neither is it the case that there is no standard to meet, that whatever we say about colour goes or—what comes to the same thing—that there is no such thing as an object's real colour. Rather, it is a perfectly objective question what, in a particular case, the deliverance of best opinion would be; and that deliverance is something with which a majority, or even a whole community, may for some reason be out of accord.

IV

Second candidate: avowals, and the Cartesian inner-perception model.

I pass now to the second set of potential examples which I want to air, namely, self-ascriptions of psychological states like sensation, emotion, mood, belief, desire, and intention—the traditional category of *avowals*. The proposal that such judgements are extension-determining is an extremely attractive one. The traditional Cartesian epistemology attempts to construe avowals in general on the model of a certain conception of what is involved in competent self-ascription of *sensation*—a model which draws heavily on a comparison with perception of material objects, but with the crucial differences that the sensation, *qua* 'inner' object, is accessible only to the subject, whose gaze is conceived as all-seeing and error-proof. As a picture of knowledge of one's own sensations, this generates problems enough, principally by its presupposition of the operability of private schemes of classification—'private languages'. But, even if we let the Cartesian account of sensation pass, it provides only the most feeble basis for an

account of the self-ascription of other kinds of psychological states. For there are no plausible introspectible processes or states which are candidates to be, for instance, beliefs and intentions. And besides—and more important—to think of, for example, intention as an introspectible episode in consciousness generates no end of difficulties and contortions when we try to make sense of the necessarily holistic character of the scheme of beliefs, desires and intentions by reference to which we explain subjects' behaviour, and of the notion of *fit* between an intention and the behaviour which implements it. One of the most central themes in Wittgenstein's later philosophy of mind is the idea that Cartesianism is based on a *grammatical* misunderstanding, a misinterpretation of the language-game of self- and other-ascription of mental states. The Cartesian takes the authority of avowals as a symptom f, as it were, a superlatively sure genre of detection. We should accomplish a very sharp perspective on the sense in which this is a 'grammatical' misunderstanding if it could be shown that avowals fail the order-of-determination test—in other words, that subjects' best opinions determine, rather than reflect what it is true to say about their intentional states, with the consequence that the notion of detection or 'inner tracking', as it were, is inappropriate. (Naturally, it would have to be part of a satisfying development in this direction that best opinions about one's own intentional states turned out to be very easily accomplished and normal; otherwise, the authority of avowals would be unaccounted for.)

Nothing introspectible is the intention, and holism.

Cartesianism as grammatical misunderstanding: no inner tracking.

Conditions for intention: self-deception as the main obstacle.

Can we, then, provide a set of conditions whose satisfaction will ensure, a priori, that subjects will believe themselves to have, for example, certain particular intentions if and only if they do? Well, how might a subject who had the conceptual resources to form a belief appropriate to the presence, or absence, of a certain intention, nevertheless fail to do so? Self-deception is one possibility, whatever the correct account of that puzzling idea. A subject may, as many people think, be simply unable to bring to consciousness the real intentions which inform certain of his courses of action. Conversely, we are familiar with the kind of weak-mindedness which can lead subjects into deceiving themselves that they have formed certain

intentions—usually ones which are desirable but difficult of implementation—when in truth they have not done so.

The three conditions: concepts, no self-deception, attentiveness.

In both these kinds of case, we typically regard the self-deceptive (lack of) belief as *motivated*. But in certain other circumstances it would be better to think of it as having a primarily physiological—perhaps a pharmacological—explanation. The cause of a subject's mistaken belief about her intentions need not reside in other aspects of her intentional psychology. Anyway, one way or another, an appropriate set of C-conditions will have to ensure that nothing of this kind is operative. And in addition, we need only include, it seems, a condition to the effect that the subject be appropriately attentive to the question what her intentions are. However, nothing in what follows will depend on whether these three conditions—grasp of the appropriate concepts, lack of any material self-deception or anything relevantly similar, and appropriate attentiveness—do indeed suffice.

The difficulty: can the no-self-deception clause be substantial?

The most salient difficulty, for our purposes, with the provisional biconditional which will emerge from these suggestions is that of making a case that it meets condition (ii)—the condition that the C-conditions be substantially specified. The motive for condition (ii) was not a distaste for triviality as such, it will be remembered, but rather for the particular kind of triviality involved in 'whatever-it-takes' formulations. Such formulations are always possible and, by prejudicially representing matters in terms of the jargon of tracking, leave us with no way of getting at the distinction which it is the point of the order-of-determination test to capture. And on the face of it, unfortunately, we are some considerable distance from a formulation of the no-self-deception condition which can count as non-trivial in the relevant sense. The problem is, first, that 'self-deception' covers, for the purposes of the biconditional, *any* motivated condition which might lead to a subject's ignorance or error concerning his or her intentions; and, second, we need, as noted, to allow for the possibility of *un*motivated conditions—chemically induced ones, or whatever—with the same effect. So we seem to be perilously close to writing in a condition to the effect that the subject be 'free of any condition which might somehow impede her ability

It has to cover any impediment, motivated or chemical.

reliably to certify her own intentions'. And that, of course, is just the sort of insubstantial, whatever-it-takes formulation which condition (ii) was meant to exclude.

No non-trivial description of the excluded states comes to mind.

Perhaps it is possible to do better, to produce some descriptions which exclude the relevant class of states but does so non-trivially. But none comes to mind, and it would of course be pointless to wait on the deliverances of empirical science if one is hoping for vindication of an extension-determining view of the beliefs expressed by avowals, and so requires something which will subserve the *a priority* of the resulting biconditional. So how to proceed?

The variant: no-self-deception is positive-presumptive.

We have, I think, to depart somewhat from the approach which emerged in the case of colour. But a possible variant of it is suggested by the reflection that the troublesome no-self-deception condition is *positive-presumptive*. By that I mean that, such is the 'grammar' of ascriptions of intention, one is entitled to assume that a subject is *not* materially self-deceived, or unmotivatedly similarly afflicted, unless one possesses determinate evidence to the contrary. Positive-presumptiveness ensures that, in all circumstances in which one has no countervailing evidence, one is a priori justified in holding that the no-self-deception condition is satisfied, its trivial specification notwithstanding. Suppose, then, that we succeed in constructing an a priori true provisional biconditional:

$$C(\text{Jones}) \rightarrow$$
$$(\text{Jones believes he intends to } \phi \leftrightarrow \text{Jones intends to } \phi),$$

Delete the clause: the restricted biconditional is a priori credible.

where C includes the (trivial) no-self-deception condition but no other trivially formulated conditions. Then if—lacking evidence to the contrary—we are a priori justified in holding the no-self-deception condition to be met, we are also a priori justified in believing the result of deleting that condition from the provisional biconditional in question. Likewise for any other positive-presumptive conditions listed under C. In this way we can eventually arrive at a restricted provisional biconditional in which all the C-conditions are substantially specified and which, in the absence of any information bearing on

whether the conditions are satisfied which we have deleted from it, it is a priori *reasonable* to believe.

Why credible, if the intention is fixed apart from the belief?

It is true that we are now dealing with something a priori credible rather than a priori true. But the question still arises: what is the *explanation* of the a priori credibility, in the relevant kind of circumstances of ignorance, of the restricted version if what determines the fact of Jones's intention, under the residual C-conditions, is something quite detached from his belief? The explanation cannot be that the C-conditions are trivially formulated, for they are all, by hypothesis, substantial.

The answer: opinion plays a defeasible extension-determining role.

Suppose that the three conditions suggested above—possession of the appropriate concepts, attentiveness, and lack of self-deception—do indeed suffice a priori for Jones's opinion about his intention to covary with the facts. But suppose also that possession of the relevant concepts and attentiveness raise, unlike lack of self-deception, no problems of triviality. Then the matter for explanation is: why is it a priori reasonable to believe that, provided Jones has the relevant concepts and is attentive to the matter, he will believe that he intends to ♀ if and only if he does? The key thought of the variant approach will be that the matter will be nicely explained if the concept of intention works in such a way that Jones's opinions, formed under the restricted set of C-conditions, play a *defeasible* extension-determining role, with defeat conditional on the emergence of evidence that one or more of the background, positive-presumptive, conditions are not in fact met.

Unlike colour, conditions that warrant acceptance, not covariance.

To elaborate a little. There are, in contrast to the case argued for colour, no conditions which can be characterised non-trivially but independently—in the sense of condition (iii) above—whose satisfaction a priori ensures covariance of a subject's beliefs about his intentions and the facts. But there are such non-trivially, independently specifiable conditions whose satisfaction ensures, courtesy of no a posteriori background beliefs, that, failing any other relevant information, a subject's opinions about his or her intentions should be accepted. And the proposed strategy of explanation is, roughly, as follows. What determines the distribution of truth values among

ascriptions of intention to a subject who has the conceptual resources to understand those ascriptions and is attentive to them are, in the first instance, nothing but the details of the subject's self-conception in relevant respects. If the assignment of truth values, so effected, generates behavioural singularities—the subject's behaviour clashes with ingredients in her self-conception, or seems to call for the inclusion of ingredients which she is unwilling to include—then the self-deception proviso, broadly interpreted as above, may be invoked, and the subject's opinion, or lack of it, overridden. But that is not because something is shown, by the discordant behaviour, about the character of some *independently constituted* system of intentions which the subject's opinions at best reflect. When possession of a certain intention is an aspect of a self-conception that coheres well enough both internally and with the subject's behaviour, there is nothing *else* that makes it true that the intention is indeed possessed.

Truth fixed by the self-conception, unless behaviour clashes.

The minimalist view, and the authority of avowals explained.

The view proposed is minimalist. Nothing leaner has any prospect, so far as I can see, of accommodating both the avowability and the theoreticity of intention. To be sure, explaining the a priori reliability of a subject's C-conditioned beliefs about her intentions will do nothing to explain the reliability of her avowals—even assuming our right to regard her as honest—unless the C-conditions in question are likely to be met. But there seems no cause to anticipate problems on that score. Attentiveness—however precisely it should be elaborated—is presumably, like lack of self-deception, a positive-presumptive condition; and a subject's possession of the appropriate concepts is prerequisite for her being able to effect the avowal in the first place. So there is every promise of a straightforward kind of explanation of the authority which avowals of intention, *qua* avowals, typically carry.

Suppose, by contrast, that subjects' best opinions about their intentions are at most extension-reflecting. Then the a priori reasonableness, when nothing else relevant is known, of the restricted provisional biconditional needs another explanation. And providing one will require explaining how it coheres with the view that subjects' best opinions track independently constituted states of affairs to suppose it a priori reasonable to think, failing evidence to the contrary,

that some of the conditions on an opinion's being best are satisfied. Why, in the absence of germane evidence, is *agnosticism* about the satisfaction of, for example, the self-deception condition not a preferable stance? The three conditions collectively formulate, on this view, what it takes to ensure a certain kind of cognitive accomplishment, a feat of detection. Why is it a priori warranted to assume, failing information to the contrary, that a subject satisfies any such conditions?

To avoid misunderstanding, I do not mean to present the question as rhetorical. There are various ways in which it might be approached by an opponent of the extension-determining view. Perhaps it could be made out that the warrant somehow flows from the subject's nature, *qua* subject, or from the nature of the states of affairs—the intentions in question are *hers*, after all—of which the conditions ensure her detection. Alternatively, perhaps other examples can be produced where best opinions are extension-reflecting and yet certain of the conditions on their being best are likewise positive-presumptive, for general reasons which might be argued to apply to the present case. But it is fair to say, I think, that the onus is now on someone who prefers to think that one's best opinions about one's own intentions, etc., are extension-reflecting. Why are we, apparently, so cavalierly optimistic about our general fitness for such detection? And what, fundamentally, when we succeed in holding best opinions, *does* determine the extension of the truth predicate among the class of judgements in question, if not those opinions themselves?

A corollary: back to the complaint against Chomsky and dispositions.

My purpose in introducing the psychological was only the limited one of canvassing a second shape which the attempt to refine and apply the order-of-determination test might assume. Perhaps enough has been said to accomplish that. But note a prospective corollary, if the extension-determining character of subjects' best opinions about their intentions and similar states, and the authority typically carried by avowals of such states, can indeed be accounted for along the lines proposed. Earlier, I criticised responses to Kripke's Sceptical Argument which—like the dispositional conception, or Chomsky's own proposal—locate the called-for meaning-constitutive facts only at the cost of obscuring subjects' non-inferential knowledge of their

own meanings. It would be appropriate to level a similar complaint against dispositional or theoretical construals of the notion of intention. But no such complaint is appropriate against an account of self-knowledge of intention further developed along the lines canvassed: one according to which subjects' best opinions about their intentions, *both past and present*, are properly conceived as provisionally extension-determining, and which explains how and why the opinions which they typically hold are indeed best. It will be, similarly, a perfect answer to Kripke's sceptic to explain how judgements concerning one's own meanings, both past and present, are likewise provisionally extension-determining in the most ordinary circumstances. Challenged to justify the claim that I formerly meant addition by 'plus', it will not be necessary to locate some meaning-constitutive fact in my former behaviour or mental life. A sufficient answer need only advert to my present opinion, that addition is what I formerly meant, and still mean, and to the a priori reasonableness of the supposition, failing evidence to the contrary, that this opinion is best.

The minimalist account doesn't obscure self-knowledge.

The perfect answer to the sceptic: my present opinion suffices.

Responding to Kripke's sceptic in this way does not require construal of meaning precisely as a species of intention; it is enough that the concepts are relevantly similar—that both sustain authoritative first-personal avowals,[40] and that this circumstance is to be explained in terms of failure of the order-of-determination test. However, I can go no further into the matter here except to record the view that it is indeed with the conception of meanings as items which our best opinions may *reflect*, rather than with their reality *tout court*, that the Sceptical Argument engages.[41]

The sceptic engages meanings as reflected, not their reality.

40. Note that the most likely reservation about the claim that they do—the thought that certain elements of *convention*, which have no counterpart in the case of intending in general, sustain my ability to mean anything in particular by a word—will undercut the sceptic's strategy in any case, since restricting the search for facts about my former meanings to aspects of my former behaviour and conscious mental states will precisely exclude the relevant conventional (presumably social) elements. I am here indebted to Bob Hale.

41. For further discussion, see Wright (1987a) (this volume Essay 5).

V

Double success: the situation and the rule, castling in chess.

Correctly applying a rule to a new case will, it is natural to think, typically involve a double success: sensitivity is necessary both to the relevant features of the presented situation and to what, in respect of those features, will fit or fail to fit the rule. Correctly castling in the course of a game of chess, for instance, will depend both on apprehension of the configuration of chessmen at the time of the move, and on a knowledge of whether that configuration (and the previous course of the game) permits castling at that point. In this respect the simple mathematical examples on which Wittgenstein's discussion concentrates are apparently untypical. Knowing how to continue a series does of course require knowledge of what point in the series one has reached, and also in certain cases some knowledge about the preceding constitution of the series beyond the identity of the immediately preceding element. But there need be no *perceptual* input, corresponding to perception of the board configuration in chess. Wittgenstein's examples thus approximate as closely as possible to cases of *pure* rule-following, where only one kind of sensitivity—that concerning the requirements of the rule—is necessary for success, and failure has to be attributed to an imperfect understanding of the rule or to some technical slip in applying it. That is, presumably, why he selects just those examples. In order to focus on questions concerning the objectivity and epistemology of rule-following, it is obviously best to consider cases where our judgements about what accords with the rule are most nearly pure and unconditioned by the cognitive effects upon us of the prevailing circumstances.

Wittgenstein's series examples need no perceptual input.

So they approximate pure rule-following, which is why he picks them.

No pure case: a slip needn't show a misunderstood rule.

Absolute purity is, however, impossible. The judgement that *this* move accords with a particular rule can never be informed only and purely by one's understanding of the rule. So much is clear when one reflects that aberrant performance, even in simple arithmetical cases and when there is no question of insincerity, can always in principle be explained in ways which conserve the hypotheses that the subject correctly understands the rule and, relative to what she took to be the

situation, applied it correctly. The subject may, for instance, have lost her place, misremembered what she had done before, or misunderstood some relevant symbol not involved in the expression of the rule.

Rule-following as holistic, and the R-informed responses.

The ability successfully to follow a rule is thus to be viewed as, at each successive instance, the product of a number of cognitive responses which interact holistically in the production of the proper step. And some of these responses—correctly perceiving the set-up on the chess board, for instance, or recollecting the expansion of the series to this point—do not strictly pertain to the rule but are possible for subjects who have no inkling of it. Where R is the rule or set of rules in question, let us call the others 'R-informed'. Now, an R-informed response need not be encapsulable in any judgement which the subject can articulate distinct from the output judgement, as it were—the judgement into which her R-informed and non-R-informed responses conjointly feed. In that respect, the chess example, in which the R- and non-R-informed components could be respectively explicitly entertained as the major and minor premises for a modus ponens step, is untypical. I cannot always have concepts *other* than those whose governing rules I am trying to observe in a particular situation in terms of which I can formulate a separate judgement of the input to which these rules are to be applied. So I cannot always extricate and articulate a judgement which, conditionally on such a separate judgement of the input, formulates my impression of the requirements of the rules in a fashion which is neutral with respect to the correctness of my R-*un*informed responses to the situation.

Chess is untypical: usually the R-informed part can't be extricated.

The fundamental question, and why holism counsels caution.

It is a platitude that the existence of a rule requires that some sort of content be given to the distinction, among responses which are informed by it, between those which are acceptable and those which are not. The fundamental question raised by the topic of rule-following concerns the ground and nature of this distinction. However, the points about holism and inextricability counsel caution in the formulation of this question. There is, for instance, because of the point about holism, no simple relation between questions concerning what makes for acceptable R-informed responses and questions to do with *truth*,

as applied to the output judgements to which those responses holistically contribute. Indeed, we cannot always—because of the point about inextricability—equate the acceptability of a rule-informed response with the truth of any judgement made by the subject—at least, not if such a judgement has to consist in an articulate belief.

The test needs extricable judgements, which we can't expect in general.

Clearly, then, there is some awkwardness in attempting to apply the distinction drawn by the order-of-determination test to the question of the ground and nature of the requirements of a rule. The test, as so far considered, calls for a class of judgements about which we can raise the question of the relation between best opinion and truth. And the existence of such judgements is just what the inextricability point counsels us not to expect in general. Still, there are extricable cases. The example of castling in chess provided one. And, most significant in the present contest, the comprehending response to a novel utterance provides another. Such a response will involve a set of beliefs about the utterance which someone could have who had no understanding of the language in question; but it will also involve a belief about what, modulo the former set, has been said—a paradigm, it would seem, of a rule-informed judgement. Rather than confront the awkwardnesses presented by inextricability, therefore, let me concentrate for our present purposes on such favourable cases: cases where the acceptability of a rule-informed response can be seen as a matter of the truth of a judgement which the responder may be thought of as making. Our question, then, is: what makes for the truth of such rule-informed judgements?

The extricable cases: castling and the novel utterance.

First option, irrealism: Kripke's Wittgenstein.

Irrealism is an option. One could hold that such judgements answer to nothing real, not even the deliverances of best opinion. Rather, making acceptable such judgements really is just a matter of keeping in step with the community of followers of the relevant rules. This would be, in effect, the opinion of Kripke's Wittgenstein.

Consensualism: true just when we agree, §242.

A variation—no longer strictly irrealist—would hold that such judgements are true just in case there is agreement *in* them, and that this is the 'agreement in judgements' which is a precondition for the possibility of language serving as a means of communication (*Investigations* §242). It might be thought that, as an interpretation of

Wittgenstein anyway, this variant of the irrealist thought is ruled out by his explicit repudiation[42] of any identification of truth with consensus. But that would be a bad reason. Consensualism about the truth-conditions of R-informed judgements would be quite consistent with repudiating consensualism concerning the output judgements to which, in association with relevant non-R-informed judgements, they contribute on particular occasions.

Platonism: a double feat of tracking, which Wittgenstein undid.

Platonism is also a possibility. Platonism is, precisely, the view that the correctness of a rule-informed judgement is a matter quite independent of any opinion of ours, whether the states of affairs which confer correctness are thought of as man-made—constituted by over-and-done-with episodes of explanation and linguistic behaviour—or truly Platonic and constituted in Heaven. For the Platonist, correctly following a rule in a particular situation involves a double-feat of tracking: one has to detect both all relevant aspects of the situation and the relevant requirements of the rule. But the principal negative point of Wittgenstein's discussion of rule-following was precisely that the ability to make acceptable rule-informed judgements allows of no coherent construal as a tracking ability, in the fashion which Platonism requires.

The option left: best opinion constitutes their truth.

That leaves the possibility for which I have been labouring to prepare. Abandoning Platonism need not involve abandoning the objectivity of rule-informed judgements. There remains the option of regarding such judgements as extension-determining, of seeing *best* opinion as constituting their truth. I would like to be in a position to offer a supported opinion about whether this idea can be approached along the lines sketched for the cases of colour and intention. But I have, at the time of writing, no settled opinion to offer about that, let alone about whether the idea can ultimately be made good.

In summary: once it is agreed that the essential Platonist thought is that best opinion about the correctness of rule-informed judgements is extension-reflecting, and that part of Wittgenstein's purpose was, in effect, to demonstrate the untenability of that view, the would-be

42. See references on p. 188 above.

Four choices; official scorn, then three answers without Platonism.

interpreter of Wittgenstein on rules faces four broad choices. First, there is 'official' Wittgensteinianism, of which the hallmark is scorn of constitutive questions. Asked what constitutes the truth of rule-informed judgement of the kind we isolated, the official Wittgensteinian will reply: 'Bad question, leading to bad philosophy—Platonism, for instance, or Kripkean scepticism'. If official Wittgensteinianism is rejected, on the other hand, and the constitutive question is allowed, there seem to be three options, each consistent with the repudiation of Platonism. First, there is the thesis that *nothing* constitutes the truth of such judgements, and that this is an important and disturbing contrast which marks them off from judgements of other kinds. This inherently unstable view is in essence that of Kripke's Wittgenstein. Second, there is consensualism, as canvassed above. And finally there is the proposal that the truth-makers for such judgements—the facts in which their correctness consists—are constituted by the deliverances of best opinion.

VI

Only Platonism sustains it? Then side with Wittgenstein.

The concern of the Central Project was to explain our recognition of the syntax and sense of novel sentences. The judgements accomplished by such recognition are, as noted, naturally taken to be rule-informed judgements par excellence, which in turn contribute holistically, in association with other judgements of features of the presenting circumstances and with certain background beliefs, to opinions concerning the proper use, or specifically the truth or falsity, of the sentence in the circumstances. So, if the options are as just described, what now are the prospects for the thought that our recognition of (agreement in judgement about) the meanings of novel sentences calls for a deep cognitive-psychological explanation, the *ur*-thought of the Central Project? Platonism might sustain it; but if that is the only way it can be sustained, then there is after all, a real collision between Wittgenstein's thought and the Central Project, and one might well want to side with Wittgenstein. 'Official'

Consensualism's dilemma: a cognitive account makes room for error.

Wittgensteinianism, as we have seen, and Kripkean Wittgensteinianism, by contrast, both undermine the thought. And the same is presumably true of the consensualist option. If the correctness of a rule-informed judgement is simply constituted by our agreement in it, there seems to be no call for a cognitive-psychological account of that agreement. On the contrary, the provision of such an account would immediately restore the idea that, since consensus in a rule-informed judgement would be the product of certain purely cognitive accomplishments, there ought to be some fact about what in any particular case, independently of our actual dispositions of response to it, the *proper* exercise of the relevant cognitive skills and informational states would culminate in; space, in other words, for the idea that an actual consensus might, in a particular case, be misplaced.[43]

Content as Lockean secondary: elaborating the C-conditions.

With Platonism out of the way, the Central Project, as a project in cognitive psychology, would thus appear to demand the view that the correctness of the relevant rule-informed judgements—semantic responses to novel utterances—is objective but best-opinion-determined—the view that semantic content is, if you like, *secondary* in Lockean terms. If that view can be sustained, then the cognitive-psychological project is properly seen as directed not at the description of the conditions for a certain kind of tracking accomplishment—that conception has to fall with Platonism—but at the detailed elaboration of the C-conditions whose realisation ensures that a subject's judgement of the content of a particular novel utterance will be *best*. The Central Project is thus prima facie compatible with Wittgenstein's thought-about rules, if the upshot of the latter is indeed (at least compatible with the claim) that judgements of content, and of the requirements of rules in general, are extension-determining.

Whatever the proper interpretation of Wittgenstein, the outstanding substantive questions are accordingly whether the order-of-determination test can be refined and developed sufficiently to allow a clear-cut application to judgements of content, with their failure of

43. Exactly the idea enshrined in the distinction between *competence* and *performance*, of course.

Two open questions: the test for content, and a place for linguistics.

the test—their classification as extension-determining—as the outcome; and whether the then emergent notion of *best* opinion will contain components whose proper description will require—or, less, allow—invoking the apparatus of theoretical linguistics. We are a long way from knowing what is correct to think about either question. But it may be some advance to be a little clearer about some of the respects in which matters are unclear.

I am very grateful to Paul Boghossian and Bob Hale for helpful discussion of an earlier draft of this essay.