The Rule-Following Considerations[1]

PAUL A. BOGHOSSIAN

This is the fifth of our commissioned State of the Art Series

INTRODUCTION

1. Recent years have witnessed a great resurgence of interest in the writings of the later Wittgenstein, especially with those passages—roughly, *Philosophical Investigations* ##138–242 and *Remarks on the Foundations of Mathematics*, section VI—that are concerned with the topic of rules. Much of the credit for all this excitement, unparalleled since the heyday of Wittgenstein scholarship in the early 1960s, must go to Saul Kripke's *Wittgenstein on Rules and Private Language*.[2] It is easy to explain why.

To begin with, the dialectic Kripke uncovered from Wittgenstein's discussion is enormously exciting on its own terms. On Kripke's reading, the passages on rule-following are concerned with some of the weightiest questions in the theory of meaning, questions—involving the reality, reducibility, and privacy of meaning—that occupy centre-stage in contemporary philosophy. Furthermore, Kripke represented Wittgenstein as defending a set of unified and extremely provocative claims concerning these questions. And, finally, he argued for these claims with power and clarity. The ensuing flood of articles and books on the subject of rule-following was both predictable and warranted.

The present paper is the result of an invitation to survey this literature. It could have been about exegetical matters, on what the recent discussions have had to teach us about the historical Wittgenstein's philosophical views. In the event, however, it is almost entirely concerned with a retrospective assessment of the *philosophical* contributions. Limitations of space dictated that a choice be made; and the philosophical assessment seemed the more fruitful thing to do.[3] Despite a lot of discussion, there is room for an improved understanding of the precise nature of Kripke's

16. †Boghossian: The rule-following considerations

[1] I am grateful to many people for helpful discussion of the issues covered in this paper, including Mark Johnston, John Burgess, Jerry Fodor, Barry Loewer, Richard Rorty, Barry Allen, Larry Sklar, Crispin Wright, Saul Kripke, Neil Tennant, Steve Yablo, Nick White, and participants in various seminars at the University of Michigan. Special thanks are due to Paul Benacerraf, Jennifer Church, and David Velleman.

[2] Cambridge, Harvard University Press, 1982. Henceforth, 'K'.

[3] The main reason is that I have actually come to despair of a satisfactory interpretation of Wittgenstein's views. I try to say why in 'The Problem of Meaning in Wittgenstein', to appear in *Meaning Scepticism*, ed. K. Puhl, De Gruyter, forthcoming.

Mind, Vol. 98 . 392 . October 1989 © Oxford University Press 1989

arguments, of their ultimate cogency, and of their relation to the wider discussion of meaning in contemporary philosophy of mind and language. Pulling on the thread that is Kripke's argument leads quite naturally to a discussion of many of the most significant issues occupying philosophers today; in that lies the main impetus behind the present essay.

I proceed as follows. In parts I and II, I lay out the essentials of Kripke's argument. In subsequent parts, I offer an extended critique of the dialectic it presents, considered on its own terms and independently of exegetical concerns. A discussion of the critical literature will be woven in as appropriate. The moral will not be recognizably Wittgensteinian: I shall argue that, *pace* Kripke's intent, the conception of meaning that emerges is a realist, non-reductionist, and judgement-independent conception, one which, moreover, sustains no obvious animus against private language.

I
KRIPKE ON MEANING AND THE SCEPTICAL PROBLEM

The sceptical problem

2. As Kripke sees it, the burden of the rule-following considerations is that it cannot literally be true of any symbol that it expresses some particular concept or meaning. This is the now-famous 'sceptical conclusion' he attributes to Wittgenstein:

> [T]here is no fact about me that distinguishes between my meaning a definite function by '+' . . . and my meaning nothing at all.[4]

How is such a radical thesis to be supported? Kripke argues, in effect, by elimination: all the available facts potentially relevant to fixing the meaning of a symbol in a given speaker's repertoire—facts about how the speaker has actually used the expression, facts about how he is disposed to use it, and facts about his qualitative mental history—are canvassed, and found wanting. Adequate reflection on what it is for an expression to possess a meaning would betray, so Kripke invites us to believe, that that fact could not be constituted by any of *those*.

The claim is, of course, indisputable in connection with facts about actual use and qualitative phenomena; it is a familiar and well-assimilated lesson of, precisely, Wittgenstein's *Investigations*, that neither of those species of fact could, either in isolation or in combination, capture what it is for a symbol to possess a meaning. Much more important and controversial, however, is Kripke's rejection of a *dispositional* account of

[4] K., p. 21.

meaning facts. Why are facts about how a speaker is disposed to use an expression held to be insufficient to determine its meaning?

Kripke develops two sorts of consideration. First, the idea of meaning something by a word is an idea with an infinitary character—if I mean *plus* by '+', then there are literally no end of truths about how I ought to apply the term, namely to just the members of this set of triples and not to others, if I am to use it in accord with its meaning. This is not merely an artefact of the arithmetical example; it holds for any concept. If I mean *horse* by 'horse', then there are literally no end of truths about how it would be correct for me to apply the term—to horses on Alpha Centauri, to horses in Imperial Armenia, and so on, but not to cows or cats wherever they may be—if I am to use it in accord with its meaning. But, Kripke argues, the totality of my dispositions is finite, being the dispositions of a finite being that exists for a finite time. And so, facts about dispositions cannot capture what it is for me to mean addition by '+'.

The second objection to a dispositional theory stems from the so-called 'normativity' of meaning. This objection is somewhat harder to state, but a rough formulation will do for now. The point is that, if I mean something by an expression, then the potential infinity of truths that are generated as a result are *normative* truths: they are truths about how I *ought* to apply the expression, if I am to apply it in accord with its meaning, not truths about how I *will* apply it. My meaning something by an expression, it appears, does not guarantee that I *will apply* it correctly; it guarantees only that there will be a fact of the matter about whether my use of it is correct. Now, this observation may be converted into a condition of adequacy on theories of meaning: any proposed candidate for being the property in virtue of which an expression has meaning must be such as to ground the normativity of meaning—it ought to be possible to read off from any alleged meaning-constituting property of a word, what is the correct use of that word. And this is a requirement, Kripke maintains, that a dispositional theory cannot pass: one cannot read off a speaker's disposition to use an expression in a certain way what is the *correct* use of that expression, for to be disposed to use an expression in a certain way implies at most that one will, not that one should.

The contents of thought

3. But what about thoughts, intentions, and other content-bearing mental states? How do they figure in the sceptical argument? More specifically: is the sceptical thesis directed against them as well, or is it confined solely to *linguistic* representation?

It is hard to see how a convincing meaning scepticism could be confined purely to the linguistic domain, given the intimate relation between thought and language. Philosophers divide, of course, on the precise

nature of this relation and, in particular, on the question of priority: Do the semantic properties of language derive from the representational properties of thought, or is it the other way round?[5] Whatever the correct answer, however, there would appear to be no plausible way to promote a *language-specific* meaning scepticism. On the former (Gricean) picture, one cannot threaten linguistic meaning without threatening thought content, since it is from thought that linguistic meaning is held to derive; and on the latter (Sellarsian) picture, one cannot threaten linguistic meaning without *thereby* threatening thought content, since it is from linguistic meaning that thought content is held to derive. Either way, content and meaning must stand or fall together.

If a sceptical thesis about linguistic meaning is to have any prospect of succeeding, then, it must also threaten the possibility of mental meaning (or content). Of course, on a Sellarsian view, that result is automatic, given a demonstration that nothing *non-mental* fixes linguistic meaning. But on a Gricean view matters are not so simple. Since the Gricean holds that linguistic items acquire their meaning from the *antecedently* fixed content of mental states, an argument to the effect that nothing non-mental fixes linguistic meaning would leave the Gricean unmoved; he needs to be given a *separate* argument against the possibility of mental content. Does Kripke see this need and does he show how it is to be met?

Colin McGinn has argued that the answer to both questions is 'no':

> My third point . . . points up a real lacuna in Kripke's presentation of his paradox. The point is that it is necessary for Kripke to apply his paradox at the level of *concepts*; that is, he has to argue that the notion of possessing a determinate concept is likewise devoid of factual foundation . . .. It cannot be said, however, that Kripke explains how this need is to be met, how this extension of the paradox to the level of concepts is to be carried out; and brief reflection shows that the exercise is by no means trivial.[6]

I think McGinn is wrong on both counts; it will be worthwhile to see why.

In fact, the suggestion that some appropriately general thought or intention constitutes the sought after meaning-determining fact comes up early in Kripke's presentation, *before* the dispositional account of meaning is considered and found wanting:

[5] In the United States, it is the Gricean view, that linguistic expressions acquire their semantic properties by virtue of being used with certain intentions, beliefs, and desires, that is most influential; whereas in Britain it appears to be the Sellarsian (Wittgensteinian?) view that thinking is a form of internalized speaking—speech *in foro interno*, as Sellars likes to put it—that tends to predominate.

For the Gricean view see H. P. Grice, 'Meaning', *Philosophical Review*, 1957; and related papers. See also, S. Schiffer, *Meaning*, Oxford, Clarendon Press, 1972. For the Sellarsian view see his 'Empiricism and the Philosophy of Mind', in his *Science, Perception and Reality*, London, Routledge and Kegan Paul, 1963. For a debate on the priority question see 'The Chisholm–Sellars Correspondence', in *Intentionality, Mind and Language*, ed. A. Marras, Urbana, University of Illinois Press, 1972.

[6] Colin McGinn, *Wittgenstein on Meaning*, Oxford, Basil Blackwell, 1984, pp. 144–6.

This set of directions, I may suppose, I explicitly gave myself at some earlier time. It is engraved on my mind as on a slate. It is incompatible with the hypothesis that I meant quus. It is this set of directions, not the finite list of particular additions that I performed in the past, that justifies and determines my present response.[7]

And his response to it seems clear (p. 16ff). The idea is that thoughts that someone may have had concerning how he is prepared to use a certain expression will help determine a meaning for that expression only if their correct interpretation is presupposed. But this is equivalent to assuming, Kripke suggests, that the sceptical challenge has been met with respect to the *expressions that figure in those thoughts*. But how was *their* meaning fixed? Not by facts about their actual or counterfactual history of use, (if the argument against a dispositional account of meaning is to be believed); and not by facts concerning associated experiential episodes. Hence—on the assumption that no other sort of fact is relevant to the fixation of meaning—by nothing.

The strategy seems clear; but is it not problematic? The trouble is that it seems to depend on the assumption that thought contents are the properties of syntactically identifiable bearers—properties, that is, of expressions belonging to a 'language of thought'. And although there may be much to recommend this view, still, does Kripke really wish to rest the sceptical conclusion on so contestable a premiss?

Fortunately for the sceptical strategy, we will see below that, although a contestable premiss about thought is involved, it is nothing so rich as a language of thought hypothesis. But we will be in a position to appreciate this properly only after we have examined McGinn's claim that, even granted a linguistic model of thinking, it is still impossible to run a Kripke-style sceptical argument against thought.

The normativity of meaning

4. McGinn writes:

> The issue of normativeness, the crucial issue for Kripke, has no clear content in application to the language of thought: what does it mean to ask whether my current employment of a word in my language of thought (i.e. the exercise of a particular concept) is *correct* in the light of my earlier employment of that word? What kind of linguistic mistake is envisaged here? . . . There is just no analogue here for the idea of linguistic incorrectness (as opposed to the *falsity* of a thought): linguistic incorrectness (of the kind we are concerned with) is using the same word with a different meaning from that originally intended (and doing so in ignorance of the change), but we cannot in this way make sense of employing a concept with a different content from that originally intended—it would just be a *different concept*.[8]

The idea of mental content cannot be threatened by Kripke, McGinn

[7] K., pp. 15–16.

argues, because the principal requirement by which putative reconstructions of that notion are to be dispatched—the normativity requirement—has no cogent application to the language of thought. The claim calls for a somewhat more searching articulation of the normativity thesis than we have attempted so far. In what does the normativity of meaning consist?

McGinn offers the following characterization:

> The notion of normativeness Kripke wants captured is a transtemporal notion . . .. We have an account of this normativeness when we have two things: (a) an account of what it is to mean something at a given time and (b) an account of what it is to mean the *same* thing at different times—since (Kripkean) normativeness is a matter of meaning now what one meant earlier.[9]

So, the later use of the expression is 'correct', according to McGinn, if it then expresses the same meaning as it did earlier; 'incorrect' if, without intending to introduce a change of meaning by explicit stipulation, it expresses a different meaning. It is in such facts as this that the normativity of meaning is said to consist.

Supposing this were the right understanding of normativity, how would it affect mental content scepticism? McGinn says that the problem is that we cannot make sense of employing a concept with a different content from that originally intended—it would just be a different concept. But although that is certainly true, it is also irrelevant: what we need to make sense of is not employing a *concept* with a different content from that originally intended, but employing an *expression in the language of thought* with a different content from that originally intended, which is a rather different matter.

As it happens, however, it is an idea that is equally problematic. The difficulty is that we do not have the sort of access to the expressions of our language of thought that an attribution to us of semantic intentions in respect of them would appear to presuppose. You cannot intend that some expression have a certain meaning unless you are able to refer to that expression independently of its semantic properties. But we have no such independent access to the expressions of our language of thought; we do not, for instance, know what they look like. So we cannot have semantic intentions in respect of them and, hence, cannot make sense of using them correctly or incorrectly in the sense defined by McGinn.

If McGinn's understanding of normativity were the correct one, then, it would indeed be difficult to see how it could operate at the level of thought (though not quite for the reasons he gives). It ought to be clear, however, that the 'normativity' requirement defined by McGinn has nothing much to do with the concept of meaning *per se* and is not the requirement that Kripke is operating with.

We may appreciate this point by observing that the requirement defined

[8] Op. cit., p. 147.

[9] Ibid., p. 174.

by McGinn could hardly act as a substantive constraint on theories of meaning, even where these are theories solely of *linguistic* meaning. *Any* theory of meaning that provided an account of what speakers mean by their expressions at arbitrary times—however crazy that theory may otherwise be—would satisfy McGinn's constraint. In particular, the main theory alleged by Kripke to founder on the normativity requirement, would easily pass it on McGinn's reading: since there are perfectly determinate facts about what dispositions are associated with a given expression at a given time—or, rather, since it is no part of Kripke's intent to deny that there are—it is always possible to ask whether an expression has the same or a different meaning on a dispositional theory, thus satisfying McGinn's requirement. How to explain, then, Kripke's claim that a dispositional theory founders precisely on the normativity requirement?

5. The answer is that the normativity requirement is not the thesis McGinn outlines. What is it then?

Suppose the expression 'green' means *green*. It follows immediately that the expression 'green' applies *correctly* only to *these* things (the green ones) and not to *those* (the non-greens). The fact that the expression means something implies, that is, a whole set of *normative* truths about my behaviour with that expression: namely, that my use of it is correct in application to certain objects and not in application to others. This is not, as McGinn would have it, a relation between meaning something by an expression at one time and meaning something by it at some later time; it is rather, a relation between meaning something by it at some time and its *use at that time*.

The normativity of meaning turns out to be, in other words, simply a new name for the familiar fact that, regardless of whether one thinks of meaning in truth-theoretic or assertion-theoretic terms, meaningful expressions possess conditions of *correct use*. (On the one construal, correctness consists in *true* use, on the other, in *warranted* use.) Kripke's insight was to realize that this observation may be converted into a condition of adequacy on theories of the determination of meaning: any proposed candidate for the property in virtue of which an expression has meaning, must be such as to ground the 'normativity' of meaning—it ought to be possible to read off from any alleged meaning constituting property of a word, what is the correct use of that word. It is easy to see how, on this understanding of the requirement in question, a dispositional theory might appear to fail it: for, it would seem, one cannot read off a disposition to use a word in a certain way what is the correct use of that word, for to be disposed to use a word in a certain way implies at most that one *will*, not that one *should* (one can have dispositions to use words *incorrectly*).[10]

[10] As we shall see below, however, the question whether dispositional accounts of meaning really do succumb to the normativity objection is much more complicated than this. I am not here trying to assess the objection, but merely to state it.

6. With this clarification of the normativity thesis in place we are finally in a position to settle the question: can Kripke develop the same sort of meaning-sceptical argument against a language of thought as he develops against public language? And the answer is: clearly, yes. For: what fixes the meaning of expressions in the language of thought? Not other thoughts, on pain of vicious regress. Not facts about the actual tokening of such expressions or facts about associated qualitative episodes, for familiar reasons. And not dispositional facts about the tokening of such expressions, for, since meaningful expressions of mentalese possess conditions of correct use in precisely the same sense as public language expressions do, because correctness cannot be reconstructed dispositionally. So, nothing fixes their meaning.

Indeed, we are also now in a position to see, as promised, that nothing so rich as a language of thought hypothesis is strictly needed. A language of thought model is composed out of two theses: (*a*) that thinking the thought that *p* involves tokening an item—a representation—that means that *p*; and (*b*) that the representation whose tokening is so involved possesses a combinatorial syntactic and semantic structure. In other words, according to a language of thought hypothesis, thought contents are the semantic properties of syntactically and semantically *structured bearers*. But it should be quite clear that nothing in the sceptical argument depends on the assumption of *structure*: even if the representation were to possess no internal syntax, we could still ask, in proper Kripkean fashion, what its correctness conditions are and in virtue of what they are determined.

It would appear, however, that the sceptical argument's strategy does presuppose that content properties have *some* sort of bearer (even if not necessarily a structured one). For, otherwise, there will be no natural way to formulate a dispositional theory of thought content, and no natural way to bring the normativity requirement to bear against it. There has to be *something*—a state, event, or particular, it need not matter which—whose disposition to get tokened under certain circumstances constitutes, on a dispositional theory, its possession of a certain content. And although this commitment is, I suppose, strictly speaking contestable, it is also very natural and plausible. After all, contents do not figure in a mental life except as subtended by a particular *mode*—belief, desire, judgement, wish—and, hence, are naturally understood as the properties of the states or events that instantiate those modes.

And so we see that the sceptical argument must, can, and does (in intent, anyway) include mental content within the scope of the scepticism it aims to promote.[11]

[11] Since nothing will hang on it, and since it will ease exposition, I shall henceforth write as if a language of thought hypothesis were true.

The constitutive nature of the sceptical problem

7. Having a meaning is essentially a matter of possessing a correctness condition. And the sceptical challenge is to explain how anything could possess *that*.

Notice, by the way, that I have stated the sceptical problem about meaning without once mentioning Kripke's notorious sceptic. That character, as everyone knows, proceeds by inviting his interlocutor to defend a claim about what he previously meant by the expression '+'. The interlocutor innocently assumes himself to have meant addition; but the sceptic challenges him to prove that the concept in question was not in fact *quaddition*, where quaddition is just like addition, except for a singularity at a point not previously encountered in the interlocutor's arithmetical practice.

It may seem, then, that the sceptical problem I have described could not be Kripke's. For Kripke's problem appears to be essentially *epistemological* in character—it concerns a speaker's ability to defend a particular meaning ascription; whereas the problem I have outlined is *constitutive*, not epistemological—its topic is the *possibility* of meaning, not our knowledge of it.

In fact, however, the two problems are the same; Kripke merely chooses to present the constitutive problem in an epistemological guise. Epistemological scepticism about a given class of judgements is the view that our actual cognitive capacities are incapable of delivering justified opinions concerning judgements in that class. Kripke's sceptic is not after a thesis of that sort. This is evident from the fact that his interlocutor, in being challenged to justify his claim that he meant addition by '+', is permitted *complete and omniscient* access to all the facts about his previous behavioural, mental, and physical history; he is not restricted to the sort of knowledge that an ordinary creature, equipped with ordinary cognitive powers, would be expected to possess.[12] Kripke's sceptical scenario is, thus, completely unsuited to promoting an epistemological scepticism. What it *is* suited for is the promotion of a constitutive scepticism. For if his sceptic is able to show that, even with the benefit of access to all the relevant facts, his interlocutor is still unable to justify any particular claim about what he meant, that would leave us no choice but to conclude that there are no facts about meaning.[13]

Pace many of Kripke's readers, then, the problem is not—not even in part—epistemological scepticism about meaning.[14] But, of course, one may agree that the problem is constitutive in character, and yet believe it to have an epistemological dimension. According to Crispin Wright, for example, Kripke is not interested in the mere possibility of correctness conditions; he is interested in the possibility of correctness conditions that may be, at least in one's own case, *known non-inferentially*.[15] The problem is essentially constitutive in character; but acceptable answers to it are to be subject to an epistemic constraint.

I do not wish to argue about this at length. It does seem to me that, once we have corrected for the distortions induced by the dialogic setting, there ought not to be any residual temptation to think that epistemological considerations are playing a critical role in Kripke's argument. In any case, whatever intention Kripke may have had, the considerations he adduces on behalf of the sceptical conclusion appear to owe nothing to epistemological constraints and can be stated without their help.[16] That, anyway, is how I shall present them.

The 'rule-following' considerations?

8. It would not be inappropriate to wonder at this point what all this has to do with the topic of *rule-following*? Where, precisely, is the connection between the concepts of meaning and content, on the one hand, and the concept of following a rule, on the other, forged? I shall argue that, in an important sense, the answer is 'nowhere', and hence that 'the rule-following considerations' is, strictly speaking, a misnomer for the discussion on offer.

Many writers seem to assume that the connection is straightforward; they may be represented as reasoning as follows. Expressions come to have correctness conditions as a result of people following rules in respect of them; hence, exploring the possibility of correctness is tantamount to exploring the possibility of rule-following.

But, at least on the ordinary understanding of the concept of following a rule, it cannot be true of *all* expressions—in particular, it cannot be true of *mental* expressions—that they come to have correctness conditions as a result of people following rules in respect of them. The point is that the ordinary concept of following a rule—as opposed to that of merely

[12] McGinn's failure to note this leads him to wonder how the constitutive and epistemological aspects of Kripke's discussion are related, 'for the epistemological claim is clearly distinct from the metaphysical claim' (op. cit., p. 149).

[13] This point is made very nicely by Crispin Wright in his 'Kripke's Account of the Argument Against Private Language', *Journal of Philosophy*, 1984, pp. 761–2. Wright, however, discerns another sort of epistemological dimension to the sceptical problem. I will discuss that below.

[14] For example, McGinn, op. cit., pp. 140–50; G. Baker and P. Hacker, 'On Misunderstanding Wittgenstein: Kripke's Private Language Argument', *Synthèse*, 1984, pp. 409–10. Neil Tennant has complained that Kripke's sceptic does not ultimately supply a convincing bent-rule reinterpretation of his interlocutor's words. See his 'Against Kripkean Scepticism', forthcoming. Tennant may well be right about this. But here again, I think, the perception that this affects the force of the sceptical problem about meaning is a result of taking the dialogic setting too seriously. The constitutive problem about meaning—how could there so much as be a correctness condition—can be stated quite forcefully without the actual provision of a convincing global reinterpretation of a person's words.

[15] See op. cit., pp. 772–5.

[16] With one relatively minor exception to be noted below.

conforming to one—is the concept of an *intentional* act: it involves the intentional attempt to bring one's behaviour in line with the dictates of some grasped rule. Crispin Wright has decribed this intuitive conception very clearly:

> Correctly applying a rule to a new case will, it is natural to think, typically involve a double success: it is necessary both to apprehend relevant features of the presented situation and to know what, in the light of those apprehended 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.[17]

As such, however, the ordinary concept of following a rule is the concept of an act among whose causal antecendents lie contentful mental states; consequently, it is a concept that *presupposes* the idea of a correctness condition, not one that can, in full generality, help explain it. Since it makes essential play with the idea of a propositional attitude, which in turn makes essential play with the idea of content, rule-following in this sense presupposes that *mental expressions* have conditions of correct application. On pain of regress, then, it cannot be true that mental expressions themselves acquire meaning as a result of anyone following rules in respect of them.

What Kripke's discussion is concerned with is the possibility of correctness; so long as we keep that clearly in mind, talk of 'rule-following' is harmless. Simon Blackburn has captured this perspective very well:

> I intend no particular theoretical implications by talking of rules here. The topic is that there is such a thing as the correct and incorrect application of a term, and to say that there is such a thing is no more than to say that there is truth and falsity. I shall talk indifferently of there being correctness and incorrectness, of words being rule-governed, and of their obeying principles of application. Whatever this is, it is the fact that distinguishes the production of a term from mere noise, and turns utterance into assertion—into the making of judgment.[18]

II
THE SCEPTICAL SOLUTION

A non-factualist conception of meaning

9. Having established to his satisfaction that no word could have the property of expressing a certain meaning, Kripke turns to asking how this

[17] Crispin Wright: 'Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics', in *Reflections on Chomsky*, ed. A. George, Oxford, Basil Blackwell, 1989, p. 255.

[18] Simon Blackburn, 'The Individual Strikes Back', *Synthèse*, 1984, pp. 281–2. My only disagreement with this passage concerns its identification of correctness conditions with truth conditions. Truth conditions are simply one species of a correctness condition; proof conditions or justification conditions supply further instances.

conclusion is to be accommodated. The question is urgent, in his view, because the conclusion threatens to be not merely shocking but paradoxical. The trouble is that we would ordinarily take a remark to the effect that there could not be any such thing as the fact that I mean something by the '+' sign, to entail that there is nothing I could mean by the use of that sign. Applied quite generally, across all signs and all people, the claim becomes the seemingly paradoxical and self-refuting thesis that no one could mean anything by their use of linguistic expressions.

Why the solution is urgent: nobody could mean anything.

A scepticism about meaning facts would appear to be, then, prima facie anyway, an unstable position. Sustaining it requires showing that what it asserts does not ultimately lapse into a form of pragmatic incoherence. What is called for, in other words, is a rehabilitation of our ordinary practice of attributing content to our thoughts and utterances, which nevertheless conserves the sceptical thesis that there are no facts for such attributions to answer to. That is what the 'sceptical solution' is designed to do. It is alleged to have the following startling consequence: the idea of a language whose meanings are constituted solely out of an individual's speaker's properties, considered 'completely in isolation from any wider community to which he may belong', is incoherent.[19]

Rehabilitate the practice, keep the no-facts thesis.

The sceptical solution has two parts that are usefully distinguished. The first consists in the suggestion that we replace the notion of truth conditions, in our intuitive picture of sentence meaning, by that of *assertibility* conditions. The second consists in a *description* of the assertibility conditions for meaning-attributing sentences, in the course of which it is argued that it is essential to such sentences that their assertibility conditions advert to the actions or dispositions of a community.

The two parts, semantic and social, kept apart.

The adjustment recommended in the first part is supposed to help because

Kripke's own words on why the first part helps.

> if we suppose that facts or truth conditions are of the essence of meaningful assertion, it will follow from the skeptical conclusion that assertions that anyone ever means anything are meaningless. On the other hand, if we apply to these assertions the tests suggested . . . no such conclusion follows. All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertible, and that the game of asserting them has a role in our lives. No supposition that 'facts correspond' to those assertions is needed.[20]

[19] Following Goldfarb, we may call this the concept of a 'solitary language'. See his 'Kripke on Wittgenstein on Rules', *Journal of Philosophy*, 1985. Goldfarb goes on to say that the idea of a solitary language is more general than that of a Wittgensteinian 'private language', for the latter essentially involves the idea of *necessary unintelligibility* to another. It is hard to assess this, because it is hard to know how to interpret 'necessary unintelligibility'. Surely it cannot mean: a language to whose predicates no two people *could* attach the same descriptive conditions. And it is not clear what it is to mean, if not that. For useful discussion see C. Wright, 'Does *Philosophical Investigations* I.258–60 Suggest a Cogent Argument Against Private Language?', in *Subject, Thought and Context*, ed. P. Pettit and J. McDowell, Oxford, Clarendon Press, 1986.

[20] K., pp. 77–8.

No line drawn round meaning-talk: non-factualism goes global.

The proposed account is, in effect, a *global* non-factualism: sentence significance is construed quite generally in assertion-theoretic terms and no invidious distinction is drawn between the sort of significance possessed by meaning-attributing sentences and that possessed by sentences of other types.

The argument against solitary language

The argument against solitary language, in Kripke's words.

10. The argument against 'solitary language' emerges, according to Kripke, from the observation that, so long as a speaker is considered in isolation we can assign no assertibility conditions to judgements to the effect that he has misapplied a symbol in his repertoire:

> [I]f we confine ourselves to looking at one person alone, this is as far as we can go . . .. There are no circumstances under which we can say that, even if he inclines to say '125', he should have said '5', or *vice-versa* . . .. Under what circumstances can he be wrong? No one else by looking at his mind or behavior alone can say something like, 'He is wrong if he does not accord with his own intention'; the whole point of the skeptical argument was that there are no facts about him in virtue of which he accords with his intentions or not.[21]

Error needs the community, so no solitary language.

The possibility of error, however, is essential to our ordinary concept of meaning, and can only be accommodated if we widen our gaze and take into consideration the interaction between our imagined rule-follower and a linguistic community. Were we to do so, Kripke continues, we could introduce assertibility conditions for judgements about error in terms of the agreement, or lack of it, between a given speaker's propensities in the use of a term and the community's. Since, however, this would appear to be the *only* way to give substance to the correlative notions of error and correctness, no one considered wholly in isolation from other speakers could be said to mean anything. And so a solitary language is impossible.

Let us turn now to an assessment of the various central aspects of Kripke's argument.

III
ASSESSMENT OF THE ARGUMENT AGAINST SOLITARY LANGUAGE

Constitutive accounts and solitary language

11. Kripke is very clear about the limited, wholly descriptive nature of the sceptical solution, at least in his 'official' explications of the view:

> We have to see under what circumstances attributions of meaning are made and what role these attributions play in our lives. Following Wittgenstein's exhortation not to think but to look, we will not reason *a priori* about the role such statements *ought* to play; rather we will find out what circumstances *actually* license such assertions and what role this license *actually* plays. It is important to realize that we are *not* looking for necessary and sufficient conditions (truth conditions) for following a rule, or an analysis of what such rule-following 'consists in'. Indeed such conditions would constitute a 'straight' solution to the skeptical problem, and have been rejected.[22]

[21] K., p. 88.

Kripke's official modesty: look, don't reason a priori.

The modesty is forced: the solution has to be sceptical.

It is important to see that the counselled modesty—we will not reason a priori about the role such statements ought to play—is compulsory. The assertibility conditions may not be understood to provide the content (or truth conditions) of the meaning-attributing sentences, on pain of falling prey to the accepted sceptical considerations. (That is why the solution on offer has to be *sceptical*: it has already been conceded that nothing could cogently amount to the fact that a meaning sentence reports). It would appear to follow from this, however, that the sceptical solution can do no more than record the conditions under which speakers in fact consider the attribution of a certain concept warranted and the endorsement of a particular response appropriate. The Wittgensteinian exhortation 'not to think but to look' is not merely (as it may be) good advice; the modesty it counsels is enforced by the fact that truth conditions for these sentences has been jettisoned. For how, in the absence of a conception of the truth conditions of meaning attributing sentences, could the project of providing an account of their assertion conditions aspire to anything more than descriptive adequacy? Were we equipped with an account of their truth conditions, of course, we might be able to reason a priori about what their assertion conditions *ought* to be and, hence, potentially, to revise the conditions for assertion *actually* accepted for them. But without the benefit of such an account there is no scope for a more ambitious project: a descriptively adequate account of the *actual* assertion conditions for such sentences is the most one may cogently aim for.

So it can only record what we actually count as warranted.

Our language isn't solitary, but that's not the modal result.

If this is correct, however, we ought to be puzzled about how the sceptical solution is going to deliver a conclusion against solitary language of the requisite modal force: namely, that there *could not* be such a language. For even if it were true that our *actual* assertibility conditions for meaning-attributing sentences advert to the dispositions of a community, the most that would license saying is that *our* language is not solitary. And this would be a lot less than the result we were promised: namely, that any *possible* language has to be communal.

Communal assertibility conditions?

12. Putting this worry to one side, let us ask whether it is in fact true that, if we accept the sceptical conclusion, we cannot introduce substantive

[22] K., pp. 86–7.

Nothing in the conclusion rules out (A), as Goldfarb notes.

assertibility conditions for meaning-attributions that do not advert to the dispositions of a community of speakers? It appears, on the contrary, that not only can we introduce such conditions, but have actually done so.[23] Consider the following:

> (A) It is warranted to assert of Jones that he means addition by '+', provided he has responded with the *sum* in reply to most arithmetical queries posed thus far.

As a description of our practice, (A) is, of course, quite rough: room has to be made for the importance of systematic deviations, the greater importance attaching to simple cases, and many other such factors. But all these refinements may be safely ignored for the purpose of raising the following critical question: what in the sceptical conclusion rules out attributions of form (A)? It had better rule them out, of course, if the argument against solitary language is to be sustained, for (A) adverts to no one other than the individual. But as Goldfarb points out, there appears to be nothing in the sceptical conclusion that will rule it out.[24] It can hardly be objected that

Using 'sum' is no cheat: this isn't a straight solution.

the interpretation of 'sum' is being presupposed in the statement of the condition, for the sceptical solution is not meant to be a *straight* solution to the problem about meaning; as Kripke himself says, in fending off a similar imagined objection to his own account of the assertibility conditions:

> What Wittgenstein is doing is describing the utility in our lives of a certain practise. Necessarily he must give this description in our own language. As in the case of any such use of our language, a participant in another form of life might apply various terms in the description (such as 'agreement') in a non-standard 'quus-like' way. . . . This cannot be an objection to Wittgenstein's solution unless he is to be prohibited from any use of language at all.[25]

The sceptic leaves arithmetical facts alone.

Nor is there any problem in the assumption that it is a genuinely factual matter what any two numbers sum to; as Kripke himself repeatedly emphasizes, the sceptical argument does not threaten the existence of *mathematical* facts. But how, then, is (A) to be ruled out, and the argument against solitary language preserved?

Is (A) parasitic on the communal conditions? The reverse.

13. Could it perphaps be argued that (A) is permissible though *parasitic* on the communal assertibility conditions Kripke outlines? As a matter of fact, just the opposite seems true.[26]

Kripke's communitarian account of meaning-attributions runs as follows:

> Smith will judge Jones to mean addition by 'plus' only if he judges that Jones's answers to particular addition problems agree with those he is inclined to give . . .

[23] This sort of rejoinder is canvassed both in Goldfarb, op. cit., and in McGinn, op. cit.
[24] Ibid.
[25] K., p. 146.
[26] This is argued in McGinn, op cit., pp. 185–7, from which this point is derived.

> If Jones consistently fails to give responses in agreement . . . with Smith's, Smith will judge that he does not mean addition by 'plus'. Even if Jones did mean it in the past, the present deviation will justify Smith in judging that he has lapsed.[27]

Smith's condition: Jones must agree with Smith's answers.

According to this account, then, I will judge that Jones means addition by 'plus' only if Jones uses 'plus' enough times in the same way I am inclined to use it. As a rough description of our practice, and many important refinements aside, this seems acceptable enough. One of the refinements that is called for, however, exposes the fact that Kripke's communitarian conditions are parasitic on the solitary conditions, and not the other way round.

The drunk or senile Smith, and amendment (B).

It would be absurd for me, under conditions where I had good reason to believe that I had become prone to making arithmetical mistakes—perhaps owing to intoxication or senility or whatever—to insist on agreement with me as a precondition for crediting Jones with mastery of the concept of addition. And this would appear to show that, at a minimum, Kripke's communitarian account must be modified to read:

> (B) It is warranted to assert of Jones that he means addition by '+', provided he agrees with my responses to arithmetical queries, *under conditions where I have been a reliable computer of sums.*

Smith drops out as idle, and only (A) remains.

But this modification would seem immediately to reveal that the reference to 'my own responses' is idle, and that the basic assertion condition I accept is just (A):

> It is warranted to assert of Jones that he means addition by '+', provided he has responded with the *sum* in reply to most arithmetical queries posed thus far.

It would appear, in other words, that the acceptability of the communitarian conditions is strongly parasitic on the acceptability of the solitary ones, and not the other way around.

Verdict: no convincing argument against solitary language.

In sum: both because it is difficult (impossible?) to generate constitutive results out of non-constitutive accounts, and because our actual assertibility conditions for meaning ascriptions appear not to be communitarian, I conclude that the sceptical solution does not yield a convincing argument against solitary language.

IV
IRREALIST CONCEPTIONS OF MEANING

14. The argument against solitary language was supposed to flow from the adjusted understanding of sentence significance forced by the sceptical conclusion. The sceptical conclusion has it that it cannot literally be true of

[27] K., p. 91.

Now the semantic claim: is non-factualism forced?

any symbol that it expresses a particular meaning: there is no appropriate fact for a meaning-attributing sentence to report. The sceptical solution's recommendation is that we blunt the force of this result by refusing to think of sentence significance in terms of possession of truth conditions, or a capacity to state facts. We should think of it, rather, in terms of possession of assertibility conditions. But is this solution forced? Are there not, perhaps, other ways of accommodating the sceptical conclusion?

Two ways it looks like an overreaction.

The solution on offer is bound to strike one as an overreaction, at least at first blush, in two possible respects. First, in that it opts for a form of *non-factualism*, as opposed to an *error* theory; and second, in that the recommended non-factualism is *global*, rather than restricted solely to the region of discourse—meaning talk—that is directly affected by the sceptical result it seeks to accommodate.

The witches: an error theory keeps truth conditions.

Semantically speaking, the most conservative reaction to the news that nothing has the property of being a witch is not to adopt a non-factualist conception of witch talk, it is to offer an *error* conception of such talk. An error conception of a given region of discourse conserves the region's semantical appearances—predicates are still understood to express properties, declarative sentences to possess truth conditions; the ontological discovery is taken to exhibit—merely—the *systematic* falsity of the region's (positive, atomic) sentences.[28]

An error theory of meaning: all attributions false.

Could not the moral of the sceptical argument be understood to consist in an error conception of meaning discourse? It could not, for an error conception of such discourse, in contrast with error conceptions of other regions, is of doubtful coherence. The view in question would consist in the claim that all meaning-attributions are false:

(1) For any S: ⌜S means that p⌝ is false.

But the disquotational properties of the truth predicate guarantee that (1) entails

(2) For any S: ⌜S⌝ has no meaning.

So (1) can't be true: falsity needs meaning.

(1) implies, that is, that no sentence whatever possesses a meaning. Since, however, a sentence cannot be *false* unless it is meaningful to begin with, this in turn implies that (1) cannot be true: for what (1) says is that some sentences—namely meaning-attributing sentences—are false.[29]

[28] See John Mackie, *Ethics: Inventing Right and Wrong*, London, Penguin, 1977, for such a conception of moral discourse.

[29] An error conception of meaning has been advocated by Paul Churchland; see his 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy*, 1981. This argument is elaborated and defended in my 'The Status of Content', *Philosophical Review*, forthcoming April 1990.

So no error theory. Does non-factualism fare better?

So it appears that Kripke was right to avoid an error conception of meaning discourse. But does his *non-factualist* conception fare any better?

15. The canonical formulation of a non-factualist view—and the one that Kripke himself favours—has it that some targeted declarative sentence is not genuinely truth-conditional. A non-factualism about meaning, consists, that is, in the view that

The non-factualism is meant globally, (3) to (4).

(3) For any S, p: ⌜S means that p⌝ is not truth-conditional.

As I noted above, however, the projectivism recommended by the sceptical solution is intended to apply globally: it is not confined solely to meaning-attributing sentences. Thus,

Why globality is forced: truth conditions depend on meaning.

(4) For any S: ⌜S⌝ is not truth conditional.

Why does Kripke adopt so extreme a view? Why does he not suggest merely that we abandon a truth-conditional model for *semantic* discourse, while preserving it, as seems natural, for at least *some* regions of the rest of language? Kripke does not say. But it may be that he glimpsed that the global character of the projectivism is in fact *forced* in the present case.[30] For consider a non-factualism solely about meaning—the view that, since there is no such property as a word's meaning something, and hence no such fact, no meaning-attributing sentence can be truth-conditional. Since the truth-condition of any sentence S is (in part, anyway) a function of its meaning, a non-factualism about meaning will enjoin a non-factualism about truth-conditions: what truth-condition S possesses could hardly be a factual matter if that in virtue of which it has a particular truth-condition is not itself a factual matter. And so we have it that (3) entails:

By disquotation again, (5) delivers (4).

(5) For all S, p: ⌜S has truth-condition p⌝ is not truth-conditional.

However, since, courtesy of the disquotational properties of the truth predicate, a sentence of the form ⌜S has truth-condition p⌝ is true if and only if S has truth-condition p, and since (5) has it that ⌜S has truth-condition p⌝ is never simply true, it follows that

(4) For any S: ⌜S⌝ is not truth-conditional

just as predicted.

It is, then, a fascinating consequence of a non-factualism about meaning, that it entails a *global* non-factualism; in this respect, if no other, a non-factualism about meaning distinguishes itself from a similar thesis

[30] Somewhat different arguments are given for this both in Crispin Wright's 'Kripke's Account', loc. cit., pp. 769–70 and in my 'Meaning, Content and Rules', in Part I of my Ph.D. Dissertation *Essays on Meaning and Belief*, Princeton, 1986.

Wright's suggestion: a self-including projectivism is unstable.

about *any* other subject matter. Crispin Wright has suggested that it also renders it irremediably problematic:

> it is doubtful that it is coherent to suppose that projectivist views could be applied quite globally. For, however exactly the distinction be drawn between fact-stating and non-fact-stating discourse, the projectivist will presumably want it to come by way of a *discovery* that certain statements fail to qualify for the former class; a statement of the conclusion of the skeptical argument, for instance, is not *itself* to be projective.[31]

It is hard not to sympathize with Wright's suggestion that there must be something unstable about a projectivist thesis that is itself within the scope of the projectivism it recommends. But it is also not entirely clear to me in what the instability consists. To be sure, a global projectivism would have to admit that it is no more than *assertible* that no sentence possesses a truth condition. But what is wrong with that? If there is an instability here, it is not a transparent one.

The real instability is about truth.

16. In fact, however, I do believe that a non-factualism about meaning is unstable, but not because of its global character. Rather, the reasons have to do with the clash between what you have to suppose about truth in order to frame a non-factualist thesis about anything, and what you have to suppose about truth as a result of accepting a non-factualism about meaning. I have developed the argument for this in some detail elsewhere;[32] here I have space only to sketch its outlines.

Consider a non-factualist thesis about, say, the good:

(7) All sentences of the form ⌜*x* is good⌝ are not truth-conditional.

Ayer's deflationary truth leaves non-factualism no room.

The point that needs to be kept in focus is that the sentence of which truth conditions are being denied is a significant *declarative* sentence. For this fact immediately implies that the concept of truth in terms of which the non-factualist thesis is framed cannot be the *deflationary* concept that A. J. Ayer succinctly described as follows:

> . . . to say that p is true is simply a way of asserting p. . . . The traditional conception of truth as a 'real quality' or a 'real relation' is due, like most philosophical mistakes, to a failure to analyze sentences correctly . . .. There are sentences in which the word 'truth' seems to stand for something real . . . [but] our analysis has shown that the word 'truth' does not stand for anything.[33]

If the concept of truth were, as Ayer claims in this passage, merely the concept of a device for semantic ascent, and not the concept of some genuine property—some 'real relation'—that a sentence (or thought) may enjoy, then non-factualism is nowhere a coherent option. For on a deflationary understanding of truth, a sentence will be truth-conditional

[31] Ibid., p. 770.
[32] In 'The Status of Content', loc. cit.
[33] A. J. Ayer, *Language, Truth and Logic*, New York, Dover, 1952, p. 89.

provided only that it is apt for semantic ascent; and it will be apt for semantic ascent provided only that it is a significant, declarative sentence. But it is constitutive of a non-factualist thesis precisely that it denies, of some targeted, significant, declarative sentence, that it is truth-conditional. It follows, therefore, that a non-factualism about any subject matter presupposes a conception of truth richer than the deflationary: it is committed to holding that the predicate 'true' stands for some sort of language-independent property, eligibility for which will not be certified purely by the fact that a sentence is declarative and significant. Otherwise, there will be no understanding its claim that a significant sentence, declarative in form, fails to possess truth-conditions.

So non-factualism needs robust, language-independent truth.

So we have it that any non-factualist thesis presupposes that truth is, as I shall henceforth put it, *robust*. But, now, notice that judgements about whether an object possesses a robust property could hardly fail to be factual. If *P* is some genuinely robust property, then it is hard to see how there could fail to be a fact of the matter about whether an object has *P*. It does not matter if *P* is subjective or otherwise dependent upon our responses. So long as it is a genuine, language-independent property, judgements about it will have to be factual, will have to be possessed of robust truth-conditions. In particular, if *truth* is a robust property, then judgements about a sentence's truth value must themselves be factual. But we saw earlier—see (5) above—that a non-factualist thesis about meaning implies that judgements about a sentence's truth cannot be factual: whether a certain sentence is true cannot be a factual matter if its meaning is not. And this exposes the contradiction we have been stalking: a non-factualism about meaning implies both that truth is robust and that it is not.

Judgements about a robust property are factual: contradiction.

17. It is hard to do justice to the issues involved within the confines of the present essay.[34] I do hope, however, that the preceding discussion has succeeded in sowing some doubts about the cogency of irrealist conceptions of meaning—whether in the form of a non-factualism about meaning, as in the sceptical solution, or an error theory, as suggested, for instance, by Churchland.

Doubts about either irrealism, Churchland's included.

The uncompromising strength of the claim is bound to arouse suspicion. Irrealist conceptions of other domains may not be particularly appealing or plausible, but they are not incoherent. Why should matters stand differently with meaning discourse?

Irrealism elsewhere is implausible, not incoherent. Why?

The source of the asymmetry is actually not that hard to track down. It consists in the fact that error and non-factualist theories about *any* subject matter presuppose certain claims about truth and truth-conditions, that an error or non-factualist conception directed precisely at our talk of meaning itself ends up denying. Not surprisingly the ensuing result is unstable.

The asymmetry: irrealism about meaning denies what it presupposes.

[34] Again, for a more detailed treatment see 'The Status of Content', loc. cit.

The verdict, and the hunt for the faulty step begins.

Thus, an error thesis about any subject matter presupposes that the target sentences are truth-conditional. But an error thesis directed precisely at our talk about meaning entails the denial of that presupposition. Thus, also, a non-factualism about any subject matter presupposes a robust conception of truth. But a non-factualism directed precisely at our talk about meaning entails the denial of that presupposition.

If these considerations are correct, then, they would show that the sceptical conclusion cannot be sustained: there appears to be no stable way of accommodating the claim that there are no truths about meaning. Something must be wrong, therefore, with the argument that appeared to lead us to it. What could it be?

V
REDUCTIVE ACCOUNTS OF MEANING

18. The sceptical argument has been faulted on a number of grounds, the most important being:

> That its arguments against dispositional accounts of meaning do not work.
> That it neglects to consider all the available naturalistic facts.
> That its conclusion depends on an unargued reductionism.

The first two objections issue from a naturalistic perspective: they claim that the sceptical argument fails to establish its thesis, even granted a restriction to *naturalistic* facts and properties. The final objection concedes the failure of naturalism, but charges that the sceptical argument is powerless against an appropriately anti-reductionist construal of meaning. In this part I shall examine the naturalistic objections, and in the next the anti-reductionist suggestion.

I should say at the outset, however, that I see no merit to objections of the second kind and will not discuss them in any detail here. All the suggestions that I have seen to the effect that Kripke ignores various viable reduction bases for meaning facts seem to me to rest on misunderstanding. Colin McGinn, for example, claims that Kripke neglects to consider the possibility that possession of a concept might consist in possession of a certain sort of capacity. Capacities, McGinn explains, are distinct from dispositions and are better suited to meet the normativity constraint.[35] This rests on the misunderstanding of normativity outlined above. Warren Goldfarb charges that Kripke neglects to consider causal/informational accounts of the determination of meaning.[36] This derives from a failure to see that, in all essential respects, a causal theory of meaning is simply one species of a dispositional theory of meaning, an account that is, of course,

[35] See McGinn, op. cit., pp. 168–74.

[36] See Goldfarb, op. cit., n. 13.

extensively discussed by Kripke. It is unfortunate that this connection is obscured in Kripke's discussion. Because Kripke illustrates the sceptical problem through the use of an *arithmetical* example, he tends, understandably, to focus on *conceptual role* versions of a dispositional account of meaning, rather than on causal/informational versions. This has given rise to the impression that his discussion of dispositionalism does not cover causal theories. But the impression is misleading. For the root form of a causal/informational theory may be given by the following basic formula:

> *O* means (property) *P* by predicate *S* iff (it is a counterfactual supporting generalization that) *O* is disposed to apply *S* to *P*.

Dispositions and meaning: finitude

19. The single most important strand in the sceptical argument consists in the considerations against dispositional theories of meaning. It would be hard to exaggerate the importance of such theories for contemporary philosophy of mind and semantics: as I have just indicated, the most influential contemporary theories of content-determination—'informational' theories and 'conceptual-role' theories—are both forms of a dispositional account.[37] In my discussion I shall tend to concentrate, for the sake of concreteness, on informational theories of the content of *mental* symbols; but the issues that arise are general and apply to any dispositional theory whatever.

The root form of an information-style dispositional theory is this:

> My mental symbol 'horse' expresses whatever property I am disposed to apply it to.

Kripke's first objection amounts, in effect, to suggesting that there will always be a serious *indeterminacy in what my dispositions are*, thus rendering dispositional properties an inappropriate reduction base for meaning properties. For, Kripke argues, if it is indeed the property *horse* that I am disposed to apply the term to, then I should be disposed to apply it to *all* horses, including horses so far away and so far in the past that it would be nonsense to suppose I could ever get into causal contact with them. Otherwise, what is to say that my disposition is not a disposition to apply the term to the property *nearby horse*, or some such? But no one can

[37] For correlational theories see: F. Dretske, Knowledge and the Flow of Information, Cambridge, MIT Press, 1981; D. Stampe, 'Towards a Causal Theory of Linguistic Representation', *Midwest Studies in Philosophy*, vol. 2, Minneapolis, University of Minnesota Press, 1977; Jerry Fodor, *Psychosemantics*, Cambridge, MIT Press, 1987. For conceptual role theories see: H. Field, 'Logic, Meaning and Conceptual Role', *Journal of Philosophy*, 1977; Ned Block, 'Advertisement for a Semantics for Psychology', *Midwest Studies in Philosophy*, vol. 10, Minneapolis, University of Minnesota Press, 1986.

have a disposition to call all horses 'horse', for no one can have a disposition with respect to inaccessible objects.

The argument does not convince. Of course, the *counterfactual*

> If I were now to go to Alpha Centauri, I would call the horses there 'horse',

is false. If I were now to go to Alpha Centauri, I probably would not be in any position to call anything by any name, for I would probably die before I got there. But that by itself need not pose an insuperable obstacle to ascribing the disposition to me. *All* dispositional properties are such that their exercise—the holding of the relevant counterfactual truth—is contingent on the absence of interfering conditions, or equivalently, on the presence of ideal conditions. And it certainly seems conceivable that a suitable idealization of my biological properties will render the counterfactual about my behaviour on Alpha Centauri true. Kripke considers such a response and complains:

> But how can we have any confidence in this? How in the world can I tell what would happen if my brain were stuffed with extra brain matter? . . . Surely such speculation should be left to science fiction writers and futurologists.[38]

If the point is supposed to be, however, that one can have no reason for accepting a generalization defined over ideal conditions unless one knows exactly which counterfactuals would be true if the ideal conditions obtained, then, as Jerry Fodor has pointed out, it seems completely unacceptable.[39] For example, no one can claim to know all of what would be true if molecules and containers actually satisfied the conditions over which the ideal gas laws are defined; but that does not prevent us from claiming to know that, if there were ideal gases, their volume would vary inversely with the pressure on them. Similarly, no one can claim to know all of what would be true if I were so modified as to survive a trip to Alpha Centauri; but that need not prevent us from claiming to know that, if I were to survive such a trip, I would call the horses there 'horse'.[40]

Still, it is one thing to dispel an objection to a thesis, it is another to prove the thesis true. And we are certainly in no position now to show that we do have infinitary dispositions. The trouble is that *not every* true counterfactual of the form.

> If conditions were ideal, then, if C, S would do A

can be used to attribute to S the disposition to do A in C. For example, one can hardly credit a tortoise with the ability to overtake a hare, by pointing out that if conditions were ideal for the tortoise—if, for example,

[38] K., p. 27.
[39] See 'A Theory of Content', Part II, p. 15 (manuscript).
[40] For a related criticism of Kripke on this score see Blackburn, op. cit.

it were much bigger and faster—then it would overtake it. Obviously, only certain idealizations are permissible; and also obviously, we do not now know which idealizations those are. The set of permissible counterfactuals is constrained by criteria of which we currently lack a systematic account. In the absence of such an account, we cannot be completely confident that ascriptions of infinitary dispositions are acceptable, because we cannot be completely confident that the idealized counterfactuals needed to support such ascriptions are licit. But I think it is fair to say that the burden of proof here lies squarely on Kripke's shoulders: it is up to him to show that the relevant idealizations would be of the impermissible variety. And this he has not done.

Dispositions and meaning: normativity

20. Few aspects of Kripke's argument have been more widely misunderstood than his discussion of the 'normativity' of meaning and his associated criticism of dispositional theories. This is unsurprising given the difficulty and delicacy of the issues involved. In what sense is meaning a normative notion? Kripke writes:

> Suppose I do mean addition by '+'. What is the relation of this supposition to the question how I will respond to the problem '68 + 57'? The dispositionalist gives a *descriptive* account of this relation: if '+' meant addition, then I will answer '125'. But this is not the right account of the relation, which is normative, not descriptive. The point is not that, if I meant addition by '+', I will answer '125', but that, if I intend to accord with my past meaning of '+', I should answer '125'. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be disposed to respond as I should, but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is *normative*, not *descriptive*.[41]

The fact that I mean something by an expression, Kripke says, implies truths about how I *ought* to use that expression, truths about how it would be *correct* for me to use it. This much, of course, is incontestable. The fact that 'horse' means *horse* implies that 'horse' is correctly applied to all and only horses: the notion of the extension of an expression just is the notion of what it is correct to apply the expression to. It is also true that to say that a given expression has a given extension is not to make any sort of *simple* descriptive remark about it. In particular, of course, it is not to say that, as a matter of fact, the expression *will* be applied only to those things which are in its extension. Kripke seems to think, however, that these observations by themselves ought to be enough to show that no dispositional theory of meaning can work. And here matters are not so straightforward.

[41] K., p. 37.

Let us begin with the very crude dispositional theory mentioned above: 'horse' means whatever property I am disposed to apply it to. This is a hopeless theory, of course, but the reasons are instructive. There are two of them, and they are closely related. The first difficulty is that the theory is bound to get the extension of 'horse' wrong. Suppose I mean *horse* by it. Then, presumably, I have a disposition to call horses 'horse'. But it will also be true that there are certain circumstances—sufficiently dark nights—and certain cows—sufficiently horsey looking ones—such that, I am disposed, under those circumstances, to call those cows 'horse' too. Intuitively, this is a disposition to make a mistake, that is, to apply the expression to something not in its extension. But our crude dispositional theory, given that it identifies *the property I mean by an expression* with *the property I am disposed to apply the expression to*, lacks the resources by which to effect the requisite distinction between correct and incorrect dispositions. If what I mean by an expression is identified with whatever I am disposed to apply the expression to, then everything I am disposed to apply the expression to is, *ipso facto*, in the extension of that expression. But this leads to the unacceptable conclusion that 'horse' does not express the property *horse* but rather the disjunctive property *horse or cow*.

There is a related conceptual difficulty. Any theory which, like the crude dispositional theory currently under consideration, simply equates how it would be correct for me to use a certain expression with how I am disposed to use it, would have ruled out, as a matter of definition, the very possibility of error. And as Wittgenstein was fond of remarking, if the idea of correctness is to make sense at all, then it cannot be that whatever seems right to me is (by definition) right.

One would have thought these points too crucial to miss; but it is surprising how little they are appreciated. In a recent, comprehensive treatment of conceptual role theories, Ned Block has written

> of a choice that must be made by [conceptual role semantics] theorists, one that has had no discussion (as far as I know): namely, should conceptual role be understood in ideal or normative terms, or should it be tied to what people actually do? . . . I prefer not to comment on this matter . . . because I'm not sure what to say . . .[42]

This ought to seem odd. If conceptual role is supposed to determine meaning, then there can be no question, on pain of falling prey to Kripke's objection, of identifying an expression's conceptual role with a subject's actual dispositions with respect to that expression.

21. The objections from normativity show, then, that no dispositional theory that assumes the simple form of identifying *the property I mean by 'horse'* with *the property I am disposed to call 'horse'*, can hope to succeed. But what if a dispositional theory did not assume this simple form? What

[42] Ned Block, op. cit., p. 631.

if, instead of identifying what I mean by 'horse' with the *entire* range of my dispositions in respect of 'horse', it identified it only with certain select dispositions. Provided the theory specified a principle of selection that picked out only the extension-tracking dispositions; and provided also that it specified that principle in terms that did not presuppose the notion of meaning or extension, would it not then be true that the objections from normativity had been disarmed?

Let us try to put matters a little more precisely. If a dispositional theory is to have any prospect of succeeding, it must select from among the dispositions I have for 'horse', those dispositions which are *meaning-determining*. In other words, it must characterize, in non-intentional and non-semantic terms, a property *M* such that: possession of *M* is necessary and sufficient for being a disposition to apply an expression in accord with its correctness conditions.[43] Given such a property, however, could we not then safely equate meaning something by an expression with: *the set of dispositions with respect to that expression that possess M*? For, since dispositions with that property will be guaranteed to be dispositions to apply the expression correctly, both of the objections from normativity canvassed so far would appear to have been met. There will be no fear that the equation will issue in false verdicts about what the expression means. And, since it is only *M*-dispositions that are guaranteed to be correct, it will no longer follow that *whatever* seems right is right: those dispositions not possessing *M* will not be dispositions to apply the expression to what it means and will be free, therefore, to constitute dispositions to apply the expression falsely.

At this point two questions arise. First, is there really such a property *M*? And, second, supposing there were, is there really no more to capturing the normativity of meaning than specifying such a property?

Now, Kripke is clearly sceptical about the existence of an appropriate *M*-property. I will consider that question below. But more than this, Kripke seems to think that even if there were a suitably selected disposition that captured the extension of an expression accurately, that disposition could still not be identified with the fact of meaning, because it still remains true that the concept of a disposition is descriptive whereas the concept of meaning is not. In other words, according to Kripke, even if there were a dispositional predicate that logically covaried with a meaning predicate, the one fact could still not be identified with the other, for they are facts of distinct sorts. A number of writers have been inclined to follow him in this. Simon Blackburn, for instance, has written:

[43] It is occasionally suggested that it would be enough if possession of *M* were *sufficient* for the disposition's correctness. But that is not right. If only sufficiency were required we would not know, simply by virtue of a definition of *M*, the expression's meaning. For although we would know what properties were definitely part of the expression's meaning we would not know if we had them all. And so we would not have even a sufficient condition for the expression's possessing a given meaning.

I share Kripke's view that whatever dispositions we succeed in identifying they could at most give us standards for selection of a function which we mean. They couldn't provide us with an account of what it is to be faithful to a previous rule. It is just that, unlike Kripke, I do not think dispositions are inadequate to the task of providing standards. Indeed, I think they must be.[44]

Blackburn here is explicitly envisaging the successful, substantive specification of dispositions that mirror the extensions of expressions correctly. But he cites the normative character of facts about meaning as grounds for denying a dispositional *reduction*. But what precisely has been left over, once the extensions have been specified correctly?

One might have a thought like this. A proper reduction of the meaning of an expression would not merely specify its extension correctly, it would also reveal that what it is specifying is an *extension*—namely, a *correctness* condition. And this is what a dispositional theory cannot do. There might be dispositions that logically covary with the extensions of expressions; so that one could read off the dispositions in question the expressions' correctness conditions. But the dispositional fact does not amount to the meaning fact, because it never follows from the mere attribution of any disposition, however selectively specified, that there are facts concerning *correct* use; whereas this does follow from the attribution of an extension. To be told that 'horse' means *horse* implies that a speaker ought to be motivated to apply the expression only to horses; whereas to be told, for instance, that there are certain select circumstances under which a speaker is disposed to apply the expression only to horses, seems to carry no such implication.

It is not clear that this is in general true. Perhaps the *M*-dispositions are those dispositions that a person would have when his cognitive mechanisms are in a certain state; and perhaps it can be non-question-beggingly certified that that state corresponds to a state of the *proper* functioning of those mechanisms. If so, it is conceivable that that would amount to a non-circular specification of how the person would *ideally* respond, as compared with how he actually responds; and, hence, that it would suffice for capturing the normative force of an ascription of meaning.

There is clearly no way to settle the matter in advance of the consideration of particular dispositional proposals. What we are in a position to do, however, is state conditions on an adequate dispositional theory. First, any such theory must specify, without presupposing any semantic or intentional materials, property *M*. This would ensure the theory's extensional correctness. Second, it must show how possession of an *M*-disposition could amount to something that deserves to be called a *correctness* condition, something we would be inherently motivated to

[44] 'The Individual Strikes Back', loc. cit., pp. 289–91. Similar concessions are made by Wright in his 'Kripke's Account', loc. cit., pp. 771–2; and by John McDowell, 'Wittgenstein on Following a Rule', *Synthèse*, 1984, p. 329.

satisfy. This would ensure the intensional equivalence of the two properties in question, thus paving the way for an outright reduction of meaning to dispositions.

What property might *M* be? There are, in effect, two sorts of proposal: one, long associated with Wittgenstein himself, seeks to specify *M* by exploiting the notion of a community; the other, of more recent provenance, attempts to define *M* in terms of the notion of an optimality condition. I shall begin with the communitarian account.

The communitarian account

22. The idea that correctness consists in agreement with one's fellows has a distinguished history in the study of Wittgenstein. Even before the current concern with a 'rule-following problem', many commentators—whether rightly or wrongly—identified communitarianism as a central thesis of the later writings. As a response to the problem about meaning, it found its most sustained treatment in Wright's *Wittgenstein on the Foundations of Mathematics*.[45] Which of the many dispositions a speaker may have with respect to a given expression determine its meaning? Or, equivalently, which of the many dispositions a speaker may have with respect to an expression are dispositions to use it correctly? Wright's communitarian account furnishes the following answer:

> . . . it is a community of assent which supplies the essential background against which alone it makes sense to think of individuals' responses as correct or incorrect . . .. None of us can unilaterally make sense of correct employment of language save by reference to the authority of communal assent on the matter; and for the community itself there is no authority, so no standard to meet.[46]

It is important to understand that, according to the proposal on offer, the correct application of a term is determined by the *totality* of the community's *actual* dispositions in respect of that term. The theory does not attempt, in specifying the communal dispositions that are to serve as the constitutive arbiters of correctness, to select from among the community's actual dispositions a privileged subset. There is a reason for this. Communitarianism is a response to the perceived inability to define a distinction, at the level of the individual, between correct and incorrect dispositions. The suggestion that correctness consists in agreement with the dispositions of one's community is designed to meet this need. The proposal will not serve its purpose, however, if the problem at the level of the individual is now merely to be replayed at the level of the

[45] Cambridge, Harvard University Press, 1980. (His more recent writings suggest that Wright no longer holds this view.) See also Christopher Peacocke, 'Reply: Rule-Following: The Nature of Wittgenstein's Arguments', in *Wittgenstein on Following a Rule*, ed. Holtzman and Leich, London, Routledge and Kegan Paul, 1981.

[46] Ibid., pp. 219–20.

community. A communitarian does not want it to be a *further* question whether a given actual communal disposition is itself correct. The proposal must be understood, therefore, as offering the folowing characterization of *M*: *M* is the *property of agreeing with the actual dispositions of the community*.

How does the proposal fare with respect to the outlined adequacy conditions on dispositional theories?

Consider first the 'intensional' requirement, that possession of the favoured *M*-property appear intuitively to resemble possession of a correctness condition. Does *communal consensus* command the sort of response characteristic of truth?

A number of critics have complained against communitarianism that communal consensus is simply not the same property as truth, that there is no incoherence in the suggestion that all the members of a linguistic community have gone collectively, but non-collusively, off-track in the application of a given predicate.[47] This is, of course, undeniable. But the communitarian is not best read as offering an *analysis* of the ordinary notion of truth, but a *displacement* of it. His thought is that the emaciated notion of truth yielded by communitarianism is the best we can hope to expect in light of the rule-following considerations. The crucial question, then, is not whether communitarianism captures our ordinary notion of truth, for it quite clearly does not; it is, rather, whether communitarianism offers *any* concept deserving of that name.

This is a large question on which I do not propose to spend a lot of time.[48] Although there are subtle questions about how much of logic will be recoverable from such a view, and whether it can be suitably non-reductively articulated (can 'non-collusive agreement' be defined without the use of intentional materials?), I am prepared to grant, for the sake of argument, that the proposal does not fare all that badly in connection with the 'intensional' requirement. Non-collusive communal agreement on a judgement does usually provide one with some sort of reason for embracing the judgement (even if, unlike truth, not with a decisive one); it thus mimics to some degree the sort of response that is essential to truth. Where communitarianism fails, it seems to me, is not so much here as with the extensional requirement.

Consider the term 'horse'. What dispositions do I have in respect of this expression? To be sure, I have a disposition to apply it to horses. But I also have a disposition, on sufficiently dark nights, to apply it to deceptively horsey looking cows. Intuitively, the facts are clear. 'Horse' means *horse* and my disposition to apply it to cows on dark nights is

[47] See Blackburn, op. cit.

[48] For a more extensive discussion see my *Essays on Meaning and Belief*, loc. cit.; see also Blackburn, *Spreading the Word*, Oxford, Oxford University Press, 1984, pp. 82ff.

mistaken. The problem is to come up with a theory that delivers this result systematically and in purely dispositional terms. The communitarian's idea is that the correct dispositions are constitutively those which agree with the community's. What, then, are the community's dispositions likely to be?

The community, I submit, however exactly specified, is bound to exhibit precisely the same duality of dispositions that I do: it too will be disposed to call both horses and deceptively horsey looking cows on dark nights 'horse'. After all, if *I* can be taken in by a deceptively horsey looking cow on a dark night, what is to prevent 17,000 people just like me from being taken in by the same, admittedly effective, impostor? The point is that many of the mistakes we make are *systematic*: they arise because of the presence of features—bad lighting, effective disguises, and so forth—that have a generalizable and predictable effect on creatures with similar cognitive endowments. (This is presumably what makes 'magicians' possible.) But, then, any of my dispositions that are in this sense systematically mistaken, are bound to be duplicated at the level of the community. The communitarian, however, cannot call them *mistakes*, for they are the community's dispositions. He must insist, then, firm conviction to the contrary notwithstanding, that 'horse' means not *horse* but, rather, *horse or cow*.

The problem, of course, is general. There are countless possible impostors under countless possible conditions; and there is nothing special about the term 'horse'. The upshot would appear to be that, according to communitarianism, none of our predicates have the extensions we take them to have, but mean something wildly disjunctive instead. Which is to say that communitarianism is bound to issue in false verdicts about the meanings of most expressions, thus failing the first requirement on an adequate dispositional theory.

It seems to me that we have no option but to reject a pure communitarianism. If we are to have any prospect of identifying the extensions of our expressions correctly, it will simply not do to identify truth with communal consensus. Even from among the *community's* dispositions, we have to select those which may be considered meaning-determining, if we are to have a plausible theory of meaning. Which is to say that we are still lacking what communitarianism was supposed to provide: the specification of a property *M* such that, possession of *M* by a disposition is necessary and sufficient for that disposition's correctness.

Of course, once we have abandoned communitarianism, we lack any motive for defining *M* over *communal* dispositions; nothing—at least nothing obvious—tells against defining *M* directly over an *individual's* dispositions. Which is precisely the way the voluminous literature on this topic approaches the problem and to a discussion of which I now propose to turn.

Optimal dispositions

23. The literature supplies what is, in effect, a set of variations on a basic theme: *M* is the property of: being a disposition to apply (an expression) *in a certain type of situation*.[49] The idea behind such proposals is that there is a certain set of circumstances—call them 'optimality conditions'—under which subjects are, for one or another reason, incapable of mistaken judgements; hence, we may equate what they mean by a given (mental) expression with, the properties they are disposed to apply the expression to, under optimal conditions. Different proposals provide different characterizations of the conditions that are supposed to be optimal in this sense. Fred Dretske, for example, holds that optimal conditions are the conditions under which the meaning of the expression was first acquired. A number of other writers subscribe to some form or other of a *teleological* proposal: optimality conditions are those conditions—defined by evolutionary biology—under which our cognitive mechanisms are functioning just as they are supposed to.[50]

Now, Kripke is very short with such possible elaborations of a dispositional theory. He briefly considers the suggestion that we attempt to define idealized dispositions and says that 'a little experimentation will reveal the futility of such an effort'.[51] But, surely, this underestimates the complexity of the problems involved and fails to do justice to the influence that such proposals currently exert. What Kripke needs, if his rejection of dispositional accounts is to succeed, but does not really provide, is a set of principled considerations against the existence of non-semantically, non-intentionally specifiable optimality conditions. What I would like to do in the remainder of this section is to begin to sketch an argument for that conclusion. Several specific problems for specific versions of an optimality theory have received discussion in the literature.[52] Here, however, I want to attempt an argument with a more general sweep: I want to argue that we have reason to believe that there could not be naturalistically specifiable conditions under which a subject will be disposed to apply an expression only to what it means; and, hence, that no attempt at specifying such conditions can hope to succeed.[53]

[49] There is one exception to this generalization: Jerry Fodor's recent proposal has it that *S*'s meaning-determining are those that serve as an 'asymmetric dependence base' for *S*'s other dispositions. See his 'A Theory of Content', Part II, forthcoming. In 'Naturalizing Content', forthcoming in *Meaning in Mind: Essays on the Work of Jerry Fodor*, Oxford, Basil Blackwell, I argue that this theory is subject to the same difficulties as confront standard optimality versions.

[50] For theories of this form see: David Papineau, *Reality and Representation*, Oxford, Basil Blackwell, 1987; J. Fodor: 'Psychosemantics', MS, MIT, 1984. I shy away from saying whether R. Millikan, *Language, Thought and Other Biological Categories*, Cambridge, MIT Press, 1987, presents a theory of this form.

[51] K., p. 32.

[52] Against Dretske see Fodor, *Psychosemantics*, loc. cit.; against teleological theories see my *Essays on Meaning and Belief*, and Fodor, 'A Theory of Content', Part I, loc. cit.

[53] This amounts to saying that such theories cannot meet the extensional requirement; so I shall not even consider whether they meet the intensional one.

24. It will be worthwhile to lay the problem out with some care. Consider Neil and a particular expression, say, 'horse', in Neil's mental repertoire. And suppose that Neil is disposed to token that expression 'in the belief mode' both in respect of horses and in respect of deceptively horsey looking cows on dark nights. Let it be clear, furthermore, that 'horse' for Neil means *horse*, and that on those occasions when he applies 'horse' to cows, this amounts to his *mistaking* a cow for a horse. Now, the thought behind an optimality version of a dispositional theory is that there is a set of naturalistically specifiable conditions under which Neil cannot make mistakes in the identification of presented items.[54] Under those conditions, then, he would believe that there is a horse in front of him only if there is one. But that in turn implies that, under those conditions, 'horse' will get tokened (in the belief mode) only in respect of the property it expresses. So, to figure out what any expression means: look at the properties Neil is disposed to apply the expression to, when conditions are in this sense optimal. The end result is a dispositional reconstruction of meaning facts: for Neil to mean horse by 'horse' is for Neil to be disposed to call only horses 'horse', when conditions are optimal. Clearly, two conditions must be satisfied: (i) the specified conditions must really be such as to preclude the possibility of error—otherwise, it will be false that under those conditions 'horse' will get applied only to what it means; (ii) the conditions must be specified purely naturalistically, without the use of any semantic or intentional materials—otherwise, the theory will have assumed the very properties it was supposed to provide a reconstruction of.

What I propose to argue is that it is impossible to satisfy both of these conditions simultaneously.

Optimal dispositions and objective contents

25. The dispositionalist is after a non-semantically, non-intentionally specifiable set of conditions *O*, which will be such as to yield true, a priori *optimality equations* of the form:

(8) For any subject *S* and concept *R*: $O \rightarrow$(*S* judges $Rx \rightarrow Rx$).

Could there be such a set of conditions?

Notice, to begin with, that where *R* is the concept of an *objective* property, we ought not to expect optimality equations for *R*, even if *O* were not required to meet the rather stiff constraints imposed by a reductive dispositionalism—namely, specification in non-semantic and non-intentional terms. For, intuitively, the very idea of a wholly objective property (or object or relation) is the idea of a property (object, relation)

[54] This restriction to perceptually fixed beliefs stems partly from a desire to simplify exposition and partly from a desire to consider such theories at their strongest.

whose nature is independent of any given person's abilities or judgements: for such a property, in other words, there is no necessary function from a given person's abilities and judgements to truths about that property.[55] The contrast is with a class of contents for which there does exist a range of circumstances such that, appropriate subjects are necessarily authoritative about those contents under those circumstances. Philosophers disagree, of course, about what contents fall where, but it is typical to think of judgements about shape as wholly objective and of judgements about pain as representing an extreme example of the contrasting class. Let us call this a distinction between *accessible* versus *inaccessible* contents.[56]

We are now in a position to see, however, that a dispositional theory of meaning, by virtue of being committed to the existence of optimality equations for every concept, is committed thereby to treating every concept as if it were accessible. It is thus committed to obliterating the distinction between accessible and inaccessible contents.

Of course, this objection will not impress anyone reluctant to countenance wholly objective, inaccessible contents in the first place. I turn, therefore, to arguing against the dispositional theory on neutral ground: for any concept, subjective or objective, it is impossible to satisfy dispositionalism's basic requirement: the specification of a set of conditions *O*, *in non-semantic, and non-intentional terms*, such that, under *O*, subjects are immune from error about judgements involving that concept.

Optimal dispositions and belief holism

26. The basic difficulty derives from the holistic character of the processes which fix belief. The point is that, under normal circumstances, belief fixation is typically mediated by background theory—what contents a thinker is prepared to judge will depend upon what other contents he is prepared to judge. And this dependence is, again typically, arbitrarily robust: just about any stimulus can cause just about any belief, given a suitably mediating set of background assumptions. Thus, Neil may come to believe *Lo, a magpie*, as a result of seeing a currawong, because of his further belief that that is just what magpies look like; or because of his belief that the only birds in the immediate vicinity are magpies; or because of his belief that whatever the Pope says goes and his belief that the Pope says that this presented currawong is a magpie. And so on. The thought that something is a magpie can get triggered by a currawong in any of an indefinite number of ways, corresponding to the potentially indefinite number of background beliefs which could mediate the transition. Now, how does all this bear on the prospects for a dispositional theory of meaning?

A dispositional theorist has to specify, without use of semantic or intentional materials, a situation in which a thinker will be disposed to think, *Lo, a magpie* only in respect of magpies. But the observation that beliefs are fixed holistically implies that a thinker will be disposed to think *Lo, a magpie* in respect of an indefinite number of *non-magpies*, provided only that the appropriate background beliefs are present. Specifying an optimality condition for 'magpie', therefore, will involve, at a minimum, specifying a situation characterized by the absence of *all* the beliefs which could potentially mediate the transition from non-magpies to *magpie* beliefs. Since, however, there looks to be a potential infinity of such mediating background clusters of belief, a non-semantically, non-intentionally specified optimality situation is a non-semantically, non-intentionally specified situation in which it is guaranteed that none of this potential infinity of background clusters of belief is present. But how is such a situation to be specified? What is needed is precisely what a dispositional theory was supposed to provide: namely, a set of naturalistic necessary and sufficient conditions for being a belief with a certain content. But, of course, if we had *that* we would already have a reductive theory of meaning—we would not need a dispositional theory! Which is to say that, if there is to be any sort of reductive story about meaning at all, it cannot take the form of a dispositional theory.

VI
ANTI-REDUCTIONIST CONCEPTIONS OF MEANING

An argument from queerness?

27. If these considerations are correct, there would appear to be plenty of reason to doubt the reducibility of content properties to naturalistic properties. But Kripke's sceptic does not merely draw an anti-reductionist conclusion; he concludes, far more radically, that there simply could not *be* any content properties. Suppose we grant the anti-reductionism; what justifies the content *scepticism*? Not, of course, the anti-reductionism by itself. At a minimum one of two further things is needed. Either an independent argument to the effect that only naturalistic properties are real. Or, failing that, a frontal assault on the irreducible property in question, showing that it is, in Mackie's phrase, somehow inherently 'queer'.

[55] See, for example, Tyler Burge, 'Cartesian Error and the Objectivity of Perception', in *Subject, Thought and Context*, ed. J. McDowell and P. Pettit, Oxford, Clarendon Press, 1986, p. 125, for a similar formulation of the concept of objective an property.

[56] It is important to appreciate that this is an *epistemological* distinction, not a constitutive one. It does not follow from the fact that a content is accessible, that it is therefore constituted by our best judgements about it. (I take it no one is tempted to conclude from the fact that we are authoritative about our pains, that pains are *constituted* by the judgements we make about them. We shall have occasion to discuss constitutive claims of this sort later on in the paper.

The single greatest weakness in Kripke's sceptical argument is that he fails to bring off either requirement. He does not even try to defend a reductionist principle about the intentional; and his brief attempt at a 'queerness' argument is half-hearted and unconvincing:

> Perhaps we may try to recoup, by arguing that meaning addition by 'plus' is a state even more sui generis than we have argued before. Perhaps it is simply a primitive state, not to be assimilated to sensations or headaches or any 'qualitative' states, nor to be assimilated to dispositions, but a state of a unique kind of its own.
>
> Such a move may in a sense be irrefutable, and if it is taken in an appropriate way Wittgenstein may even accept it. But it seems desperate: it leaves the nature of this postulated primitive state—the primitive state of 'meaning addition by "plus"'—completely mysterious. It is not supposed to be an introspectible state, yet we supposedly are aware of it with some fair degree of certainty whenever it occurs. For how else can each of us be confident that he does, at present, mean addition by 'plus'? Even more important is the logical difficulty implicit in Wittgenstein's sceptical argument. I think that Wittgenstein argues, not merely as we have said hitherto, that introspection shows that the alleged 'qualitative' state of understanding is a chimera, but also that it is logically impossible (or at least that there is a considerable logical difficulty) for there to be a state of 'meaning addition by "plus"' at all.
>
> Such a state would have to be a finite object, contained in our finite minds. It does not consist in my explicitly thinking of each case of the addition table. . . . Can we conceive of a finite state which *could* not be interpreted in a quus-like way? How could that be?[57]

There are several problems with this passage. In the first place, it misconstrues the appropriate anti-reductionist suggestion. I take it that it really is not plausible that there are 'primitive states' of meaning *public language* expressions in certain ways, one state per expression. The process by which the inscriptions and vocables of a public language acquire meaning is a manifestly complex process—involving an enormous array of appropriate propositional attitudes—the outlines of which may arguably be found in the writings of Paul Grice and others.[58] A plausible anti-reductionism about meaning would not wish to deny that there is an interesting story to be told about the relation between *linguistic content and mental content*; what it maintains, rather, is that there is no interesting reduction *of mental content properties* to *physical/functional properties*. According to anti-reductionism, in other words, at some appropriate level mental content properties must simply be taken for granted, without prospect of identification with properties otherwise described.

Does Kripke manage to create a difficulty for *this* suggestion? The passage contains a couple of considerations that may be so construed.

The first charge is that we would have no idea how to explain our ability

[57] K., pp. 51–2.

[58] See the papers cited under n. 3 above.

to know our thoughts, if we endorsed a non-reductionist conception of their content. Now, no one who has contemplated the problem of self-knowledge can fail to be impressed by its difficulty.[59] But I think that we would be forgiven if, before we allowed this to drive us to a dubiously coherent irrealism about content, we required something on the order of a *proof* that no satisfactory epistemology was ultimately to be had.

Kripke, however, provides no such proof. He merely notes that the non-phenomenal character of contentful states precludes an introspective account of their epistemology. And this is problematic for two reasons. First, because there may be non-introspective accounts of self-knowledge.[60] And second, because it does not obviously follow from the fact that a mental state lacks an individuative phenomenology, that it is not introspectible.[61]

Kripke's second objection to the anti-reductionist suggestion is that it is utterly mysterious how there could be a finite state, realized in a finite mind, that nevertheless contains information about the correct applicability of a sign in literally no end of distinct situations. But, again, this amounts merely to insisting that we find the idea of a contentful state problematic, without adducing any independent reason why we should. We *know* that mental states with general contents are states with infinitary normative characters; it is precisely with that observation that the entire discussion began. What Kripke needs, if he is to pull off an argument from queerness, is some substantive argument, distinct from his anti-reductionist considerations, why we should not countenance such states. But this he does not provide.

None of this should be understood as suggesting that an anti-reductionism about content is unproblematic, for it is far from it. There are, for example, familiar, and serious, difficulties reconciling an anti-reductionism about content properties with a satisfying conception of their causal efficacy.[62] But in the context of Kripke's dialectic, the anti-reductionist suggestion emerges as a stable response to the sceptical conclusion, one that is seemingly untouched by all the considerations adduced in the latter's favour.

[59] For discussion of some of the difficulties see my 'Content and Self-Knowledge', *Philosophical Topics*, Spring 1989.

[60] See, for example, Tyler Burge, 'Individualism and Self-Knowledge', *Journal of Philosophy*, November 1988, and D. Davidson, 'Knowing One's Own Mind', *Proceedings of the APA*, January 1987.

[61] It is interesting to note, incidentally, that one of the more striking examples of the introspective discernment of a non-qualitative mental feature is provided by, of all things, an experiential phenomenon. I have in mind the phenomenon, much discussed by Wittgenstein himself, of seeing-as. We see the duck-rabbit now as a duck, now as a rabbit; we see the Necker cube now with one face forward, now with another. And we know immediately precisely how we are seeing these objects as, when we see them now in one way, now in the other. But this change of 'aspect', although manifestly introspectible, is nevertheless not a change in something qualitative, for the qualitative character of the visual experience remains the same even as the aspect changes.

[62] See below.

McDowell on privacy and community

28. If we endorse a non-reductionist conception of meaning, does that mean that the rule-following considerations disturb nothing in our ordinary conception of that notion? A number of writers who have found an anti-reductionist suggestion attractive have certainly not thought so; they have discerned in those considerations important lessons for the correct understanding of the possibility of meaning, while rejecting substantive reductive answers to the constitutive question: in virtue of what do expressions possess meaning?

John McDowell, for example, has written that:

> By Wittgenstein's lights, it is a mistake to think we can dig down to a level at which we no longer have application for normative notions (like 'following according to the rule').[63]

We have to resist the temptation, according to McDowell's Wittgenstein, to form a picture of 'bedrock'—'of how things are at the deepest level at which we may sensibly contemplate the place of [meaning] in the world'—which does not already employ the idea of the correct (or incorrect) use of an expression.

Oddly, however, McDowell does not take this to commit him to a *quietism* about meaning, a position from which no substantive results about the conditions for the possibility of meaning can be gleaned. On the contrary, he claims that it is the discernible moral of the rule-following considerations that correctness, and hence meaning, can exist only in the context of a *communal practice*, thus precluding the possibility of a private language. He writes:

> Wittgenstein warns us not to try to dig below 'bedrock'. But it is difficult, in reading him, to avoid acquiring a sense of what, as it were, lies down there: a web of facts about behavior and 'inner' episodes, describable without using the notion of meaning. One is likely to be struck by the sheer contingency of the resemblances between individuals on which, in this vision, the possibility of meaning seems to depend . . ..[64]

And:

> It is true that a certain disorderliness below 'bedrock' would undermine the applicability of the notion of rule-following. So the underlying contingencies bear an intimate relation to the notion of rule-following . . ..[65]

This is, of course, McDowell's characterization of the familiar Wittgensteinian claim that a certain measure of agreement in communal responses is a precondition for meaning. But how is such a thesis to be motivated? How, in light of the rejection of substantive answers to the constitutive question, is it to be argued for? The claim that communal practice is necessary for meaning is a *surprising* claim; mere reflection on the concept of meaning does not reveal it. And what, short of a substantive constitutive account, could conceivably ground it?

[63] 'Wittgenstein on Following a Rule', loc. cit., p. 341.
[64] Ibid., p. 348.
[65] Ibid., p. 349.

Consider the contrast with the communitarian view considered above. That view engages the constitutive question, offers a substantive answer to it, and generates, thereby, a straightforward argument for the necessity of a communal practice: since correctness is said to *consist* in conformity with one's fellows, correctness, and with it meaning, are possible only where there are others with whom one may conform. But McDowell, rightly in my view, rejects the suggestion that correct application might be analysed in terms of communal dispositions. Indeed, as I have already noted, he rejects the very demand for a substantive account of correctness: norms are part of the 'bedrock', beneath which we must not dig. But if we are simply to be allowed to take the idea of correctness for granted, unreduced and without any prospect of reconstruction in terms of, say, actual and counterfactual truths about communal use, how is the necessity of an 'orderly communal' practice to be defended? From what does the demand for orderliness flow? And from what the demand for community? McDowell's paper contains no helpful answers.[66]

Wright on the judgement-dependence of meaning

29. Crispin Wright has written about the anti-reductionist conception that:

> [t]his somewhat flat-footed response to Kripke's Sceptic may seem to provide a good example of 'loss of problems.' . . . 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 given can *be* correct.[67]

Wright's intriguing suggestion is that there are important constitutive results to be gleaned from the epistemological question we shelved some pages back: namely, how, if content properties are simply to be taken for

[66] Though see his remarks—which I am afraid I do not understand—on a 'linguistic community [that] is conceived as bound together, not by a match in mere externals (facts accessible to just anyone), but by a meeting of minds'. McDowell's problems here echo, I think, Wittgenstein's own. The main difficulty confronting a would-be interpreter of Wittgenstein is how to reconcile his rejection of substantive constitutive accounts—especially of meaning, see *Zettel* #16: 'The mistake is to say that there is anything that meaning something consists'—with the obvious constitutive and transcendental pretensions of the rule-following considerations. It is fashionable to soft-pedal the rejection of constitutive questions, representing it as displaying a mere 'distrust' on Wittgenstein's part. But this ignores the fact that the rejection of analyses and necessary and sufficient conditons is tied to extremely important first-order theses about meaning, including, most centrally, the family-resemblance view of concepts.

[67] 'Wittgenstein's Rule-Following consideration and the Central Project of Theoretical Linguistics', loc. cit., p. 237.

granted, without prospect of reconstruction either in experiential or dispositional terms, can they be known? As we saw, Kripke attempted to use this question to embarrass his anti-reductionist opponent. Wright, however, has a more constructive project in mind. Pressing the epistemological question will reveal, so he claims, that facts about content are essentially 'judgement-dependent'.

What does it mean for a class of facts to be judgement-dependent? Wright's explanation is framed in terms of a failure to pass the 'order-of-determination test':

> The order-of-determination test concerns the relation between *best* judgements—judgements made in what are, with respect to their particular subject matter, *cognitively ideal* conditions of both judger and circumstance—and truth. . . . 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.[68]

We may explain the contrast Wright has in mind here by recurring to the idea of an accessible content (see above). An accessible content is one about which subjects are necessarily authoritative under cognitively optimal circumstances. Now, a question may be raised about the correct explanation for this authority: is it that, under those optimal circumstances, subjects are exceptionally well-equipped to *track* the relevant, independently constituted facts? or is it, rather, that judgements under those circumstances simply *constitute* the facts in question? A fact is judgement-independent if the former, judgement-dependent if the latter.

The contrast, then, is between facts which are constituted independently of our judgements, however optimal, and facts which are constituted precisely by the judgements we would form under cognitively ideal circumstances. And the claim is that facts about content have to be construed on the latter model. *Pace* Kripke, the target of the rule-following considerations is not the reality of content facts, but, rather, a judgement-independent (or *Platonist*, if you think these come to the same thing) conception of their *constitution*. Best judgements constitutively determine the truth-value of sentences ascribing content to mental states; they do not *track* independently constituted states of affairs which confer truth or falsity upon them.

Wright argues for this 'judgement-dependent' conception of content by attacking the epistemologies available on the alternative model. Drawing extensively on Wittgenstein's actual text, Wright reconstructs an interesting set of considerations against both *introspective* and *inferential* conceptions of self-knowledge, thus, presumably, exhausting the epistemologies

[68] Ibid., p. 246.

available to his opponent. So long as facts about our mental states are construed as independent of, and, hence, as tracked by our self-regarding judgements, we can have no satisfactory explanation of our ability to know them. On the assumption, then, that Kripke's unstable content irrealism is to be avoided at all costs, that leaves the judgement-dependent conception as the only contender. So goes Wright's argument.

Wright's discussion raises a number of interesting and difficult questions. Is it really true that Wittgenstein's discussion destroys all 'cognitive accomplishment' theories of self-knowledge? Supposing it does, does this inevitably drive us to a judgement-dependent conception of content? Are there not other conceptions that would equally accommodate the rejection of a tracking epistemology? Unfortunately, none of these questions can be adequately addressed within the confines of the present essay. Here I have to settle for raising a question about whether a judgement-dependent conception of *content* could *ever* be the cogent moral of any argument.

30. The suggestion is that we must not construe facts concerning mental content as genuine objects of cognition, and that this is to be accomplished by regarding them as constituted by truths concerning our best judgements about mental content. Well, what does this amount to? For illustrative purposes, Wright offers the case of colour. What would have to be true, if facts about colour are to judgement-dependent? We would need, first and foremost, to secure the *accessibility* of colour facts, and so a biconditional of the following form:

> if *C*: *S* would judge *x* to be blue↔*x* is blue.

But not just any biconditional of this form will serve to secure the accessibility of colour. For example, unless restrictions are placed on the permissible specifications of *C*, *every* property will turn out to be accessible; just let *C* be: conditions under which *S* is infallible about colour. So, it must be further required that *C* be specified in substantial terms, avoiding a 'whatever-it-takes' formulation.

Now, what it would take to ground not merely the accessibility of colour facts, but their *judgement-dependence*? What is needed, as Wright points out, 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.[69]

This seems right. For unless the specification of the *C*-conditions, or, indeed, of anything else on the left-hand-side, is precluded from presupposing facts about the colours of objects, it will remain entirely open whether subjects' judgements, formed under the relevant *C*-conditions,

[69] Ibid., pp. 247–8.

really did determine facts about colour. For satisfaction of the conditions described on the left-hand-side would always presuppose some antecedently fixed constitution of colour facts, thus undermining the claim that it is precisely truths about best judgement that fix those facts.

No doubt, other requirements are in order as well.[70] But it is, I trust, already clear that there is a serious difficulty seeing how facts about mental content could conceivably satisfy the stated requirements on judgement-dependence. For it is inconceivable, given what *judgement-dependence* amounts to, that the biconditionals in the case of mental content should satisfy the requirement that their left-hand-sides be free of any assumptions about mental content. For, at a minimum, the *content of the judgements* said to fix the facts about mental content have to be presupposed. And that means that any such biconditional will always presuppose a constitution of mental content quite independent of constitution by best judgement.

In a way, an intuitive difficulty should have been clear from the start. A 'judgement-dependent' conception of a given fact is, by definition, a conception of that fact according to which it is *constituted* by our *judgements*. The idea is clearly appropriate in connection with facts about the *chic* or the *fashionable*; familiar, though less clearly appropriate, in connection with facts about colour or sound; and, it would appear, impossible as a conception of facts about mental content. For it cannot in general be true that facts about content are constituted by our judgements about content: facts about content, constituted independently of the judgements, are presupposed by the model itself.

Conclusion: robust realism—problems and prospects

31. Let *robust realism* designate the view that judgements about meaning are factual, irreducible, and judgement-independent. Then the moral of this paper—if it has one—is that the major alternatives to robust realism are beset by very serious difficulties.

Irrealism—the view, advocated by Kripke's Wittgenstein, that judgements about meaning are non-factual—appears not even to be a coherent option. (An error-theoretic variant, as promoted, for example, by Paul Churchland, seems no better.)

Reductionist versions of realism appear to be false. The proposal that judgements about meaning concern communal dispositions is unsatisfactory not merely because, implausibly, it precludes the possibility of communal error, but because it appears bound to misconstrue the meaning of every expression in the language. The rather more promising (and rather more popular) proposal, that judgements about meaning concern a

[70] For a very illuminating discussion of the conditions that would have to be met, see ibid., pp. 246–54.

certain sort of *idealized* disposition, also appears to confront serious difficulties: it is hard to see how the idealizations are to be specified in a non-question-begging way.

And, finally, a *judgement-dependent* conception of meaning seems not to be a stable option, because the very idea of constitution by best judgement appears to presuppose a judgement-*independent* conception of meaning.

It is sometimes said that an anti-reductionist conception is too facile a response to the problem about meaning. It is hard not to sympathize with this sentiment. But if the considerations canvassed against the alternatives are correct, and if it is true that the 'rule-following' considerations leave an anti-reductionist conception untouched, it is hard, ultimately, also to agree with it. Meaning properties appear to be neither eliminable, nor reducible. Perhaps it is time that we learned to live with that fact.

I do not pretend that this will be easy. Robust realism harbours some unanswered questions, the solutions to which appear not to be trivial. There are three main difficulties. First: what sort of room is left for *theorizing* about meaning, if reductionist programs are eschewed? Second: how are we to reconcile an anti-reductionism about meaning properties with a satisfying conception of their causal or explanatory efficacy? And, finally: how are we to explain our (first-person) knowledge of them?

I cannot, of course, hope to address any of these questions adequately here. A few brief remarks will have to suffice.

To begin with the last question first, I cannot see that an anti-reductionist conception of content has a *special* problem about self-knowledge. As far as I am concerned, no one has a satisfactory explanation of our ability to know our own thoughts.[71] But I do not see that the anti-reductionist need feel any special embarrassment about this. If anything, it seems to me, the prospects are better for him than for his opponent. A reductionist would have it that meanings are fixed by certain kinds of dispositional fact, the sort of fact that could hardly be known observationally. It would appear to follow that the reductionist is committed, if he is to have a substantial epistemology of self-knowledge, to an inferential conception—a conception that may be, as I have argued elsewhere, worse than implausible.[72] The anti-reductionist labours under no comparable burden.

As for the charge that there would be nothing left for a theory of meaning to be, if reductionism is eschewed, it seems to me simply false. Let me here mention just a few of the questions that survive the rejection of reductionist programmes. For one thing, as I have stressed, a non-reductionism about meaning is best understood as a thesis about *mental* meaning, not about linguistic meaning. So anti-reductionism, as I understand it, is not only consistent with, but positively invites, a theory about

[71] See my 'Content and Self-Knowledge', loc. cit.

[72] Again see my ibid.

the relation between thought and language. How do public language symbols come to acquire meaning and what role does thought play in that process? Secondly, anti-reductionism in my sense is consistent with wanting a general account of the principles by which we interpret other people. The important work of Quine, Davidson, Lewis, Grandy, and others on the theory of radical interpretation neither needs, nor is best understood in terms of, reductionist aspirations. Its proper goal is the articulation of the principles we evidently successfully employ in interpreting the speech and minds of others. And, finally, an anti-reductionism about mental content is perfectly consistent both with substantive theories of the nature of the propositional *attitudes*—that is, of what makes a given mental state a *belief*, as opposed to a wish or a desire; and with the claim that the grasping of certain mental contents depends on the grasping of others, and so with theories of the *compositional structure* of mental content.

There is hardly any fear, then, that we shall run out of things to do, if we forego reductionist programmes in the theory of mental content.

Finally, though, there is the question of mental causation: how are we to reconcile an anti-reductionism about content properties with a satisfying conception of their causal efficacy? It is a view long associated with Wittgenstein himself, of course, that propositional attitude explanations are not causal explanations. But, whether or not the view was Wittgenstein's, it has justifiably few adherents today. As Davidson showed, if propositional attitude explanations are to rationalize behaviour at all, then they must do so by causing it.[73] But propositional attitudes rationalize partly by virtue of their content—it is partly because Neil's belief is *that there is wine in his glass*, that he reaches for it; so, propositional attitude explanations commit us to holding that content properties have a genuine causal role in the explanation of intentional action. But, now, how is an anti-reductionist about content properties to accord them a genuine causal role without committing himself, implausibly, to the essential incompleteness of physics?

This is, I believe, the single greatest difficulty for an anti-reductionist conception of content. It may be that it will eventually prove its undoing. But the subject is relatively unexplored, and much interesting work remains to be done.[74]

Department of Philosophy PAUL A. BOGHOSSIAN
University of Michigan
Ann Arbor, MI, USA 48109

[73] See 'Actions, Reasons and Causes', in his *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

[74] For some recent papers see E. LePore and B. Loewer, 'Mind Matters', *Journal of Philosphy*, 1987, and Jerry Fodor, 'Making Mind Matter More', *Philosophical Topics*, Spring 1989.