1998

Skepticism about Meaning: Indeterminacy, Normativity, and the Rule-Following Paradox

SCOTT SOAMES
Princeton University

Quine and Kripke's Wittgenstein both present "skeptical" arguments for the conclusion that there are no facts about meaning.[1] In each case the argument for the conclusion is that (i) if there are facts about meaning (and propositional attitudes), then they must be determined by some more fundamental facts, but (ii) facts about meaning (and propositional attitudes) are not determined by any such facts. Consequently there are no facts about meanings (or propositional attitudes). Within this overall framework, Quine and Kripke's Wittgenstein differ substantially — both in their reasons for thinking that facts about meaning (and propositional attitudes) are not determined by other facts, and in their responses to the alleged elimination of these facts. Despite this, I believe that their arguments fail for essentially the same reason; each equivocates about what it means for one set of facts to determine another. Once the equivocation is eliminated, the arguments lose their plausibility.

---

1 In discussing 'Kripke's Wittgenstein' I have in mind the philosophical position presented by Saul Kripke in chapters 2 and 3 of *Wittgenstein On Rules and Private Language*, (Cambridge, MA: Harvard University Press 1982). Kripke presents this position both as an interpretation of the leading ideas of Wittgenstein's *Philosophical Investigations* and as a philosophical point of view of independent interest. My discussion of the position will be concerned only with its content and merits, not its origin. The question of whether the position outlined by Kripke is an accurate interpretation of Wittgenstein is not directly relevant to my discussion.

Kripke's Skeptical Argument

I begin with Kripke's Wittgenstein. His argument is based on an unde-niable truism: What we mean by a word is not exhausted by the cases in which we, or those who have taught us the word, have actually used it. Rather, what we mean somehow determines the correct application of the word to an indefinite range of new, so-far unencountered, cases. Kripke's skeptical argument challenges us to explain how this comes about. What is it about us that determines that the word, as we now use it, already applies in a definite way to cases we have not yet considered? If we mean anything at all by the word, then something must determine this, for otherwise we would be free to apply the word in new cases in any way we liked, without changing its previous meaning (or saying anything false). But surely we are not free to do this. So, if it is a fact that we mean so and so by a given word *w*, then some fact about us must determine in advance how *w* properly applies in new cases. This much seems undeniable. The surprise comes when we examine potential candidates for such a determining fact and find that none fills the bill. Because of this, the skeptic concludes, we have no choice but to admit that it is not a fact that we mean anything by *w* after all.

Kripke uses '+' to illustrate the point. This symbol, as we normally understand it, stands for the addition function, which assigns a unique natural number to each of infinitely many pairs of natural numbers. Thus, its range of application far exceeds the (relatively) small number of cases in which we have actually used it in computing sums. Conse-quently if we really do mean the addition function by '+,' then some-thing about us must determine that when '+' is applied to a pair of numbers we have not previously considered it always yields the sum of those numbers, rather than nothing, or the result of applying some other arithmetical operation. But, the skeptic argues, nothing about us does determine this.

Wait a moment, one might object: Isn't the skeptic's position self-defeating? In setting up the case, he uses the words *sum* and *the addi-tion function* to pick out a certain function, which he himself seems to characterize as applying in a definite way to infinitely many pairs of arguments. The problem, of course, is that if he is right, this is pre-cisely the sort of thing that cannot be done. Thus, doesn't his own ar-gument presuppose the falsity of its conclusion?

The short answer is that, yes, in my opinion, there is an important sense in which the skeptic's argument is self-undermining; if one suc-ceeds in stating the argument, or in using it to justify one's belief in the conclusion, then that very act falsifies the conclusion.[2] The real chal-lenge is to find where the argument goes wrong; and to do this it is helpful to have a formulation which is not (immediately) self-defeat-ing. Kripke provides one. In his formulation we assume that there is no doubt about what we *now* use our words to mean. In particular, we all agree that at present we use '+' to stand for the addition function, which applies in a definite and prescribed way to infinitely many cases. All arithmetical facts about this function, including all results of ap-plying it to particular numbers, are taken as given, and are not in dis-pute. What is in dispute is what we meant by '+' *in the past.*[3]

Here we face the familiar problem. Although I performed lots of calculations in the past using '+,' there are infinitely many pairs of ar-guments that I never applied it to. Pick one such pair — say, 68 and 57. Doing the calculation now, I will tell you '68 + 57 = 125.' Everyone grants that this is arithmetically correct. However the skeptic raises the question whether it is correct in a certain metalinguistic sense.[4] Is it the case that in the past I used the '+' sign to denote a function that gives 125 as value to the pair of arguments 68 and 57? More generally, is it the case that in the past I used '+' to denote the same function I now do, the addition function?[5] If so, then some fact about my past

---

2 Stating the skeptic's argument is not just a matter of uttering the words, but rather involves taking up genuine propositional attitudes toward the contents expressed by those words.

3 The advantage of this way of formulating the problem is that it allows us to take the meanings of the words we use in stating the argument for granted while the argument is being given and evaluated. Hence we do not have to rely on the meanings of certain words in stating the argument while at the same time questioning what those words mean.

4 Kripke introduces this sense of metalinguistic correctness on 8.

5 The skeptic's challenge is made graphic by his suggestion that perhaps in the past I used '+' to denote not the addition function but the quaddition function, where the latter differs from the former in assigning the value 5 to all pairs of arguments greater than any arguments I had explicitly used in calculations. It is assumed, for the sake of discussion, that 68 and 57 are such a pair.

understanding of the '+' sign must have determined that '125' was the right answer to the question 'What is 68 + 57?' even though I had never considered that case. The skeptic challenges us to find such a fact.

Certain natural responses are immediately disqualified. If, outside the context of the skeptical challenge, one were asked what past facts determined what we meant by a particular word, it might be natural to answer that the determining facts were facts about the beliefs and intentions connected with our use of sentences containing the word. For example, one might claim that my having meant addition by '+' is a consequence of the fact that in the past when I used sentences containing '+' I did so with the intention of communicating my belief concerning the result of *adding* certain numbers, rather than the result of applying some other function. Because of this, it was addition that I meant by '+.'

Whether or not such a response is true, it is of no help in dealing with Kripke's skeptical challenge. The skeptic raises a question about the content of the words we use. He does so knowing all the nonintentional facts there are about our past uses of the word — our linguistic performances, our behaviour, and so on. Any attempt to answer his question by citing our past beliefs and intentions will be met with a mere reformulation of the original challenge. Granted, in the past we used the words *9 + 16 = 25* to express a belief. But what belief? A belief about addition, or a belief about some other function? To settle this question we must find some fact that determined the contents of our past beliefs and intentions. But this is just our original problem all over again. Having started by challenging us to find facts that determined the content of our words, the skeptic continues by challenging us to find facts that determined the contents of our mental states. Thus, if we are ever going to be able to answer the skeptic, we will have to find facts that determine content in general — linguistic and otherwise.

The same sort of difficulty plagues other natural responses to the skeptic. For example, in the case of '+' it is probably true that for most of us coming to understand the term was associated with learning a rule or algorithm for applying it. When confronted with a new calculation, we simply apply the algorithm and obtain the desired result. Since, in principle, the algorithm applies to infinitely many cases, it might seem ideally suited to answering the skeptic. One might say, for example, that the fact about me that determined that in the past I used '+' to stand for a function that assigns the value 125 to the arguments 68 and 57, even before I performed the actual calculation, was simply the fact that the algorithm I associated with '+' gives the result 125 in that case. The same point obviously can be made for all other cases. Thus, one might say, the fact that I meant *addition* by '+' was determined by the fact that the algorithm I associated with it in the past always yields the *sum* of the arguments to which it is applied.

But what is an algorithm anyway? Is it a collection of symbols (the symbols we use to express it), or is it the content of those symbols (what those symbols express)? If we think of the algorithm as just a collection of symbols, without any particular interpretation, then it won't determine the value of any function at any argument, and so will provide no answer to the skeptic. If, on the other hand, we think of the algorithm as the content of the symbols used in stating it, then the skeptic will question whether we did associate that content with '+' in the past. He does not doubt that in the past we associated the same symbols with '+' as we do now — symbols which, under their present interpretation, express an algorithmic content that determines the addition function. However he does doubt that we interpreted those symbols in the past in the same way we do now. In particular he doubts that, as we interpreted the symbols of the algorithm in the past, they provided a procedure that always yields the sum of a pair of numerical arguments.

Our problem, then, is this. We think that our understanding of '+' determines how it applies to an indefinitely large range of cases, even though at any given time we have considered only a small number of these. Thus, the fact that we mean one thing by '+' rather than another cannot consist simply in the uses we have already made of it in particular calculations. What then does it consist in? A natural thought is that it consists in associating '+' with a linguistic rule that determines its correct use. However this only postpones the problem, since we can now ask what our understanding of the symbols used in formulating the rule consists in. At some point this process of using symbols to interpret other symbols must stop. And when it does, we have to face the skeptic's challenge directly. In general, if we ever mean anything by any expression, then there must be some expressions whose meanings are not determined either by the particular

occasions we have used them in the past, or by any linguistic rules we have associated with them, or by any introspectable, content-bearing, mental images or psychological representations that may accompany their use.[6]

This realization leads to a different attempt to answer the skeptic — one that Kripke calls the dispositional view. Applied to '+,' it says that for us to use '+' to denote one function rather than another is for us to have certain mental and behavioural dispositions. Specifically, in the past I used '+' to denote the addition function in virtue of the fact that I was disposed when asked any question of the form *What is m + n?* to answer by giving the sum of *m* and *n*. Thus, even though in the past I never was asked the question *What is 68 + 57?*, if I had been asked, I would have answered *125* — just as I do now . Although the particular calculations I have performed have changed over time, my dispositions to answer arithmetical questions in specific ways has not. Consequently, the dispositionalist concludes, I meant the same thing by '+' in the past as I do now.

Paraphrasing Kripke, we may take (D) to be a preliminary statement of the dispositional analysis of '+.'

(D)   One uses '+' to denote a 2-place numerical function f iff for all natural numbers m, n, and z, and numerals $\underline{m}$, $\underline{n}$, and $\underline{z}$ denoting

---

6   Kripke's skeptic raises a challenge about content-bearing (i.e. intentional) facts. The challenge is to locate non-content-bearing (i.e. nonintentional) facts that determine the content-bearing ones. In attempting to meet this challenge, the nonintentional facts that one is allowed to appeal to are not restricted to behavioural facts, publicly observable facts, or even physical facts. Mental images, representations, feelings, and sensations can all be appealed to, provided that their 'interpretations,' or contents, (if any) are not taken for granted, but are themselves given a thoroughly non-intentional explanation.

In certain cases, it is quite plausible to suppose that factors such as linguistic rules or representations really do play a role in determining the meanings of some expressions. A case in point is '+' itself. However in this, as in other cases, the role is dependent upon a prior explanation of the contents of the relevant rules, algorithms, images, or representations.

---

them, one is disposed to answer $\underline{z}$ when asked the question 'What is $\underline{n} + \underline{m}$?' iff $f(m,n) = z$.[7]

It is worth noting that on this account, the fact about me that determined what I meant by '+' was an intentional fact that is itself subject to skeptical worries of two different sorts. First, the disposition to *answer* a question in a certain way is a disposition to *assert* a certain proposition in answer to the question. But what, the skeptic will demand, justifies our characterizing an event that consists of my producing a numeral $\underline{z}$ in response to the utterance of an interrogative sentence, 'What is $\underline{n} + \underline{m}$?', as my answering the question — by asserting the proposition that n + m = z — as opposed to emitting a mere verbal reflex, or giving an arbitrary response without meaning anything? A thoroughgoing skeptic about meaning and propositional attitudes would not take this for granted. Consequently our statement of the dispositional view already presupposes more than the skeptic would allow. Second, in order to use the analysis to determine the function meant from my verbal performances one would have to be given the denotations of the numerals and other symbols used in questions and answers. These are just as subject to skeptical doubt as the interpretation of '+.'

For these reasons, the dispositional analysis (D) stated here is itself too intentional to succeed as a response to the skeptic. However, I will put this objection aside temporarily in order to review two other objections discussed by Kripke — the error and finitude objections. After looking at these, our worries about the intentionality built into (D) will resurface in the discussion of Kripke's normativity objection.

The error objection is that the analysis does not make room for dispositions to misapply the expression. Since, according to it, any answer I am disposed to give to a particular query, 'What is $\underline{n} + \underline{m}$?', is definitive of what I mean by '+,' all such answers must be correct. But this is wrong. One can mean addition by '+' even if one is disposed to make a few systematic mistakes in particular cases, and hence is disposed to give incorrect answers to some questions of the form 'What is

---

7   Kripke, 26

$n + m$?'. The finitude objection is that the analysis cannot account for the fact that our dispositions are finite, and will give out before the application of the term does. Certainly, one can mean addition by '+' despite the fact that at some point the numbers to be added get too large for us to deal with. When this happens we have no dispositions to answer the relevant questions 'What is $n + m$?', even though the addition function denoted by '+' continues to be defined.[8]

For these reasons the dispositional analysis, in the form we have stated it, is surely inadequate.[9] Nevertheless, it is difficult to resist the idea that our dispositions to apply a term in particular cases play some

8 Both the error and the finitude objections can be generalized to cover dispositional analyses of other terms. In this connection, it is important to note that the finitude objection to the dispositional analysis of '+' is just one instance of a more general objection. The objection is that a term may apply to certain objects even in cases in which those objects are, for one reason or another, epistemically inaccessible to us, and hence are not objects about which we have any normal dispositions. In the case of '+' the infinity of natural numbers ensures that at some point they will get too big for us to consider. In the case of other words the reasons for epistemic inaccessibility may be quite different.

9 One possible response to these objections would be to give up the dispositional analysis for '+,' while retaining it for simpler notions like 'successor,' in terms of which '+' could then be defined. Certainly, there is less room for error in applying the successor function than in applying the addition function. However, some room for error may remain, and, in any case, the finitude objection still seems to apply. Another possibility, considered by Kripke, is to reformulate the analysis so as to appeal to idealized dispositions — dispositions to answer questions 'What is $n + m$?', not in just any circumstances, but in certain idealized circumstances, in which one scrupulously 'checks' one's work, and in which one's mental capacities have somehow been enhanced to allow one to consider arbitrarily large numbers. Kripke's criticism of this response is telling. In order for the analysis to be non-circular, one must not characterize the idealized dispositions in a way that presupposes in advance what we mean by '+.' For example, they cannot be characterized as dispositions to answer the relevant questions when we have been provided with sufficient mental capacities to allow us to correctly *add* any two numbers. However, if our characterization of the ideal dispositions is non-circular, then we will have little reason to be confident that they will in fact determine the desired function. It is also worth noting that the finitude and error objections cannot be overcome by appealing to the dispositions of the entire linguistic community of which the speaker is a part. These

role — perhaps along with other factors — in determining what the term means. With this in mind, we turn to Kripke's normativity objection. Applied to the dispositional analysis, it is designed to show how far from providing an answer to the skeptic the analysis really is. Applied more generally, the normativity objection is seen as the most fundamental obstacle to any analysis that attempts to answer the skeptical challenge.

The objections to the dispositional analysis based on error and finitude show that my dispositions to calculate with '+' do not correspond perfectly to the addition function. The normativity objection maintains that even if my dispositions did correspond perfectly to the addition function, that would not be enough to show that I meant addition by '+.' The mere fact that I am now, and have been in the past, disposed to answer a question in a certain way does not show that I should answer it in that way. But it is a crucial aspect of meaning something by a term that one has adopted a standard to which one tries to adhere, and which provides a basis for judging responses to be correct or incorrect. Kripke contends that the dispositional analysis doesn't provide such a standard.

According to Kripke, the main reason that it doesn't is that it fails to explain how I can be *justified* in giving one answer rather than another when presented with a previously unconsidered calculation.[10] The

dispositions are just as subject to the objections as are the dispositions of an individual.

These considerations constitute serious obstacles to the dispositional analysis. I do not maintain that they rule out all possible reformulations. Perhaps there are reformulations of the analysis that are capable of avoiding the error and finitude objections. Even if this is so, the normativity arguments will remain, and will, I suspect, require appeal to some factors over and above dispositions. See below.

10 A relevant passage is the following:

To a good extent this reply [the dispositional analysis] immediately ought to appear to be misdirected, off target. For the skeptic created an air of puzzlement as to my *justification* for responding '125' rather than '5' to the addition problem as queried. He thinks my response is no better than a stab in the dark. Does the suggested reply advance matters? How does it *justify* my choice of '125'? What it says is: "'125' is the response you are

picture that emerges from his discussion is something like this: The skeptic gives me a new calculation that I have never done before. He asks me *What is 68 + 57?* I answer that 68 + 57 = 125. The skeptic asks me to cite some past fact that justifies this answer, in the sense of showing that it is the answer I ought to give provided that I am now using '+' with the same meaning as I did in the past. In effect, the skeptic is imposing the requirement that no fact will suffice to answer his challenge unless we can conclude from that fact that I ought to give '125' as answer to the question *What is 68+57?*, provided that I now intend to use '+' with the same meaning as I did before.

Let me formulate this requirement as follows:

(N$_E$) If the fact that *F* determined that (in the past) one meant addition by '+,' then knowing that *F* would, in principle, provide one with a sufficient basis for concluding that one ought to give

---

disposed to give, and (perhaps the reply adds) it would also have been your response in the past." Well and good, I know that '125' is the response I am disposed to give (I am actually giving it!), and maybe it is helpful to be told — as a matter of brute fact — that I would have given the same response in the past. How does any of this indicate that — now *or* in the past — '125' was an answer *justified* in terms of instructions I gave myself, rather than a mere jack-in-the-box unjustified and arbitrary response? (23)

The emphasis here on justification echoes the way in which Kripke initially sets up the skeptical problem.

In the discussion below, the challenge posed by the skeptic takes two forms. First, he questions whether there is a *fact* that I meant plus, not quus, that will answer his skeptical challenge. Second, he questions whether I have any reason to be so confident that now I should answer '125' rather than '5.' The two forms of the challenge are related. I am confident that I should answer '125' because I am confident that this answer also accords with what I *meant*. Neither the accuracy of my computation nor of my memory is under dispute. So it ought to be agreed that if I meant plus, then unless I wish to change my usage, I am justified in answering (indeed compelled to answer) '125' not '5.' An answer to the skeptic must satisfy two conditions. First, it must give an account of what fact it is (about my mental state) that constitutes my meaning plus, not quus. But further, there is a condition that any putative candidate for such a fact must satisfy. It must, in some sense, show how I am justified in giving the answer '125' to '68 + 57.' (11)

the answer '125' to the question *What is 68 + 57?*, provided one intends to use '+' with the same meaning it had in the past.

It is important to understand what must be presupposed in order for this requirement to make sense. The reason one ought to answer '125' is that, if one did mean addition by '+,' then '125' is the *correct* answer to the question *What is 68+57?*; and the correct answer is the answer one ought to give. The reasoning here is as follows: If knowing that *F* allows us to conclude that in the past one meant addition by '+,' then it allows us to conclude that the sentence *68 + 57 = 125* means that the *sum* of 68 and 57 is 125, and hence is true iff the sum of 68 and 57 is 125, provided, of course, that one is now using '+' with the same meaning as before.[11] Next we appeal to the arithmetical fact that 125 *is* the sum of 68 and 57, which the skeptic grants, and is not in dispute.[12] From these two facts, one about meaning and one about numbers, it follows that the sentence *68 + 57 = 125* is true, and hence that one speaks the truth if one responds '125' to the question *What is 68+57?* Since the skeptic's requirement clearly presupposes that one ought to speak the truth in this case, it follows that one ought to answer '125.' In this way knowledge of the putative meaning-determining fact that *F*, together with undisputed knowledge of nonlinguistic facts and the general presumption that one ought to speak the truth, provide a basis for concluding what answer one ought to give to the question *What is 68+57?*

---

11 I am here taking the meanings of the other symbols — the numerals and identity sign as given. This simplification does not affect the point at hand.

12 On page 11 Kripke insists *"Neither the accuracy of my computation nor of my memory is under dispute. So it ought to be agreed that if I meant plus, then unless I wish to change my usage, I am justified in answering ... '125,' and not '5.'"* Similarly, on page 13, in explaining why he sets up the paradox as a metalinguistic problem about the past use of an expression, Kripke says that when this is not done, *"some listeners hear it as a skeptical problem about arithmetic: "How do I know that 68 + 57 is 125?" (Why not answer this question with a mathematical proof?) At least at this stage, skepticism about arithmetic should not be taken to be in question: we may assume, if we wish, that 68 + 57 is 125."* I take these passages to indicate that in Kripke's formulation of the paradox, my present (true) belief that 68 + 57 = 125 can be appealed to (without further justification) in attempting to meet the skeptic's demand that I show how I am justified in thinking that '125' is the answer required by my past understanding of the term.

Applying the normativity requirement (N_E) to the dispositional view yields the first premise, (P1), of an obvious normativity argument.

(P1)   If one's past dispositions to answer questions 'What is $\underline{n} + \underline{m}$?' determined that one meant addition by '+,' then knowing one's past dispositions to answer such questions would, in principle, provide one with a sufficient basis for concluding that one ought to give the answer '125' to the question *What is 68 + 57?*, provided one intends to use '+' with the same meaning it had in the past.

The normativity objection to the dispositional analysis consists in (P2).

(P2)   Knowing one's past dispositions to answer questions 'What is $\underline{n} + \underline{m}$?' does not, in principle, provide one with a sufficient basis for concluding what answer one ought to give to any such question, provided one intends no change in meaning.

Together, (P1) and (P2) entail that no matter how closely my past dispositions to answer questions of the relevant form may, in fact, have corresponded to the results of applying the addition function, those dispositions did not determine that in the past I meant addition by '+.' Thus, if (P1) and (P2) are correct, then the dispositional response to the skeptic is incorrect.

Much of Kripke's discussion is aimed at supporting (P2), which is itself quite plausible. Suppose I had amnesia and didn't remember what I meant by '+' in the past. When asked, at present, to calculate '68 + 57' I might feel inclined to give the answer '125.' Still, I might wonder whether this answer was justified, in terms of what I meant by '+' in the past. Would it be sufficient to be told that this was the answer I was disposed to give in the past? It doesn't seem that it would, for I might wonder whether in the past this was one of the calculations on which I was disposed to make mistakes. A similar observation holds regarding the point of view of someone viewing me from the outside and wondering whether my present answer of '125' is correct. Simply being given a specification of the relevant past dispositions is not enough to allow such an observer to conclude that my present answer accords with my past meaning or intentions.

Since (P2) is true, and since (P1) and (P2) entail that the dispositional response to the skeptic is false, the success of the argument against the dispositional view depends on our evaluation of the Normativity Requirement (N_E), and the particular instance of it given in (P1). How plausible are these principles? Quite plausible, it seems, provided that we understand what it is for one empirical truth to determine another in accord with the following constraint.

(Det$_E$) P determines Q only if, given P, one can demonstrate Q without appealing to any other empirical facts — i.e. only if Q is an apriori consequence of P.

With this understanding of determination, (N_E) and (P1) will hold.[13] If the claim that in the past one meant addition by '+' were an apriori

13 Is it necessary, in order for the Normativity Requirement (N_E) to play its proper role in the skeptic's argument, for the meaning-determining fact that F to provide the basis for an apriori, demonstrative inference to the relevant conclusion, or would it be enough for the fact that F to provide the basis for any sort of inference to the conclusion? Although the point is arguable, there are reasons for opting for the stronger interpretation. First, Kripke's skeptic seems to be looking for nonintentional facts that both metaphysically necessitate and epistemologically demonstrate the relevant meaning facts. Nonintentional facts that have meaning facts as apriori consequences would, presumably, do this; whereas nonintentional facts which provide only the basis for an empirical inference to the meaning facts presumably would not. Second, when we cast our net wider, and include nonintentional facts other than the dispositional facts presently under consideration, the skeptic's claim that meaning facts are not apriori consequences of any such set of facts remains plausible, whereas a corresponding claim, to the effect that meaning facts cannot be inferred from nonintentional facts, even by an empirical inference to the best explanation, say, is not nearly so plausible.

A different question about the normativity requirement is whether, as it is currently stated, it is too weak. Perhaps, in addition to being required to have meaning facts as apriori consequences, meaning-determining facts should also be required to be readily accessible to the speaker — so that the speaker is in a position to draw the relevant conclusions from those facts. I have no objection to this strengthening of the requirement; all the philosophical points I want to make regarding (N_E) would hold for this strengthening of (N_E). My reason for preferring the weaker version in the text is that it leads to a conclusion of greater generality. I am indebted to George Wilson for a discussion of this point.

consequence of the relevant dispositional truths, then knowing those dispositional truths would, in principle, make it possible to conclude that in the past one meant addition by '+.' But then, since it is accepted as uncontroversial that 125 *is* the result of adding 68 and 57, one could conclude that '125' is the correct answer to the question *What is 68 + 57?*, and hence is the answer that one ought to give, provided one intends to use '+' with the same meaning as it had in the past.

It is worth emphasizing that on this account the substantive constraint placed on the meaning-determining fact F by the normativity requirement (N_E) is that, in principle, knowledge of F must allow one to demonstrate that in the past '+' was used to mean addition. This is essentially a descriptive matter. The normative prescription about what answer one ought to give in a particular case results from (i) the descriptive fact about what one meant, (ii) the arithmetical facts involving the addition function, and (iii) the normative presumption that one should speak the truth. The substantive requirement on F is just that knowledge of it be sufficient to demonstrate the first of these.[14]

The advantage of this way of construing the normativity objection is that the resulting argument against the dispositional analysis turns out to be sound. However, there is a corresponding disadvantage — the conclusion thereby established is weak. Suppose that the proposition that one meant addition by '+' really isn't an *apriori consequence* of true propositions about one's past dispositions to calculate. For all we know it might still be a *necessary consequence* of those propositions. And if it is, then there is an important sense in which what one means is determined by facts about one's dispositions. Despite all that has been shown up to now, facts about meaning may yet turn out to be meta-

---

14 The point here is a general one, and applies to versions of the rule-following paradox involving all sorts of different words. For example, consider a word like 'fossil,' which applies to concrete physical objects. Given some object *o*, the normative prescription regarding whether one should apply 'fossil' to *o* depends on (i) the meaning of the word 'fossil,' (ii) the nonlinguistic facts about the nature of *o* — i.e. whether or not it is a fossil and (iii) the general presumption that one should apply a term to an object only if doing so would involve speaking the truth. As Michael Thau has emphasized to me, examples like this highlight the need for (ii) over and above (i) particularly clearly.

There is, however, another issue involving facts like (ii) that is worth mentioning, even though I don't have sufficient space here to go into detail. Suppose I apply the word 'fossil' to a certain object *o*, and then I am challenged by a skeptic to justify this new application of the word. Suppose further that part of my response involves citing some fact F about my past use of the word. Imagine for the sake of argument that knowledge of F does allow me to demonstrate that in the past I used the word to mean fossil. Still, this by itself doesn't justify my application of the word to *o*, even if I am now using the word with the same meaning as before, and it is granted that I ought in this case to speak the truth. To complete my justification I have to show that I am justified in thinking that *o* is a fossil. And how am I to do that? Perhaps the explanation of what it is for me to think that *o* is a fossil is simply for me to understand the

word 'fossil' and be disposed to apply it to *o* on the basis of some reasonable, empirical examination of, or inquiry about, *o*. If so, then the justification for my thinking that *o* is a fossil may amount to little more than my now understanding the word 'fossil' and being disposed, after appropriate investigation and reflection, to apply it to *o*.

If this is the situation, then in some ultimate sense I may have no justification for applying the term to a new item *o*, other than my own confident, informed, and reflective inclination to do so, plus my status as a competent user of the term. This does seem to be an important part of what Wittgenstein was trying to show in his deployment of the rule-following paradox, and it is also present in Kripke's discussion. Moreover, there are, I think, cases in which something like this point is correct. However, this observation about justification does not undermine the response given in the text to Kripke's version of the skeptical paradox; nor does it lead to any defensible skepticism about meaning.

What it may do is point to an alternative route to one of Wittgenstein's conclusions about language — a route that does not require any fundamental skeptical recasting of our ordinary conception of meaning. The conclusion is that, at least in some cases, the explanation of a speaker's understanding of a term *T* (including his ability to apply it to newly considered objects) does not involve associating *T* with an independently apprehended property *P* (and judging those objects to have *P*). Understanding a term, or using it to mean a certain thing, is not always like deciding to attach a new proper name to an object with which one is already familiar. See my 'Semantic Competence,' *Philosophical Perspectives* 3 (1989) 575-96, at 587-91 for a brief discussion of this idea. This idea is discussed in more detail in my, 'Facts, Truth Conditions, and the Skeptical Solution to the Rule-Following Paradox,' *Philosophical Perspectives* 12 (1998), forthcoming.

I am indebted to James Pryor and Michael Thau for helpful discussions of the material in the footnote.

physically necessary consequences of facts about dispositions, even if they are not apriori consequences of such facts.

However, this is not all that Kripke has to say about the normativity objection to the dispositional analysis, as is indicated by his discussion of the following example.

> Assuming determinism, even if I mean to denote *no* number theoretic function in particular by the sign '*,' to the same extent as it is true for '+,' it is true here that for any two arguments *m* and *n*, there is a uniquely determined answer *p* that I would give (I choose one at random, as we would normally say, but causally the answer is determined). The difference between this case and the case of the '+' function is that in the former case, but not in the latter, my uniformly determined answer can properly be called 'right' or 'wrong.' (p. 24)[15]

Kripke's idea can be made more graphic by imagining that as a result of some special brain surgery, I might be programmed to verbally respond to an utterance of the interrogative sentence *What is x * y?* by uttering a definite numeral z for each pair of numerals x and y used in the question. For any pair of numerals you give me, using '*,' I will respond by uttering a definite numeral. Because of this we may presume that some definite function f on the natural numbers corresponds to the verbal responses I am disposed to give to interrogatives containing '*.' Despite this, it might still be that I don't *mean* anything by '*.' My responses may be nothing more than verbal reflexes. I might notice that I always give determinate responses to queries using '*.' But I don't regard these as correct or incorrect — since I don't intend to assign any meaning to '*.'

The point here is that, by themselves, mere verbal dispositions to respond to interrogative sentences involving a symbol, either '+' or '*,' are not enough to endow it with meaning — even if the dispositions can be seen as corresponding to a definite function. Since it is possible to have these dispositions without meaning the function, meaning the function is not a necessary consequence of having the specified dispositions. Some other element must be present in order for us to have genuine meaning. So far, we haven't found it.

---

15 See also Kripke's discussion in footnote 18, 24.

It should be noticed that in giving this argument, we have quietly replaced Kripke's original formulation (D) of the dispositional analysis with some nonintentional understanding of it. In the original formulation, the allegedly meaning-determining dispositions were characterized as dispositions to *answer* arithmetical questions, presumably by *asserting* that a particular number is the value of the relevant function at a given pair of arguments. By contrast, in Kripke's '*' example I don't mean anything by '*'; I do not *assert* anything or *answer* any question. Here the dispositions — both with '*' and with '+' — are characterized simply as dispositions to utter specific numerals when prompted by utterances of particular interrogative sentences containing '*' or '+.'

When dispositions are characterized in this way, Kripke's discussion of the '*' example can be seen as constituting a second normativity argument against the dispositional analysis. Like the first, it is based on a general requirement, in this case (N_M).

(N_M) If the fact that F determined that (in the past) one meant addition by '+,' then in any possible world in which it is the case that F, one means addition by '+'; hence, in any possible world w in which it is the case that F, '125' is the answer one ought to give to the question *What is 68 + 57?*.

Applied to the nonintentional version of the dispositional analysis, this general requirement yields (P1').

(P1') If a certain complex dispositional fact — namely the fact that in the past one was disposed to verbally respond in such and such ways to utterances of interrogative sentences 'What is $n + m$?' — determines that (in the past) one meant addition by '+,' then in any possible world w in which that dispositional fact obtains one means addition by '+' in w, and hence ought to answer '125' to the question *What is 68 + 57?*

The second normativity objection to the dispositional analysis (formulated nonintentionally) is (P2').

(P2') It is possible *not* to mean addition (or anything at all) by '+' even though one is disposed to verbally respond in such and such ways (the ways mentioned in (P1')) to utterances of interrogative sentences 'What is $\underline{n} + \underline{m}$?'

From (P1') and (P2') it follows that the fact that in the past one was disposed to verbally respond in the ways mentioned in (P1') to utterances of interrogative sentences 'What is $\underline{n} + \underline{m}$?' does *not* determine that (in the past) one meant addition by '+.'

The argument is clearly valid. Moreover, the discussion of the '*' example shows that (P2') is true, provided that the dispositions spoken of are nonintentionally stated and restricted simply to the production of numerals in response to utterances of interrogative sentences. (P1') is also true, provided that the determination relation it speaks of is constrained by the following principle.

(Det$_M$) *P* determines *Q* only if *Q* is a necessary consequence of *P*

We have, then, two versions of the normativity objection. One version, the necessary-consequence version, incorporates (Det$_M$); the other version, the apriori-consequence version, incorporates the earlier requirement, (Det$_E$), on the determination relation. Both versions are sound, and establish their conclusions, when applied to the nonintentional form of the dispositional analysis. However, recognition of the difference between the two gives rise to an important question. If, in addition to the verbal dispositions already considered, we cast our net wider, so as to include more and more nonintentional facts, will the two versions of the normativity objection continue to apply together, or will they come apart at some point?

In order to investigate this question, we let *T* be the set of nonintentional truths about my dispositions to produce numerals in response to questions 'What is $\underline{n} + \underline{m}$?'. In our discussion of the '*' example, we have seen that it is possible for all members of *T* to be true even though I don't mean anything by '+,' because my responses are not real answers but only verbal reflexes. Are there other nonintentional truths which, together with *T*, would change this result? Suppose we formed a new set, *T'*, which included all members of *T* together with all additional nonintentional truths about my dispositions to verbal

behaviour involving '+' — for example, dispositions covering cases in which I 'check and revise' my work, dispositions to insist on one and only one 'answer' for any given question, dispositions to strive for agreement between my own answers and those of others, and so on. Would it be possible for all members of this larger set *T'* to be true, even though I didn't mean anything by '+'?

I should emphasize that I am not here trying to identify all the relevant verbal dispositions; nor am I trying to characterize them precisely in a thoroughly nonintentional way. Rather I am assuming that some such characterization is possible, and I am asking whether, by taking a broad enough look at such a characterization of our linguistic behaviour and dispositions, we would at some point reach a stage in which we were confident that anyone conforming to all those facts meant something by his words? Is there a possible world in which all one's nonintentionally characterized dispositions to verbal behaviour, including those just mentioned, match my dispositions in the actual world, yet one does not mean addition, or anything else, by '+'? If so, how is this possible? Would the result change if we enlarged the set of potential meaning-determining truths still further to include not only all truths about my dispositions to verbal behaviour, but also all truths about (i) the internal physical states of my brain, (ii) my causal and historical relationships to things in my environment, (iii) my (nonintentionally characterized) interactions with other members of my linguistic community, (iv) their dispositions to verbal behaviour, and so on? Is there a possible world in which someone conforms to all those facts — precisely the facts that characterize me in the actual world — and yet that person does not mean anything by '+'?

I think not. Given my conviction that in the past I did mean addition by '+,' and given also my conviction that if there are intentional facts, then they don't float free of everything else, I am confident that there is no such a world. Although I cannot identify the smallest set of nonintentional facts about me in the actual world on which meaning facts supervene, I am confident that they do supervene. Why shouldn't I be? I start out convinced that in the past I meant addition by '+.' Next, I imagine some large set of nonintentional facts about the past remaining fixed, so that they obtain in some possible world w. I then ask whether the individual who corresponds to me in w — the individual who is the focus of the fixed nonintentional facts in w — means some-

thing by '+.' The answer, of course, is that since I meant something in the actual world, he means something in *w*. And if he means something by '+,' surely he means what I meant by '+' in the actual world — addition.

Of course, this won't convince the skeptic. He isn't willing to grant that I meant anything in the past. He wants to be shown that there is a nonempty set of meaning facts about which the question of supervenience can be raised; and he won't be satisfied until I specify a set of nonintentional truths from which the existence of such meaning facts can be demonstrated. In effect, he demands that I establish that purported truths about what I meant are apriori consequences of nonintentional truths the existence of which we both accept.

But why should I acquiesce in this demand? It would be very interesting if it turned out that I could refute the skeptic using only uncontroversial premises that even he accepts. However this is not generally required of responses to skeptical problems, and I see no special reason to require it in this case. What is required is that I rebut any skeptical challenge that purports to show that our ordinary, pretheoretic beliefs about meaning conflict with other fundamental beliefs in a way that makes our beliefs about meaning unsustainable. In general, the power of skeptical views lies not in their rejection of what we commonly take for granted, but in their finding reasons for such rejection that arise directly from other fundamental convictions of ours. If the skeptic can find such reasons, he can show that our own most deeply held views are sufficient to undermine the beliefs (about meaning) that he challenges.

Has Kripke's skeptic about meaning done this? I don't think so. He has insisted that if I meant anything in the past, then what I meant must be determined by nonintentional facts; and I have agreed, provided that the determination relation is one of necessary consequence. I grant that if I meant anything in the past, then what I meant must be a necessary consequence of nonintentional truths about me, my environment, my community, and so on. But it is not evident that there is a problem here, since none of the skeptic's arguments show that such a relation fails to hold. Indeed, they scarcely even attempt to show this. Instead, they try to establish that no collection of nonintentional truths will allow us to *demonstrate* the truth of the relevant intentional claims. This, I have suggested, is tantamount to an attempt to convince us that

claims about meaning (and propositional attitudes) are not apriori consequences of any set of nonintentional truths.

On this point it must be admitted that the skeptic has a strong case. The task of deriving claims about meanings (and propositional attitudes) from nonintentional truths, together perhaps with apriori definitions of intentional notions like meaning and belief, is daunting. I don't know how to give such a derivation, and I am not sure that any is possible.[16] Consequently, I am willing to grant that the skeptic might be right in maintaining that claims about what I meant are not apriori consequences of nonintentional truths.[17]

If it were clear that any necessary consequence of a set of claims *P* was also an apriori consequence of *P*, then this admission would provide the skeptic with just what he needs; for then he could force me to admit that claims about meaning may not be necessary consequences of nonintentional truths. That would conflict with my conviction that meaning facts must supervene on nonintentional facts, and so would threaten my pretheoretic commitment to meaning facts. However, this argumentative strategy fails. Thanks to the work of Kripke and others, it has become clear that many necessary consequences of propositions are not apriori consequences of them. Consequently, my admission that claims about meanings may not be apriori consequences of nonintentional truths need not undermine my belief that they are necessary consequences of those truths.

Could the skeptic undermine my claims about past meanings by appealing directly to the thesis that if there are facts about what I meant

---

16 One important reason why it is difficult to be certain on this point is that it is not completely clear what apriori definitions of intentional notions are possible. For a good discussion of this issue, and an argument for the conclusion that facts about meaning (and belief) are not apriori consequences of nonintentional facts, see Alex Byrne, *The Emergent Mind*, unpublished dissertation, Princeton University, 1993.

17 If the required epistemological relationship between the nonintentional truths and the claims about meaning were weakened to include empirical inferences to the best explanation, then the skeptical thesis that meaning claims are not epistemological consequences of nonintentional truths would be far more questionable. In any case, Kripke's skeptic does not argue in this way. (See note 13 above.)

in the past, then they must be apriori consequences of nonintentional truths? I don't think so. That thesis is not one of my beliefs about meaning, pre-theoretic or otherwise. Since the skeptic has done nothing to establish it in its own right, his challenge fails.

To sum up, the general form of the skeptic's argument is the following:

P_1 If in the past there was a fact about what I meant by '+' , in particular, if there was a fact that I meant addition by '+,' then either:

    (i) this fact was determined by nonintentional facts of such and such kinds — facts about my past calculations using '+,' the rules or algorithms I followed in doing calculations involving '+,' my past dispositions to respond to questions 'What is $\underline{n}$ + $\underline{m}$?', the totality of my past dispositions to verbal behaviour involving '+,' etc.

    or

    (ii) the fact that I meant addition by '+' was a primitive fact, not determined by nonintentional facts.

P_2 Nonintentional facts of type (i) did not determine that I meant addition (or anything else) by '+.'

P_3 What I meant by '+' was not a primitive fact.

C_1 Thus, in the past there was no fact that I meant addition (or anything else) by '+.'

C_2 By parity of reasoning, there never was a fact about what I, or anyone else, meant by any word; *ditto* for the present.

In my view, the argument suffers from equivocation. If the determination relation is one of necessary consequence, then P_3 is plausible, but P_2 is not.[18] If the determination relation is one of apriori consequence, then P_2 is plausible, but P_3 is not. What makes the argument seductive is the failure to distinguish these alternatives. Once this distinction is made, the argument loses much of its force. There is no in-

---

18 Given an appropriately broad listing of nonintentional facts in P_1.

terpretation in which P_1, P_2, and P_3 are jointly true; nor do I know of one in which they are even jointly plausible.

Quine's Indeterminacy Thesis

I will now try to show that Quine's arguments for the indeterminacy of translation (and inscrutability of reference) suffer from a similar defect. We may begin by noting the relationship between Kripke's skeptical problem and Quine's. Kripke sets up the challenge as that of specifying nonintentional facts which determine that what I mean now by '+' is the same as what I meant by '+' in the past. This can be rephrased as the challenge of specifying nonintentional facts which determine that a translation from '+' as I used it in the past to '+' as I use it now is correct, whereas a translation from '+' as I used it in the past to, say, 'quus' as I use it now (to stand for the quaddition function defined in note 5) is incorrect. Quine's question is a generalization of this challenge: What nonintentional facts determine that it is correct to translate an expression A, as used by a person p, or a community c, as meaning the same as an expression B, as used by a person p', or community c'? Like Kripke's skeptic, Quine answers that in a wide range of cases there are no meaning-determining facts.

Quine's position can be set out as a series of theses. The first of these is The Underdetermination of Translation by Data, which is an instance of his view that empirical theories of all sorts typically are underdetermined by observational evidence.

> *The Underdetermination of Translation by Data*
> Let L_1 and L_2 be arbitrary languages, and let D be the set of all observational truths (known and unknown) relevant to translation from one to the other. For any theory of translation T for L_1 and L_2, compatible with D, there is a theory T', incompatible with T, that is equally well supported by D.[19]

---

19 The strength of this thesis, as well as the more general thesis of the underdetermination of empirical theories by observational data, depends on one's conception of what it is for a class of data statements to support a theory.

This thesis can be fleshed out by saying a little about what theories of translation are, and what Quine takes the evidence for them to be. First consider theories of translation. A theory of translation for two languages correlates individual words of each language with words or phrases of the other language; this correlation is used in the theory to correlate the sentences of the two languages.[20] Any system of establishing such correlations can be counted as a translation manual, or theory. We may take such a theory as yielding (infinitely many) theorems of the form:

Word or phrase w in L_1 means the same as word or phrase w^* in L_2.

Sentence S in L_1 means the same as sentence S^* in L_2.

According to Quine, the empirical data relevant to theories of translation are statements about the stimulus meanings of sentences. The stimulus meaning of a sentence S (for a speaker at a time) is a pair of classes — the class of situations in which the speaker would assent to S, and the class in which the speaker would dissent from S. Stimulus meanings are particularly important in evaluating translations of what Quine calls 'occasion sentences' and 'observation sentences.' An occasion sentence is one assent to, or dissent from, which depends on the

speaker's current observation. An observation sentence is an occasion sentence for which assent or dissent depends only on observation — with no, or only a minimum in the way of, background information required. For example, 'He is a bachelor' is an occasion sentence, since assent or dissent in a given case depends on whom the subject is observing; but it is not an observation sentence in Quine's sense because whether or not one assents in a particular case depends in part on having special background information about the person perceived. 'That is red,' on the other hand, does not depend (to the same extent) on having such background information, and so counts as an observation sentence for Quine. Quine's way of approximating this intuitive notion of an observation sentence within his behaviouristic framework is to define an observation sentence in a language L to be an occasion sentence for speakers of L the stimulus meaning of which varies little from one speaker to another.

The stimulus meanings of observation and occasion sentences play a prominent role in Quine's conception of the main observational predictions made by translation theories. These predictions are summarized by the following three principles for extracting testable claims from theories of translation (which otherwise wouldn't entail any such predictions via their form alone):[21] (i) if a translation theory states that an observation sentence S in L_1 means the same as a sentence S^* in L_2, then the theory predicts that the two sentences have the same stimulus meanings in their respective linguistic communities; (ii) if S and S' are occasion sentences of L_1, and if a translation theory states both that S means the same in L_1 as S^* in L_2 and that S' means the same in L_1 as $S^{*'}$ in L_2, then the theory predicts that S and S' have the same stimulus meaning in L_1 iff S^* and $S^{*'}$ have the same stimulus meaning in L_2; (iii) if a theory translates an expression n of a language L as meaning the same as 'not' in English, then adding n to sentences of L must reverse stimulus meaning; similar claims are made regarding other truth functional operators. On this conception, the observational data for theories of translation consist mainly of behavioural evidence regarding the stimulus meanings of occasion sentences.

For present purposes I will follow what appears to be Quine's latitudinarianism on this subject. Theories, together with auxiliary observational statements, make (entail) observational predictions. (Which statements count as observational for this purpose will not be an issue here.) A set of such observational predictions supports a theory to the extent that the theory, supplemented by true auxiliary observation statements, entails the members of the set. Two theories (appropriately supplemented with auxiliary observational statements) that entail the same members of the set, are equally well supported by the set.

20 I am assuming, in order to simplify the argument, that words are the minimal meaning bearing units, that languages contain finitely many such words, and that the translation of the infinitely many phrases and sentences of the two languages is the result of (i) the translation of the words that make them up plus (ii) combinatorial principles specifying the translations of syntactically complex expressions in terms of the translations of their parts.

21 See chapter 2 of *Word and Object* (Cambridge, MA: MIT Press 1960).

Quine also considers a possible constraint relevant to the translation of what he calls standing sentences — sentences assent to, or dissent from, which is independent of current sensory stimulation. The possible constraint is that sentences assented to (dissented from) in every situation by the community of speakers of L_1 must be translated onto those assented to (dissented from) in every situation by the community of speakers of L_2. But this constraint is itself problematic, and, in any case, adding it would not significantly change the overall picture. It is clear that if the observational data for theories of translation are restricted to behavioural evidence of the sort Quine has in mind, then profoundly different theories of translation will be supported equally well (in his sense) by all observational data, known and unknown, in virtually all interesting cases.

For example, the set of all behavioural data concerning the stimulus meanings of sentences for me, past and present, is equally compatible with theories of translation which claim that (i) the term 'rabbit' as used by me in the past means the same as the term 'rabbit' as used by me now, (ii) the term 'rabbit' as used by me in the past means the same as the phrase 'set of undetached rabbit parts' as used by me now, or (iii) the term 'rabbit' as used by me in the past means the same as the phrase 'temporal stage of a rabbit' as used by me now. Since the expressions 'rabbit,' 'set of undetached rabbit parts,' and 'temporal stage of a rabbit,' as used by me now mean, and refer to, different things, alternative translation theories that map these different expressions onto the term 'rabbit' as I used it in the past conflict with one another, and can be regarded as incompatible. Quine concludes from this that theories of translation are underdetermined, in the sense defined above, by the observational data for them.[22]

Since empirical theories of all sorts are typically underdetermined by the observational evidence bearing on them, the extension of this result to theories of translation is not, in itself, a radical development. What makes it striking is Quine's use of it as the basis for the truly radical doctrine of the Indeterminacy of Translation.

The Indeterminacy of Translation
Translation is not determined by the set N of all truths of nature, known and unknown. For any pair of languages, there are incompatible theories of translation for those languages that accord equally well with N. All such theories are equally true to the facts; there is no objective matter of fact on which they disagree, and no objective sense in which one is true and the other is not.[23]

---

22 When Quine speaks of different, *incompatible* theories (or 'theory formulations') all equally supported by the same possible observational evidence, he seems to have in mind *logically incompatible* theories (or theory formulations). (See 'On the Reasons for Indeterminacy of Translation,' *Journal of Philosophy* 67 (1970) 178-83, at 179; and 'On Empirically Equivalent Systems of the World,' *Erkenntnis* 9 (1975) 313-28, at 322.) However, despite the obvious difference in meanings between 'rabbit,' 'set of undetached rabbit parts,' and 'temporal stage of a rabbit' as I use them now, the following claims are not *logically* incompatible: (i) the term 'rabbit' as I used it in the past means the same as the term 'rabbit' as I use

it now (ii) the term 'rabbit' as I used it in the past means the same as the phrase 'set of undetached rabbit parts' as I use it now, (iii) the term 'rabbit' as I used it in the past means the same as the phrase 'temporal stage of a rabbit' as I use it now. Consequently, translation theories making these different claims need not be *logically* incompatible with one another.

Logical incompatibility will result if translation theories are embedded in larger background theories containing the following claims: (a) Rabbits are not sets of undetached rabbit parts & sets of undetached rabbit parts are not temporal stages of rabbits & rabbits are not temporal stages of rabbits ; (b) 'Rabbit' (as I use it now) refers to an object iff it is a rabbit & 'set of undetached rabbit parts' (as I use it now) refers to an object iff it is a set of undetached rabbit parts & 'temporal stage of a rabbit' (as I use it now) refers to an object iff it is a temporal stage of a rabbit; (c) if two words refer to different things then they don't mean the same; (d) if a means the same as b & a means the same as c, then b means the same as c. Let T_1 be a translation theory containing statement (i), T_2 be a translation theory containing statement (ii), and T_3 be a translation theory containing statement (iii). The union of T_1, T_2, and a set containing (a)-(d) is logically inconsistent; as are corresponding unions involving the other relevant combinations. The justification for appealing to these auxiliary claims is that (a) states an obvious fact and (b)-(d) are axiomatic to any overall theory that makes significant use of the concepts of meaning and reference.

23 See Quine, 'Reply to Chomsky,' D. Davidson and J. Hintikka, eds., *Words and Objections* (Dordrecht: Reidel 1969), 303.

This doctrine is a consequence of two more basic Quinean views.

Physicalism
All genuine truths (facts) are determined by physical truths (facts).

The Underdetermination of Translation by Physics
Translation is not determined by the set of all physical truths (facts), known and unknown. For any pair of languages, there are incompatible theories of translation for those languages that accord equally well with all physical truths (facts).

Quine's route to the Indeterminacy of Translation is as follows: He begins with the behaviourist premise that since we learn language by observing the linguistic behaviour of others, the only facts relevant to determining linguistic meaning must be publicly observable behavioural facts — in particular facts about stimulus meaning. But his discussion of the Underdetermination of Translation by Data shows that these facts do not determine which translations of our words are correct, and so do not determine what our words mean. It follows that *no* physical facts determine word meaning or correct translation. In the presence of the doctrine of Physicalism this means that claims about what our words mean — e.g., claims like *'rabbit' as I used it in the past means the same as 'rabbit' as I use it now* — never state genuine facts, and never count as expressing genuine truths. Hence, the Indeterminacy of Translation.

How persuasive is this argument? Many object, quite correctly, to its behaviourist premise, which allows Quine to move from the Underdetermination of Translation by Data to the Underdetermination of Translation by Physics.[24] However, there is another, more fundamental, point to be made. The contents of Quine's central claims are unclear.

---

24 In this paper I will assume, without argument, that Quine's behaviourism about language is false. It is noteworthy, however, that the appeal of the indeterminacy thesis, and the challenge posed by it, have been strongly felt by many philosophers who have not been prepared to accept behaviourism independently. My task here is to diagnose the source of that appeal, and defuse the challenge felt by those philosophers.

In particular, it is unclear precisely what determination relation is invoked in the three central theses of Physicalism, the Underdetermination of Translation by Physics, and the Indeterminacy of Translation. Because of this, equivocation threatens. On certain construals of the determination relation Physicalism is plausible; on other construals the Underdetermination of Translation by Physics is plausible; but there is no construal on which both Physicalism and the Underdetermination of Translation by Physics are plausible, and no interpretation in which the Indeterminacy of Translation can be sustained.

What is it for one set of claims to determine another? It is not for the claims in the second set to be logical consequences of the claims in the first. Certainly translation theories are not logical consequences of the set of all physical truths. But this is trivial, since whenever an empirical theory of any interest includes vocabulary not found in the truths of physics it will fail to be a logical consequence of those truths. For example, not all the truths of chemistry and biology are logical consequences of the set of true sentences of the language of an ideal physics. But chemistry and biology are supposed by Quine to be determined by physics; so the determination relation cannot be that of logical consequence. If it were, Physicalism would obviously be false.

A different way of specifying the determination relation would be to say that a set P of statements determines a set Q of statements iff it would be (metaphysically) impossible for all the statements in P to be true without all the statements in Q being true — i.e. iff Q is a (metaphysically) necessary consequence of P. On this construal physicalism is quite plausible; it states that all genuine truths (facts) supervene on the physical truths (facts). However, the Underdetermination of Translation by Physics now turns out to be implausible. The point here parallels the one made in response to Kripke's skeptical argument about meaning. Prior to any skeptical argument, Kripke's or Quine's, we naturally assume at the outset that there are facts about meaning and translation of the sort the skeptic denies. Given this conviction, we may ask directly whether a physically identical twin — someone (in a physically identical possible world) whose utterances, behaviour, brain states, causal and historical relations to the environment, and interactions with other speakers completely and exactly matches mine (in the actual world) — could mean by 'rabbit,' what I mean (in the actual world) by, say, 'set of undetached rabbit parts.' The natural answer to this question

is 'No.' Hence on this interpretation of the thesis of The Under-determination of Translation by Physics, it should be rejected.

The thesis could, of course, be saved if it could be shown both that (i) theories of translation are not apriori consequences of the set of physical truths, and that (ii) whenever a claim is not an apriori consequence of a set of statements it is not a necessary consequence of those statements. As we shall see something like (i) is reasonably plausible. Thus if one were confused about the relationship between necessity and apriority, or if one failed to distinguish them, one might wrongly conclude that both Physicalism and The Underdetermination of Translation by Physics are jointly true, when determination is taken to be necessary consequence. The error here is the implicit reliance on the false claim (ii) — that any necessary consequence of a set of statements is an apriori consequence of that set. Once this error is removed there is no plausible route to the theses of the Underdetermination of Translation by Physics and the Indeterminacy of Translation, on the interpretation in which determination is construed as necessary consequence.

In light of this, it may be worthwhile to put the notion of necessity aside and examine a more epistemological conception of the determination relation. One might say that a set P of claims determines a set Q iff it is in principle possible, given the claims in P, for one to demonstrate the truth of the claims in Q, appealing only to logic and obvious apriori principles or definitions. The idea here is that for P to determine Q is for P to provide a theoretical basis for establishing Q that is absolutely conclusive, and that rules out any possibility of falsehood. In effect, determination is here construed as apriori consequence.

On this construal, the thesis of the Underdetermination of Translation by Physics is both interesting and plausible. Given the total set of behavioural evidence about stimulus meanings of the sort Quine identifies, I cannot absolutely establish that what a speaker means by one of his terms is what I now mean by 'rabbit,' as opposed to what I now mean by 'set of undetached rabbit parts' or 'temporal stage of a rabbit' (even if the speaker is me in the past). The claim that a speaker means one of these things rather than another is not an apriori consequence of the total set of claims about Quinean stimulus meanings. Moreover, it is hard to see how adding more behavioural facts (beyond Quinean facts about stimulus meaning), or more physical facts — about the neurological events in the person's brain or his physical interactions with

his environment — would, by itself, change the situation. Thus, it may turn out that theories of translation are epistemologically underdetermined by the set of all physical truths.

But how serious would this be? Is there some reason to believe that all genuine truths must be not just necessary consequences, but also purely apriori consequences, of the set of all physical truths? As far as I can tell this is not one of our pretheoretic convictions; nor has Quine given a theoretical explanation of why it must be maintained.[25] Hence, if theories of translation do turn out to be underdetermined by physics in this epistemological sense, such a result should be taken to show that the corresponding epistemic version of Physicalism is false, even though the metaphysical version of Physicalism, in which the determination relation is that of necessary consequence, remains true. Certainly, such a position is preferable to the radical and paradoxical rejection of facts about meaning and reference implied by Quine's Indeterminacy thesis.[26]

---

25 It is important not to confuse what is a genuine pretheoretic conviction — namely that we do know what our words, and those of our neighbors, mean — with what is not such a conviction — namely that we arrive at this knowledge by deriving true claims about the meanings of our words, and those of our neighbors, as apriori consequences of purely physical truths. Whereas the former claim is clearly true, the latter is almost certainly false.

26 Quine's doctrines of Physicalism, the Indeterminacy of Translation and the Inscrutability of Reference entail that claims of the sort, *Person P's word w refers to rabbits (as opposed to sets of undetached rabbit parts, etc.)*, are not determined by the totality of physical facts and so do not express genuine truths; ditto for any open formula *P's word w refers to x* relative to an assignment of an object as value of the variable 'x.' Given Quine's usual understanding of the existential quantifier, one can conclude from this that *∃x P's word w refers to x* never expresses a genuine truth. Supposing that it is nevertheless meaningful, we may conclude that *~∃x P's word w refers to x* will always be true; in effect, no one's words ever refer to anything. This is eliminativism about reference (as ordinarily understood). Pretty paradoxical, especially for someone who clearly is attempting to use words to refer to, and make claims about, things.

    Quine nowhere explicitly acknowledges such starkly paradoxical consequences of his views. The closest he comes is in the essay 'Ontological Relativity,' in Quine, *Ontological Relativity* (New York and London: Columbia University Press 1969), 26-68, at 47-51. There he notes the paradoxical consequences of

If this is right, there may be an interesting epistemological construal of the determination relation according to which the Underdetermination of Translation by Physics is true, while Physicalism and the Indeterminacy of Translation are false. However, this point must be qualified by a complication that I have neglected up to now. In order to derive any empirical theory T from the set of truths of physics, one must appeal to theoretical identifications, or bridge principles, relating the vocabulary of T to the vocabulary of the underlying physics. What conditions must these identifications or bridge principles satisfy in order to be available for such derivations? Is it enough that they be true, or must they also be apriori (or necessary)? In my epistemic characterization of the determination relation, I allowed only obvious apriori truths and definitions. With this understanding the Underdetermination of Translation by Physics is at least plausible (though it has not, of course, been demonstrated).

However, with this same understanding, Physicalism is threatened, not just by theories of meaning or translation, but by ordinary instances of theoretical reduction, such as the reduction of the biological concept of a gene to a physical construction involving the concept DNA. The relevant theoretical identity statement relating the two seems to be an empirical, a posteriori truth. Thus, if bridge principles relating the biological vocabulary to the physical vocabulary are restricted to apriori definitions, then our theoretical identity statement will be excluded, and the derivation of genetics from physical theory will be placed in jeopardy. Surely no one would conclude from this that genetics in particular, or biology in general, fail to state genuine truths; rather, if a choice has to be made, we will reject the (strong) epistemological version of Physicalism.[27]

---

applying the Indeterminacy and Inscrutability doctrines to our (present) selves, and attempts (unsuccessfully) to avoid these consequences by invoking his mysterious doctrine of reference relative to a coordinate system (48) and "acquiescing in our mother tongue" (49). Unfortunately, space limitations prevent me from providing a thorough discussion of these passages here.

27 This result could be avoided if it could be shown that there are genuinely apriori semantic definitions of 'gene' and 'DNA' from which, together with the set of purely physical truths, the theoretical identification of genes with DNA can be

---

In light of this, we could opt to weaken the epistemological determination relation so as to allow P to determine Q provided that Q is a logical consequence of P together with bridge principles, or theoretical identities, consisting of any truths — contingent, aposteriori, or otherwise — that relate the vocabulary of P to that of Q.[28] On this weakened

---

derived. Since I am not certain whether this is possible in principle, I am not certain that the case of genetics provides a genuine counterexample to the strong epistemological version of physicalism. By the same token, I am not certain that no genuinely apriori semantic definitions of notions like meaning and reference exist from which, together with all physical truths, claims about meaning and reference can be derived.

28 This conception of determination is closely related to familiar conceptions of theoretical reduction, which are used by Michael Friedman in 'Physicalism and the Indeterminacy of Translation,' *Noûs* 9 (1975) 353-73, to characterize Quine's theses of Physicalism and the Indeterminacy of Translation. There Friedman recognizes two kinds of reduction, strong and weak. Strong reduction is reduction in the classical sense. A theory T_2 is classically reducible to a theory T_1 iff the theorems of T_2 are logical consequences of T_1 together with a set D containing a 'definition' for each primitive predicate of T_2. A 'definition' is a universally quantified biconditional establishing the extensional equivalence of an n-place primitive predicate of T_2 with a corresponding formula of arbitrary complexity of the language of T_1.

The same point can be expressed in another way by noting that the 'definitions' appealed to in a reduction can be taken as establishing a mapping D from primitive predicates of T_2 onto coextensive open formulas of T_1. Given this, we may define the notion of an n-place (primitive) predicate P of T_2 being *satisfied by an n-tuple in an arbitrary model* M *relative to a mapping* D as consisting in the image of P under D being satisfied by that n-tuple in M. Classical (strong) reduction obtains when there is a mapping D such that every model of T_1 is a model-relative-to D of T_2. (See Friedman 357-8.)

Weak reduction is just like strong reduction except that the mapping D associates each primitive predicate of T_2 with a set of corresponding open formulas of the language of T_1. The set of formulas D associates with each primitive predicate P must be coextensive with P — i.e., as a matter of fact an n-tuple will satisfy P iff it satisfies at least one formula in the image of P under D. The notion of an n-place (primitive) predicate P of T_2 being satisfied by an n-tuple in an arbitrary model M *relative to such a mapping* D is then defined as consisting in there being at least one member of the set of formulas associated with P by D being satisfied by that n-tuple in M. As before, reduction obtains when there is a mapping D of this sort such that every model of T_1 is a model-relative-to D of

epistemological conception the usual reduction of the notion of a gene to DNA will pose no difficulty for the thesis of Physicalism. However, on this weakened conception, even theories of translation will be determined by the set of physical truths.

To see this let Sx be some formula specifying a set of physical facts satisfied by me and only me — so that Sx is true relative to an assign-

---

T_1. (Weak reduction differs from strong reduction only in cases in which the sets associated with the primitive predicates of T_1 are infinite.)

Friedman's stated reason (358) for allowing weak reduction to count as a genuine type of theoretical reduction is to make room for positions such as functionalist theories of mind which identify each token of a mental type with a particular physical realization, while recognizing arbitrarily many different ways in which the given type *might be physically realized*. Note, however, the modal notion here. Its use in characterizing the relevant functionalist theories points up a modest puzzle having to do with Friedman's position. Reduction, as he officially characterizes it, does not require the 'definitional' mapping D to pair the predicates of T_1 with formulas, or sets of formulas, that are intensionally equivalent to them in any sense. In particular D is not required to produce pairs that are extensionally equivalent in arbitrary counterfactual, or apriori imaginable, circumstances. Because of this the different merely possible, or merely imaginable, ways in which a mental type might be physically realized are, strictly speaking, irrelevant to the existence of 'definitional' mappings D satisfying Friedman's stated conditions for reduction. Since, as far as I know, physicalist functionalists never maintain that there actually exist infinitely many physically different kinds of realizations of a given mental type, they presumably ought to be reasonably confident in asserting the strong reducibility (in Friedman's official sense) of their theories to physics. Why then is there a need for the notion of weak reducibility? Does Friedman's use of the notion reflect an implicit desire to require the 'definitional' mappings in genuine reductions to provide more than actual coextensiveness? Do they also have to provide coextensiveness in all counterfactual (or in all apriori imaginable) situations as well? If so, then couldn't we define the determination relations needed to evaluate Quine's theses directly in terms of necessary, or apriori, consequence, as above?

Putting these and other subsidiary issues aside, I would like to acknowledge the essential correctness of some of Friedman's central points. In particular, he makes a plausible case for interpreting Quine's thesis of the Indeterminacy of Translation as the doctrine that theories of translation are not reducible (in either his strong or his weak sense) to the set of physical truths. He then argues for the correct (but understated) conclusion that Quine has given no compelling argument for the Indeterminacy Thesis, understood in this way.

ment of me as value of x, and is false relative to other assignments. It doesn't matter what this formula looks like — whether it is complex, whether we can identify it as applying just to me, etc. It only matters that it exists. For completeness we may suppose that it includes a specification of all physical facts about me relevant to my use of language — facts about my present and past brain states, facts about my behavioural dispositions, facts about my physical interactions with the environment, including causal and historical relations connecting my uses of individual words with other speakers and things, and so on. Let Lx be a similar formula applying uniquely to a certain Spanish speaker, Luisa. (For the sake of simplicity, suppose that I am a monolingual speaker of English and Luisa is a monolingual speaker of Spanish.) Now imagine a claim of the following sort, exhaustively listing the translation of the finitely many individual words of my language into words and phrases of Luisa's language.[29]

TSL: $\exists x \, \exists y \, [Sx \,\&\, Ly \,\&$ for all words w (of English) and words or phrases w^* (of Spanish), w as used by x means the same as w^* as used by y iff (i) w = 'woman' and w^* = 'mujer,' or (ii) w = 'headache' and w^* = 'dolor de cabeza,' or …

A corresponding claim lists the translation of the finitely many individual words of Luisa's language onto words and phrases of my language; similar claims may be imagined for each actual pair of language users — past, present and future. Since it is extremely plausible to suppose that there are only finitely many such pairs of speakers, it seems safe to assume that some extremely long and complicated general formula of the following sort exists which encompasses all the individual cases.

---

29 In this discussion I ignore certain practical complications such as the fact that some speakers speak more than one language, the fact that words of the language may be ambiguous, and the possibility that sometimes there may be no translation of a word in one language onto a word or phrase of the other language. Although these are real factors in translation, they are peripheral to Quine's philosophical claims about translation.

GT: For all speakers x and y, and for all words w in x's language and
words or phrases w^* in y's language, w as used by x means the
same as w^* as used by y iff (i) Sx & Ly &(w = 'woman' & w^* =
'mujer,' or w = 'headache' & w^* = 'dolor de cabeza,' or ...; or (ii)
Lx & Sy &(w = 'semaforo' & w^* = 'traffic light,' or ...; or (iii) Sx
& Gy & ...; or ...

GT may be regarded as a bridge principle providing a coextensive
physicalistic counterpart of the predicate *means the same as* applied to
words and phrases in theories of translation. Next we need a general
bridge principle that encompasses the combinatorial rules used in theo-
ries of translation to combine translations of parts into translations of
whole sentences. Presumably there are only finitely many such combi-
natorial rules for each pair of speakers. If, as we are assuming, there
are only finitely many such pairs of actual speakers, these rules can in
principle be exhaustively listed. This list together with GT can then be
used to formulate another bridge principle that provides a coextensive
physicalistic counterpart to the predicate *means the same as* as used be-
tween sentences in theories of translation. But then, theories of transla-
tion will be derivable from the set of all physical truths together with
these bridge principles, and so will count as determined by physics in
our weakened epistemological sense.

It might, of course, be observed that this trivialization just shows
that stronger conditions are needed to characterize the required deter-
mination relation. Two possibilities suggest themselves. First, it might
be claimed that what we want is not just a reduction of theories of
translation to physical truths, but rather a single reduction to the set of
physical truths of all theories making use of semantic notions such as
meaning and reference. Surely, if these notions are legitimate they will
have significant theoretical uses well beyond theories of translation in
the narrow sense considered here. The fact that a trivial reduction is
theoretically possible when translation theories are considered in iso-
lation does not show that such a reduction is possible in a context which
is properly more inclusive.

The point is well taken. But the problem with this suggestion is that
it takes us far beyond Quine's own discussion, and into uncharted
waters. Before we can make any progress on the question of whether a
single physicalistic reduction of all legitimate uses of semantic notions

is theoretically possible, we need a reasonably precise and exhaustive
characterization of the range of theoretical uses of semantic notions.
Until we have this, we can't evaluate the case for skepticism about
meaning (and reference) because no sufficiently articulated case for
skepticism about these notions has even been made.

The second possible response to the trivial reduction sketched above
is to stick to a physicalistic reduction simply of translation theories,
but to claim that what is wanted are bridge principles that provide
definitions yielding physical formulas coextensive with their
nonphysical counterparts in all counterfactual circumstances. The
bridge principles in the trivial reduction do not satisfy this demand,
since, for example, someone could have meant the same by a particu-
lar term as I do by 'woman' even though that person did not satisfy the
physical predicate identifying any actually existing individual.

I have no quarrel with this strengthened, modal constraint on the
determination relation. However, two points should be noted. First, if,
unlike Quine, we grant the legitimacy of modal notions, then we can
characterize determination directly in terms of necessary consequence.
But then, as I have maintained above, we have reason to believe that
theories of translation *are* determined by the physical truths. Second, if
one insists characterizing determination in terms of a strengthened re-
duction relation that requires physicalistic formulas which are neces-
sarily coextensive with the predicates and other vocabulary items used
in the theories undergoing reduction, then it is not clear what our atti-
tude should be towards the resulting strengthened versions of
Physicalism and the Underdetermination of Translation by Physics.
On this interpretation the latter thesis requires the existence of physi-
cal formulas (or sets of formulas)[30] necessarily coextensive with the
predicates *expression E as used by x means the same as expression E' as used
by y* and *sentence S as used by x means the same as sentence S' as used by y*.

---

30 On this interpretation Friedman's relation of weak reduction, strengthened by
the requirement that the mapping from predicates in T_2 to sets of formulas of
the language of T_1 produce pairs that are *necessarily coextensive* (in the sense of
note 28), should count as a genuine instance of determination in the sense
presently under consideration.

Are there such formulas (or sets of formulas)? There may well be, though it is hard to say for sure. Certainly Quine has given no compelling arguments to the contrary. Thus, on this interpretation, the case for the Underdetermination of Translation by Physics has not been made.

A similar point can be made regarding Quine's other premise — namely Physicalism — in the argument for Indeterminacy. On the present strengthened interpretation of the determination relation, it is not evident what we should think of the resulting version of Physicalism. I have already granted the truth of a weaker version of Physicalism, which states that all genuine truths must be necessary consequences of the physical truths. Is it obvious that we should add the further requirement that they be logical consequences of the Physical truths plus definitions that provide necessarily coextensive physical translations of all vocabulary items used in genuine truths? Since this point hasn't been established, the present interpretation of the determination relation is one in which neither of the two premises for the Indeterminacy Thesis has been secured. In light of this, the conclusion to be drawn is that Quine's argument for the Indeterminacy of Translation, like Kripke's skeptical argument about meaning and following a rule, fails to provide a compelling challenge to our pretheoretic convictions about meaning and translation.

This does not foreclose the possibility that something could be done to strengthen, or revive such a skeptical challenge. But surely the burden of proof is on those who wish to persuade us to adopt a radically skeptical attitude toward our ordinary semantic notions. Although I do not believe that any skeptical challenge of this sort could succeed, that has not been the burden of my argument. Instead, I have tried to defuse the particular skeptical arguments of Kripke and Quine by showing that their initial power is due in substantial part to an equivocation about what it is for one set of claims to determine another. Once this equivocation is removed, the original skeptical arguments lose their force, and it becomes highly dubious and problematic that any nonequivocal replacements could be found that would provide good reasons for a thorough-going skepticism.[31]