

Explaining the Disquotational Principle

JEFF SPEAKS
University of Notre Dame
Notre Dame, IN 46556
USA

Questions about the relative priorities of mind and language suffer from a double obscurity. First, it is often not clear *which* mental and linguistic facts are in question: we can ask about the relationship between any of the semantic or syntactic properties of public languages and the judgments, intentions, beliefs, or other propositional attitudes of speakers of those languages. Second, there is an obscurity about what 'priority' comes to here.

We can approach the first problem by way of the second. Often, talk about one class of facts being prior to another is glossed in terms of the claim that we can give an *account* of the latter in terms of the former. This gloss might seem to be a small advance, since it is not clear what sort of account we want. Intuitively, what is required is an account of the *nature* or *essence* of the facts in question — one property is prior to another, in the relevant sense, if the correct account of the essence of the latter includes the former. But what are the criteria by which we can judge the success of accounts of this sort? And why think that there are any such accounts to be had?

Recent work has suggested partial answers to these questions: accounts of the essence of some property should explain necessary truths involving that property.¹ This both suggests a justification for the

1 For an argument that the notion of essence should be regarded as explanatory

idea that such accounts are possible (since it is plausible that necessary truths require some explanation) and provides a test for the adequacy of such accounts (that they yield explanations of the relevant necessary truths). This suggests a way to raise questions about the priorities of mind and language more sharply: find some necessary truth connecting linguistic and mental facts and ask whether that necessary truth is to be explained in terms of the essence of the linguistic, or of the mental, facts.

An example of such a necessary truth is the following (simplified) disquotational principle, which connects the meanings of expressions in a language with the contents of the beliefs of users of that language:

If S means p and a speaker understands and sincerely accepts S , then the speaker believes p .²

Given the necessity of this principle (about which more below), we can formulate one clear version of the intuitive question about the relationship between linguistic meaning and the content of belief by asking which of the following two explanations of its necessity is correct:³

Explanation 1. The meanings of sentences are explained in terms of the contents of beliefs formed as a result of accepting those sen-

of, rather than defined by, modal notions, see Fine 1994. This is also related to the sorts of primitive grounding relations invoked in Schaffer 2009, and contrasts with more ‘modal’ ways of understanding priority in terms of supervenience relations. This view of essence is put to work in Setiya 2003, on which the method of the present paper is modeled. For an earlier, less explicit use of this method of using necessary truths to discern the nature of properties, see the discussion of the impossibility of self-blindness in Shoemaker 1994.

- 2 The principle is simplified because I ignore the need to relativize it to contexts of utterance. This principle corresponds to the weaker of the two disquotational principles discussed in Kripke 1979, though Kripke formulates the principle differently.
- 3 It’s worth being clear about a few limitations of this strategy. The first is that while an explanation of the disquotational principle might give us a satisfactory view of the relationship between linguistic meaning and the beliefs of agents, it would not follow that this result can be generalized to the relationship between linguistic meaning and other types of mental states — though it does seem plausible that, given the centrality of belief in understanding other sorts of mental states, the result would have some generality. The second limitation is that an explanation of the disquotational principle, to be successful, would also have to be consistent with the explanations of whatever other necessary truths there might be in the vicinity. My own view is that there are few necessary truths of this sort, but I do not argue for that here.

tences: details aside, what it is for a sentence to mean p is for it to be the case that someone who accepted the sentence would thereby believe p . On this view, the disquotational principle is necessary because sentences inherit their meanings from the contents of the beliefs with which they are correlated.

Explanation 2. The contents of beliefs are explained in terms of the meanings of sentences: details aside, what it is for an agent to believe p is for that agent to be disposed to accept a sentence of some public language which means p . On this view, the disquotational principle is necessary because beliefs inherit their contents from the meanings of the sentences the agent of the belief is disposed to accept.

Because the first of these explains meaning in terms of mental content, I shall call it the *mentalist* direction of explanation; because the second explains thought in terms of social facts about meaning, I shall call it the *communitarian* direction of explanation.⁴

To most recent philosophers, the view of language and mind corresponding to the mentalist direction of explanation of the disquotational principle has seemed by far the more promising one.⁵ The communitarian direction of explanation faces two apparently decisive objections: (i) many agents (including animals and infants) have beliefs without having dispositions to express those beliefs with sentences of public languages; (ii) accepting a sentence is a kind of intentional action, and intentional actions are themselves to be explained in terms of beliefs and other mental states; so an explanation of belief in terms of dispositions to accept meaningful sentences is circular.

The purpose of this paper is to defend a version of the communitarian explanation of the disquotational principle, and hence also the view that the foundations of mental content are to be found, in part, at the level of social facts about public language meaning. I will argue, first, that the disquotational principle must be explained by one or the other of the mentalist and communitarian explanations; second, that the mentalist explanation fails; and third, that the above objections to the com-

4 One might avoid either option by taking the disquotational principle to be explicable in terms of the interdependence of meaning and belief; I return to this possibility in III below.

5 Prominent defenses of broadly mentalist views of meaning include Grice 1957, Grice 1969, Lewis 1969, Schiffer 1972, Lewis 1975, and Loar 1981. The view is also presupposed by most work on the foundations of mental content; for a representative sample, see the essays in Stich and Warfield 1994.

munitarian explanation are not as decisive as they appear, because the assumptions about mental states and the nature of intentional action on which they rely are not as obvious as they are usually taken to be.

I Is the Disquotational Principle a Necessary Truth?

So far I have set up the question about the relationship between the contents of beliefs and the meanings of sentences as a question about the proper explanation of the necessity of the disquotational principle. But does this principle really express a necessary truth?

Some doubts about the disquotational principle might be occasioned by the obvious falsity of the following superficially similar principles:

If S means p and a speaker utters S , then the speaker believes p .

If S means p and a speaker believes S to be true, then the speaker believes p .

The first of these is falsified by cases in which the speaker's utterance is not sincere, as when the utterance is a joke, a line in a play, or a lie. The second principle corrects for these cases, but is falsified by cases in which the speaker fails to understand the sentence he believes to be true. I can trust a friend enough to come to believe that a certain German sentence is true without thereby acquiring a belief with the content of that sentence. The falsity of these two formulations shows that we need to take the qualifications built into the disquotational principle seriously. We need the qualification that the agent be sincere to rule out cases of deception and non-literal uses of language, and the qualification that the agent understand the sentence to rule out cases in which an agent endorses a sentence without knowing its meaning.

It is also worth keeping in mind that when we talk about accepting a sentence, we sometimes have in mind a restricted interpretation of 'acceptance', as when we speak of accepting a sentence for purposes of argument. But, fairly clearly, these restricted notions of acceptance are just that: restrictions of a more fundamental notion of acceptance *simpliciter*. With these qualifications in mind, it is hard not to agree with Kripke that 'taken in its obvious intent ... the principle appears to be a self-evident truth.'⁶

6 Kripke 1979, 249. One might, of course, worry that the rest of the discussion in 'A Puzzle About Belief' shows that the principle is far from self-evident — indeed, that it is false! I don't think that this is the right lesson to draw from the examples

II Analyzing Sincere Acceptance

But though emphasis on these qualifications might allay doubts about the necessity of the disquotational principle, it might give rise to the suspicion that the principle is trivial, and can be explained without any deep theorizing about the nature of meaning or belief. If we are building enough into the notions of understanding and sincere acceptance to secure the disquotational principle against counterexamples, it might seem that we are also building enough into these notions to explain the principle's necessity. The relevant notion of sincere sentence acceptance, it might be thought, just *presupposes* belief in what the sentence says.

But here we need to be clear about what 'presupposes' means. There is, of course, a sense in which anyone who thinks that the disquotational principle expresses a necessary truth thinks that sentence acceptance presupposes belief in the relevant proposition: if it didn't *entail* belief in the relevant proposition, the disquotational principle wouldn't be a necessary truth. So, if the view that the relevant notion of sentence acceptance presupposes belief is meant to be an objection to the view that the disquotational principle needs explanation, 'presupposes' must mean something other than 'entails.'

This suggests a better way to read the objection. The serious worry here is that sentence acceptance presupposes belief in such a way that the nature of sentence acceptance — independent of any analysis of meaning or belief — can provide a sufficient explanation of the necessity of disquotational principle. Put another way, the objection is that there is an analysis of 'understands and sincerely accepts' from which the necessity of the disquotational principle follows. If this were true, then the necessity of the disquotational principle wouldn't tell us much about the relationship between meaning and belief.⁷

Kripke discusses, however. One can derive a contradiction from Kripke's examples only by assuming, in addition to the disquotational principle, either the converse of the principle or some principle to the effect that anyone with contradictory beliefs will be in a position to discover this a priori. Though the matter is beyond the scope of this paper, I think that both of these principles are considerably less plausible than the disquotational principle. I discuss the latter principle further in II in connection with the question of whether it is possible to understand a pair of synonyms without knowing that they are synonymous.

⁷ The simplest way to raise this worry is to say that acceptance is a propositional attitude rather than a sentential attitude, and so that talk about sentence acceptance just is talk about belief in propositions. But while it is plausible that we do talk about accepting propositions, it is not plausible to deny that we also genuinely talk about accepting sentences. It seems as though 'accepts', like 'asserts', can be

To see how such a view of the disquotational principle might be worked out, suppose that we roughly analyze the sincere acceptance condition as the requirement that the speaker believe the sentence in question to be true, and roughly analyze the understanding condition as requiring that the speaker know the meaning of that sentence. Then, it seems, the disquotational principle is merely a disguised formulation of the following:

- [1] If S means p , and a speaker knows that S means p and believes that S is true, then that speaker believes p .

So far, this might not appear to be a necessary truth, as it must be if its necessity is to explain the necessity of the disquotational principle. But one might argue for its necessity by noting, first, that knowledge of meaning brings with it knowledge of truth conditions, so that the following is a necessary truth:

- [2] If a speaker knows that S means p , then that speaker believes that S is true iff p is true.

Second, one might note that every instance of the following schema is a trivial, necessary truth:

- [3] $(A \text{ believes that } s \ \& \ A \text{ believes that } (s \equiv s')) \rightarrow A \text{ believes that } s'$

But if [2] is a trivial necessary truth, and every instance of [3] is a trivial necessary truth, it follows that the following consequence of the conjunction of [2] with the relevant instance of [3] is a necessary truth:

- [4] If a speaker knows that S means p and believes that S is true, then that speaker believes that p is true.

But, supposing (as is plausible) that it is a necessary truth that if one believes that p is true one believes p , [4] entails [1]. So we might seem to have established the intended conclusion that [1] is a trivial necessary truth: [2] and (every instance of) [3] seem to be trivial necessary truths and, as we've just seen, from these [1] follows. So, if the disquotational

used as a term for either a sentential or a propositional attitude — the proponent of the idea that the disquotational principle is a substantial necessary truth will have the former interpretation of the term in mind.

principle really is just a disguised version of [1], we seem to have the desired explanation of its necessity.

Though initially attractive, this trivializing explanation of the necessity of the disquotational principle is not, I think, successful. The fundamental problem is that instances of [3] are simply not, in general, necessary truths; as is well-known, we don't believe all of the logical consequences of what we believe. We often have a pair of beliefs that entail a third but, for whatever reason, fail to see this consequence.

One dramatic way to illustrate this possibility is to show that an agent may believe two propositions without even *being in a position* to know a third proposition which is a trivial consequence of the first two. Cases in which an agent understands two synonymous expressions without knowing that they are synonyms provide ready examples. To borrow a case from Nathan Salmon, imagine that Sasha satisfies the standards for competence with both 'catsup' and 'ketchup' without knowing that they have the same meaning; suppose, further, that Sasha has encountered bottles labeled 'catsup' only at breakfast, and bottles labeled 'ketchup' only at lunch. Suppose that Sasha walks into his favorite breakfast restaurant, where he's eaten catsup for breakfast many times; he'd certainly be willing to accept, and hence, by the disquotational principle, believe, the proposition expressed by,

[5] There's some catsup on the table.

As always, Sasha is also willing to accept

[6] There's some ketchup on the table if and only if it is time for lunch.

But of course, 'catsup' and 'ketchup' being synonyms, [6] expresses the same proposition as

[7] There's some catsup on the table if and only if it is time for lunch.

Since, given the disquotational principle, Sasha believes the proposition expressed by [6], it follows that he believes the proposition expressed by [7]; despite this, Sasha is clearly in no position to see, a priori, that it is a consequence of these propositions that

[8] It is time for lunch.

So, given only the disquotational principle and the assumption that Sasha can understand two synonyms without knowing that they are

synonymous, it follows that Sasha believes the propositions expressed by [5] and [7] without believing, or indeed even being in a position to see any reason to believe, [8]. But this is a counterexample to the claim that every instance of [3] is true.⁸

This is enough to undercut the argument given above for the necessity of [1] — but one might wonder whether we can argue directly that [1] is false. I think that we can. One sort of counterexample is a case in which I fail, for whatever reason, to integrate two of my beliefs. Suppose that I am over at the apartment of two German friends, and I hear one say to the other, ‘Zwiebelkuchen ist köstlich.’ Since I know that they are very trustworthy people, I immediately come to believe the proposition expressed by

[9] ‘Zwiebelkuchen ist köstlich’ is true.

Later, I decide to enroll in German classes to help me unlock the secrets of German cuisine, and I learn that

[10] ‘Zwiebelkuchen ist köstlich’ means that onion cake is delicious.

As I am rather immersed in my learning of the German language, I don’t connect this new bit of knowledge with the beliefs about true sentences I’ve gained from my German friends until the lesson is over. Of course, I then realize that my beliefs jointly entail that onion cake *is* delicious, and come to hold this new belief. But, prior to the integration of these two beliefs, it seems clear that there is a certain sentence — namely, ‘Zwiebelkuchen ist köstlich’ — which is such that I believe that it is true, and know what it means, but don’t believe the proposition that it expresses. This is a counterexample to [1], which is therefore not a necessary truth, and hence not the correct analysis of the disquotational principle. This sort of case, I take it, makes it clear that counterexamples to [1] are not just possible, but commonplace; they are just familiar examples of a lack of integration of two beliefs where the subject matter of the relevant beliefs is a sentence.

But even if this blocks one way of explaining the necessity of the disquotational principle in terms of the nature of sincere acceptance, it doesn’t undercut every such explanation. For example, suppose that sincere acceptance of a sentence is not, as suggested above, a matter of

⁸ For discussion of this sort of example, see Salmon 1990. Thanks to anonymous referees for suggestions that led to the improvement of the preceding argument.

believing a sentence to be true while knowing its meaning, but rather a matter of there being some proposition p such that one both believes that the sentence means p , and accepts the sentence because one believes p .⁹ Given that causal 'because' claims are factive, then, if we combine this view of sincere acceptance with an analysis of understanding as knowledge of meaning (ignoring cases in which a speaker has more than one belief about the meaning of the sentence in question), we have a pair of analyses from which the disquotational principle is a logical consequence.

A worry about this strategy is that it succeeds in explaining the disquotational principle only by placing overly strong conditions on the sincere acceptance of sentences. It seems as though we often sincerely accept sentences we understand without this being caused by a belief with the content of the sentence in question. Consider, for example, coming to accept a sentence on the basis of an argument. I might believe the premises of the argument to be true, and come to accept the conclusion of the argument because I see that it follows from the premises. My acceptance of the conclusion is perfectly sincere, but is not caused by my having a prior belief with the content of the conclusion.

But there is a more fundamental problem with this sort of analysis of sincere acceptance. The analysis treats sincere acceptance as the compound of two independent elements: the act of accepting a sentence, and the property of being sincere, where the latter is understood as there being some proposition p such that the agent believes that the sentence means p , and accepts it because he believes p . These elements are not just independent in the sense that we could have acts of sentence acceptance which are not sincere; they seem to be independent in the sense that it is possible that *every* act of sentence acceptance in a given world or community could fail to be sincere.¹⁰ After all, on the present view, the act of sentence acceptance is one which is constitutively independent of questions about sincerity, and insincere sentence acceptance (again, on the present view) is just a matter of there being no proposition such that you believe the sentence to express that proposition and you accept the sentence because you believe that proposition. So imagining a world in which every act of sentence acceptance is insincere is just a matter of imagining acts of a certain type — acts of sentence acceptance — being permuted in a certain way with the beliefs of agents. When-

9 Thanks to Kieran Setiya for pointing out this possibility.

10 The argument which follows is based on Sarah Stroud's discussion of the possibility of global akrasia in her excellent paper 'Weakness of the Will and Practical Judgment.'

ever the act-type in question can be understood without recourse to the relevant beliefs of agents, this seems possible. An example here might help. We can understand the act of raising one's arm intentionally independently of questions about the purpose of that arm-raising; and, for this reason, it is coherent to imagine customs of arm-raising where these serve virtually any purpose. Just so, if the act of accepting a sentence can be understood independently of acts of sincere acceptance, it should be coherent to imagine a linguistic community in which acts of the former take place, but not acts of the latter type.

But this sort of linguistic community does not seem possible. Insincere utterances of sentences — whether cases of lying, or of the non-literal use of language — presuppose a background of serious, literal uses of language.¹¹ The relationship between sincere and insincere acceptance is not like the relationship between red and green jellybeans, which are simply two subspecies of a common kind, either of which could exist without the other; it is more like the relationship between real and counterfeit money.¹² Just as it is impossible for there to be counterfeit money without there being real money to counterfeit, so it is impossible for there to be insincere accepting of sentences without a background of acts of sincere acceptance. If this is right, then this is an indication that insincere acceptance should be understood in terms of the prior notion of sincere acceptance, rather than the two being taken as different and independent categories of the more general act type of sentence acceptance. But if this is right, then the sorts of explanations of the disquotational principle in terms of analyses of sincere acceptance we're considering — which, by treating sincere and insincere acts of accepting sentences as on a par, entail the possibility of insincere sentence acceptance without since sentence acceptance — can't be right.

Perhaps there is some analysis of sincere acceptance that does not have this form, and can explain the necessity of the disquotational principle. But at this point I think we can say that the most intuitively plausible candidate analyses fail, and that there's no particular reason to think that another will succeed.

11 See, for example, the discussion of the population of inveterate liars in Lewis 1975, 182-3.

12 The example is due to Geach 1956, and is discussed in Stroud 2003. See also the discussion of arguments from illusion in Dancy 1995.

III Interdependence

One way of escaping the requirement that we explain the necessity of the disquotational principle in terms of the essence of belief or of meaning is to deny that the principle is necessary, and another is to claim that, given the nature of understanding and sincere acceptance, it is trivial; yet a third is to grant that it is a substantial necessary truth, but deny that its necessity requires explanation in terms of the essence of one of either meaning or belief.

The response I have in mind is based not in a wholesale rejection of the demand for explanation of necessary truths, but rather in the fact that some necessary truths seem not to admit of explanation in terms of the essence just one property figuring in those truths. Consider, for example, the following:

$$\forall x \forall y (x \text{ is to the east of } y \iff y \text{ is to the west of } x)$$

Though this is plausibly a necessary truth, it is not plausible to require that it be explained in terms of the essence of either of the relations expressed by 'is to the east of' or 'is to the west of.' This is a good example of two properties whose necessary connection is plausibly explained in terms of the essence of neither, but rather by their being interdefined in terms of some third property — in this case, the direction of the earth's axis. Could the necessity of the disquotational principle be explained by belief and meaning being related in this way?

That belief and meaning are so related was, in essence, the view of Donald Davidson. In his view, belief and meaning are both defined in terms of the principles which should govern a radical interpreter: in particular, the principle that we should interpret an agent's beliefs and utterances so as to maximize the truth of each (while making allowances for explicable errors). The dilemma that Davidson sees as leading to this conclusion is as follows:

A central source of trouble is the way beliefs and meanings conspire to account for utterances. A speaker who holds a sentence to be true on an occasion does so in part because of what he means, or would mean, by an utterance of that sentence, and in part because of what he believes. If all we have to go on is the fact of honest utterance, we cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief.¹³

13 Davidson 1973a, 314. See also Davidson 1973b, 134-5. For a definitive later statement of Davidson's views on this topic, see Chapter 3 of his *Truth & Predication*.

Transposed from Davidson's epistemological idiom, the dilemma can be put as follows: if we had an independent account of what it is for an agent to have a belief with a certain content, we could ascend from there to an account of what it is for a sentence to have a meaning; if we had an independent account of what it is for a sentence to have a meaning, we could ascend from there to an account of what it is for an agent to have a belief with a certain content; but in fact neither sort of independent account is available. Davidson's solution, of course, is that we must define belief and meaning together, in terms of an independent third fact: the fact that the beliefs of an agent, and the meanings of her words, are whatever they must be in order to maximize her true beliefs (modulo qualifications about explicable errors). This suggests an alternate explanation of the necessity of the disquotational principle, which is on par with the kind of explanation of the necessity of the east-west principle suggested above: belief and meaning are 'interdefined' in terms of events of agents holding true, or accepting, sentences.

A simplified version of Davidson's interdefinition of belief and meaning, which abstracts from, among other things, the need to make room for explicable error, might be stated as follows:

(An agent believes p & S means p for the agent) \iff (the agent holds true S iff p is true)

It is an interesting and important question whether some suitably qualified biconditional of this sort might be true. But the important point for present purposes is that, even if it were true, it would not explain the necessity of the disquotational principle. Any explanation of the necessity of the disquotational principle must explain why the following state of affairs is impossible: an agent understands a sentence which means p , and accepts it sincerely and reflectively, and yet fails to believe p . Davidson's interdefinition of belief and meaning does not explain this. That interdefinition says only that the following two classes of facts necessarily covary:

- (i) An agent means p by some sentence S , and believes p
- (ii) The agent holds S true iff p is the case.

But, even if the relevant facts of classes (i) and (ii) *did* covary, this does not rule out the possibility that the agent could accept a sentence which means p *without* believing p . The Davidsonian analysis only makes a claim about what must be the case when an agent *both* means p by a sentence *and* believes p ; it makes no claim about what follows from the

truth of the first conjunct, along with the fact that the agent accepts the relevant sentence.¹⁴

I conclude that the disquotational principle, suitably restricted, is necessary; that it is not trivial, and hence needs some explanation; and that it cannot be explained via a Davidsonian interdefinition of belief and meaning.¹⁵ Hence the necessity of the disquotational principle must be explained either in terms of the nature of meaning or in terms of the nature of belief.

This way of putting things brings out one important benefit of this way of raising the question about the relationship between mind and language via the explanation of the necessity of the disquotational principle: it, so to speak, forces a choice on the question. So long as one agrees that the necessity of the disquotational principle requires explanation in terms of the nature of meaning or of belief, one cannot simply rest content with the position that there is no substantial theory of the nature of belief and meaning, or with the position that belief and meaning are 'basic' or 'primitive.' For, if there were no substantial theory of the nature of belief and meaning, the necessity of the disquotational principle would be without explanation. It is a constraint on any view

-
- 14 The claim that the Davidsonian interdefinition fails to explain the necessity of the disquotational principle can also be expressed as the claim that the following argument, with the Davidsonian analysis as premise and the disquotational principle as conclusion, is invalid:

$$\frac{(A \text{ believes } p \text{ and } S \text{ means } p \text{ for } A) \iff (A \text{ accepts } S \text{ iff } p \text{ is true})}{(A \text{ accepts } S \text{ \& } S \text{ means } p) \rightarrow A \text{ believes } p}$$

Here all variables should be understood as bound by universal quantifiers with wide scope over everything in the formulae except the necessity operator.

- 15 Could some other, non-Davidsonian interdefinition of belief and meaning fare better? Here is one option, which might be suggested by some things Davidson says. Perhaps we should understand 'believes', 'means' and other such terms as 'theoretical terms' to be defined in the manner of Lewis 1970; if this were right, then plausibly the disquotational principle would count as one of the platitudes used in constructing the definition. This would amount to quite a different approach to the disquotational principle and other such necessary truths than the one pursued in this paper: on this sort of view, they would not be truths in need of explanation so much as that in terms of which explanations of the meanings of terms like 'believes' are to be constructed. My own view is that this strategy is unlikely to succeed because there are simply not enough platitudes around to define the relevant terms, but discussion of this is well beyond the scope of this paper. The point of mentioning this is to make clear that this strategy, which has some affinities with the 'interdefinition' strategy discussed in the text, is an alternative not closed off by any arguments in this paper. Thanks to an anonymous referee for helpful discussion of this point.

of the nature of mental content and linguistic meaning that it be able to provide such an explanation.¹⁶

Which way does the explanation go? We can begin by considering the view about the relationship between mind and language which, largely for the reasons sketched at the outset, has been dominant in recent philosophy: the mentalist view that the contents of mental states like beliefs constitute, and hence are more fundamental than, facts about the meanings of expressions in public languages.

IV The mentalist direction of explanation

In his 1927 paper, 'Facts and Propositions,' Frank Ramsey had this to say about meaning:

The essence of pragmatism I take to be this, that the meaning of a sentence is to be specified by reference to the actions to which asserting it would lead, or, more vaguely still, by its possible causes and effects.¹⁷

Discussions earlier in the paper indicate that when he linked the meaning of a sentence with the causes and effects of asserting it, Ramsey had in mind especially the propositional attitudes — in particular, the beliefs — which a competent speaker would have by virtue of asserting a sentence. Ramsey's remarks suggest giving an account of the meaning of a sentence in terms of the content of the belief a speaker would acquire by accepting that sentence in that context.

This is the kind of 'theory of meaning' a proponent of the mentalist direction of explanation should want. If we can find a way of spell-

16 Philosophical positions which run afoul of this constraint might include the quietism of John McDowell, the minimalism espoused by Mark Johnston in 'The End of the Theory of Meaning,' and, perhaps, for different reasons, the Schiffer of *Remnants of Meaning* and the Schiffer of *The Things We Mean*. Nothing I've said, however, rules out the possibility that *either* belief or meaning might be primitive, in the sense of there being no substantial theory of what it is for an agent to have a certain belief or a sentence to have a certain meaning. A theory of one of the two might be enough to explain the necessity of the disquotational principle.

Another (less common) position which is threatened by the present line of argument is the idea that both meaning and belief are susceptible to analysis, but that the two analyses are independent: meaning is defined in terms which make no mention of belief, belief is defined in terms which make no mention of meaning, and the two are not interdefined in terms of some third thing. Millikan seems to hold this view; see especially Millikan 2001.

17 Ramsey 1927, 51

ing out the idea that the meanings of sentences are fixed by the beliefs with which they are correlated, then this will provide the wanted explanation of the necessity of the disquotational principle in terms of the essence of meaning: it is a necessary truth that if someone accepts a sentence which means p , then that person believes p because this is just what it is for a sentence to express a proposition: for individuals who accept it to believe that proposition.

The burden of such a theory will be to say what it takes for a sentence to be 'correlated' with a belief. The simplest attempt at doing this is to say that a sentence is correlated with a belief just in case *if* an agent accepts the sentence, *then* the agent will form the belief. But of course this 'if-then' can't be understood as expressing a material conditional, since then for any sentence S that no member of the relevant community accepts, then for *any proposition* p , the analysis would entail that S means p . But of course a sentence may be a meaningful sentence of a language, even if every speaker of the language thinks it false.

The natural reply is that, even if no speaker accepts the sentence in question, it remains true that *were* a speaker to accept it, the speaker *would* have a belief with the content of the sentence. We can incorporate this into our mentalist theory of meaning by making the consequent of the right-hand side into a counterfactual conditional, giving us the following mentalist theory of meaning [M]:

$$[M] \quad (S \text{ means } p) \equiv \forall a (a \text{ is a member of the relevant community} \rightarrow (a \text{ accepts } S \square \rightarrow a \text{ believes } p))$$

There are delicate issues about the interpretation of the counterfactual on the right-hand side of [M]. It is tempting to argue that it is falsified by sentences like '0=1,' which are so obviously false that everyone who understands them rejects them. Consider the nearest possible world in which I accept the '0=1.' Isn't a world in which the sentence has a different meaning more similar to the actual world than one in which I believe, absurdly, that zero and one are identical?

It is plausible that counterexamples of this sort can be blocked by stipulating away 'backtracking' evaluations of the counterfactual in [M], in which one rejects a counterfactual on the grounds that, had the event mentioned in the antecedent of the counterfactual occurred, something in the past (relative to that event) would also have been different, where this latter makes the consequent of the counterfactual false.¹⁸ In the present case, it is natural to think that, if I had accepted '0=1', this

18 For discussion, see Lewis 1979.

could only have been because '0=1' expressed some non-trivially-false proposition; and this latter consideration is what makes it false that, in the world under consideration, I believe that 0=1. But, by ruling out backtracking evaluations of this sort, perhaps we can put cases like this to the side.¹⁹

The real problem with [M] is that facts about the meanings of sentences are, in a certain sense, more fine-grained than facts about the beliefs of agents; and this makes it difficult to give an account of the former in terms of the latter.

Belief distributes over conjunction; that is, it is a necessary consequence of the truth of the proposition expressed by a sentence of the form 'α believes that σ and σ'' that, in that context, the propositions expressed by the following sentences are also true: 'α believes that σ'' and 'α believes that σ''. Now consider a conjunctive sentence *S* whose meaning is the conjunction of the propositions *p* and *q*. Then each of the following claims are true:

□∀*a* (*a* accepts *S* → *a* believes *p* & *q*)

□∀*a* (*a* accepts *S* → *a* believes *p*)

□∀*a* (*a* accepts *S* → *a* believes *q*)

The problem is that it follows from [M], along with the truth of these three claims, that *S* is three ways ambiguous: *S* means *p* & *q*, *S* means *p*, and *S* means *q*. After all, if [M] is to be true, then anyone who accepts *S* must believe the conjunctive proposition *p*&*q*; but the distribution of belief over conjunction entails that they will also believe *p* and believe *q*. So all three beliefs are correlated with the conjunctive sentence; and if meaning is defined in terms of correlated beliefs, the sentence has as its meaning each of these three propositions. But this is incorrect: conjunctive sentences are not, in general, three ways ambiguous just by virtue of being conjunctions.²⁰

19 A residual worry here is that in making this stipulation, we are illegitimately building facts about what sentences mean into the class of facts held fixed for purposes of evaluation of the counterfactual. If we stipulate that possible worlds relevant to the evaluation of the counterfactual in [M] share their history with the history of the actual world, then we are stipulating that expressions in those worlds mean the same thing as they do in the actual world. But facts about the meanings of expressions are the facts for which [M] is supposed to provide an explanation. I am unsure whether this is a genuine circularity or just the appearance of one.

20 Actually, matters are a bit worse than this for [M]. Normal cases of ambiguous sentences are ones in which the sentence is sometimes correlated with one belief,

One might object at this stage: is this really so clear? Isn't there a sense of 'means' according to which a sentence $\ulcorner\sigma \text{ and } \sigma'\urcorner$ *does* mean that σ , and that σ' ?²¹ To answer this question, it is important to be clear that we are talking about the semantic contents of sentences (relative to contexts), rather than, for example, what a speaker means by uttering a sentence in a given context. (It is plausible that speaker-meaning *does* distribute over conjunction.) So the question is not whether there is any sense of 'means' on which this is correct, but rather whether, given that by 'means' we mean 'semantically expresses,' this is possible. Given this clarification, we can argue as follows that conjunctions do not, in general, semantically express the propositions expressed by their conjuncts. On any approach to semantics, the meaning of a sentence in a context determines its truth conditions. So, if two sentences have different truth conditions, they have a different meaning – in the sense of meaning were interested in. But in the standard case conjunctions will have different truth conditions than at least one of the conjuncts; hence (in at least these cases) the conjunction can't semantically express the same proposition as that conjunct.

But, while this does show that [M] is false, it seems that there is an easy way to revise it in answer to this objection. We can limit the account of meaning to non-conjunctive sentences, and give a separate account of the meanings of conjunctive sentences in terms of the meanings of their parts.

But this is only a superficial solution; there are counterexamples similar in form to those based on the distribution of belief over conjunction even for simple sentences. To generate such a counterexample, one needs only a sentence S and two non-conjunctive propositions p, q such that every speaker in the relevant community who would believe p upon accepting S would also believe q . For simplicity, I'm focusing on cases in which believing q is a metaphysically necessary consequence of believing p ; but the class of pairs of propositions which can generate counterexamples to [M] is much larger than this.

But even if we restrict ourselves to cases in which having one belief entails having another, there are plenty of plausible examples. Consider, for example, the proposition that Lassie was a brown dog (or collie, or

and sometimes with another; 'I am going to the bank' is sometimes correlated with beliefs about financial institutions, and sometimes with beliefs about the sides of rivers. But in this case we have a sentence that is always correlated in the relevant way with three beliefs. So the result is not just that conjunctive sentences are ambiguous; it's that they are ambiguous in a heretofore unimagined way. Thanks to Tom Feeney for pointing this out.

21 Thanks to an anonymous referee for pressing this worry.

puppy) and the proposition that Lassie was a dog. Is it possible for an agent who accepts, and therefore understands, the sentence 'Lassie was a brown dog,' and who believes that Lassie was a brown dog, not to believe that Lassie was a dog? If not, [M] entails that 'Lassie was a brown dog' is ambiguous between meaning that Lassie was a brown dog and meaning that Lassie was a dog; this is clearly incorrect. Colors provide a similar example — is it, e.g., a necessary truth that if someone believes that an object is scarlet, she also believes that the object is red? If so, then we can generate a similar counterexample in the obvious way. Another such class of cases exploits the trivial equivalence — paradoxical cases aside — of a proposition p and the proposition which attributes truth to p . Suppose that an agent accepts the sentence 'It is true that Lassie was a dog', and so by the disquotational principle believes that it is true that Lassie was a dog. Is it possible that such an agent could yet fail to believe that Lassie was a dog? Again, it seems unlikely.

One response to these kinds of cases is to try to come up with some criterion for separating simple from complex sentences, and to hold that (i) [M] is only applicable as a theory of meaning for simple sentences, (ii) the meanings of complex sentences are a function of the meanings of the simple sentences, and (iii) all the above cases turn on illicit applications of [M] to complex sentences. This strategy faces two main challenges. First, we need some criterion which classes together all of the sentences which generate counterexamples of the above form; second, we need this criterion to be loose enough to count as simple enough sentences to provide the resources to supply the missing theory of meaning for complex sentences. The prospects for finding a criterion to meet both constraints do not appear to me to be good.²²

22 Both challenges seem to me to pose serious problems. I am reluctant to think of 'is scarlet,' for example, as being syntactically complex; and even if it were, I am unsure how the assignments of meanings to simple sentences could be extended to a complete theory that applies to complex sentences as well.

A related move for the mentalist is to try to solve the problem via a distinction between the various propositions that are candidate meanings for a sentence. One proposal which several people have suggested to me is that, if it is a necessary truth that someone who accepts a sentence S believes a set of propositions $p_1 \dots p_n$, we should assign as the meaning of S the strongest of these propositions. This works nicely for some of the cases discussed above, such as the conjunctive sentence. Because the conjunctive proposition entails but is not entailed by its conjuncts, the present proposal assigns, correctly, the conjunctive proposition as the meaning of the conjunction. The solution isn't general though, as is shown by a small variant on the preceding one: if one of the conjuncts is a necessary truth, then neither the conjunction nor the other conjunct will be uniquely strongest; but,

It is tempting to try to solve this problem by appealing to a distinction between beliefs gained ‘directly’ as a result of accepting a sentence, and beliefs gained ‘indirectly.’ In one sense, this is clearly correct; if S means p , then there is, intuitively, a more direct link between the acceptance of S in by an agent and that agent’s believing p than between the agent’s acceptance of S and the agent’s coming to believe any other propositions. The problem, though, is not to find intuitions of directness; the problem is to explain what these intuitions are tracking in a noncircular way. A first natural idea is that directness should be spelled out in causal terms. But this use of causation appears misplaced; it is not as though, in the sorts of cases described above. The agent comes to believe p and then, a moment later, comes to believe q . Rather, it is a fact about belief that the former is metaphysically sufficient for the latter.

One natural response to these problems is to say that [M], while fundamentally correct, is just a bit over-simple; perhaps meaning can be defined in terms of belief, but not by the sort of simple pairing of sentences and beliefs suggested by [M]. One natural way to develop this idea is to pursue an analysis along the lines of (Lewis 1975), who seeks to analyze the meanings of sentences not just in terms of what speakers who accept those sentences believe, but also in terms of mutual beliefs held by other speakers of the language about those beliefs. Adding Lewis’ requirement of mutual knowledge to [M] might give us

- [L] $(S \text{ means } p) \equiv (1) \forall a(a \text{ is a member of the relevant community} \rightarrow (a \text{ accepts } S \square \rightarrow a \text{ believes } p)) \ \& \ (2) \text{ this is mutually known by the members of the relevant community.}$

But it seems that the sorts of counterexamples to [M] discussed above are easily transformed into problem cases for [L]. We can, after all, imagine a linguistically sophisticated community, in which, for example, all speakers know — and know that each other know — that belief distributes over conjunction. (Or fill in your favorite of the above examples.) For such a community, the examples above will work as well against a convention-theoretic account of Lewis’ sort as they do against the simpler [M].²³

for all that, the conjunction is not ambiguous. The example of a proposition and the proposition that ascribes truth to it is similar.

23 Here I am ignoring details about how, exactly, clause (2) of [L] is to be interpreted; see Lewis’ discussion of the distinction between knowledge *in sensu diviso* and knowledge *in sensu composito* in Lewis 1969, 64-68. While [L] is inspired by the theory of Lewis 1975, that theory was a bit more complicated:

Perhaps we can do better by invoking the idea of an internalized semantic theory for the language.²⁴ In the present context, this promises to provide the following kind of explanation of the difference between beliefs directly and indirectly gained by accepting sentences:

Everyone who understands the language is in possession of an internalized language-processing system which, upon hearing a sentence *S*, maps *S* onto a sentence *M* of the internal processing system which means the same thing as *S* does in the agent's public language (or idiolect). In cases where accepting a sentence *S** yields more than one belief (as with conjunctions), we can distinguish between the belief gained directly by accepting the sentence (the one with the content of the sentence paired with *S** by the internal language processor) and beliefs gained indirectly.

There are really two ways to read this invocation of an internalized semantic theory. On the one hand, it might be a replacement for our mentalist theory [M]; on the other, it might be a supplement to it. Fol-

[L*] *x* means *p* in a population *G* ≡

- (1a) ordinarily, if a member of *G* utters *x*, the speaker believes *p*,
- (1b) ordinarily, if a member of *G* hears an utterance of *x*, he comes to believe *p*, unless he already believed this,
- (2) members of *G* believe that (1a) and (1b) are true,
- (3) the expectation that (1a) and (1b) will continue to be true gives members of *G* a good reason to continue to utter *x* only if they believe *p*, and to expect the same of other members of *G*,
- (4) there is among the members of *G* a general preference for people to continue to conform to regularities (1a) and (1b)
- (5) there is an alternative regularity to (1a) and (1b) which is such that its being generally conformed to by some members of *G* would give other speakers reason to conform to it
- (6) all of these facts are mutually known by members of *G*

However, the ways in which [L*] diverges from [L] are either detrimental to [L*] or irrelevant to the point at hand. In particular, I take it that clauses (3) and (4) fail to block the sorts of counterexamples being considered, and that, as Burge 1975 argues, clause (5) leaves the theory open to counterexamples. For interesting further discussion of problems with the formulation of convention-based accounts of meaning related to the meaningfulness of very long or complex sentences of a language which are never used, see the series of articles Hawthorne 1990, Lewis 1992, O'Leary-Hawthorne 1993.

24 Thanks to Kieran Setiya for the idea that the mentalist might take this route.

lowing the first option, we might be inclined to think that, once we have the internalized semantic theory on the table, counterfactual connections between acts of accepting sentences and beliefs are superfluous. Why not just offer the following theory of meaning: a sentence S means p in a group if most or all members of the group have internalized semantic theories which map S onto some sentences of their internal language which has the content p ? Despite its appealing simplicity, this theory is a nonstarter. We are trying to find a theory of meaning that is fit to explain why the disquotational principle is a necessary truth. [M] does that: the disquotational principle is derivable from it. But the present theory does not make any mention of belief; so it is very hard to see how it could explain the necessity of a principle connecting meaning and belief.

It might seem more promising to use the notion of an internalized semantic theory (and the attendant distinction between beliefs gained directly and indirectly by accepting sentences) as a supplement rather than a replacement for [M]. The theory of meaning which results would then be something like this:

- [M+] S means $p \equiv \forall a$
- (i) a is a member of the relevant community \rightarrow
(a accepts $S \square \rightarrow a$ believes p), &
 - (ii) a 's internalized semantic theory translates S onto a sentence of a 's internal language which has the content p

We are looking for an explanation of a necessary truth; the kind of explanation currently on offer is an explanation of this necessary truth in terms of the nature of linguistic meaning. Since the current explanation makes essential use of the fact that each of us possesses an internalized theory that translates sentences of our public language onto synonymous sentences of an internal language, it had better be a necessary truth that any speaker of a language should possess such an internalized theory. But this seems to be, at best, a contingent truth: perhaps it is the best explanation of our ability to understand novel utterances, but we do not in general expect that the best explanation of a phenomenon should provide a metaphysically necessary condition for the occurrence of that phenomenon.²⁵

25 Here I'm following David Lewis, who claimed that 'it makes sense to say that languages might be used by populations even if there were no internally represented grammars.' See Lewis 1975, 178.

A simple example may make the point more explicit. The following seems to be a coherent scenario: a possible community speaks a language with the semantics of English. Each member of the community possesses an internalized theory which maps sentences of English onto at least one sentence of a mental language. Sometimes, however, that theory maps the English sentences onto a number of distinct sentences of the mental language. In particular, that theory may map conjunctions onto three sentences of the mental language: one with the content of the conjunction, and two with, respectively, the contents of each of the conjuncts. If this appears to you to be a coherent scenario — and you are not tempted to reject the initial hypothesis that they speak a language with the same semantics as English — then you agree that possession of an internalized semantic theory of the form employed in [M+] is not a necessary condition on speaking a language. And then it is clear that possession of such an internalized theory cannot be used in an explanation of the necessity of the disquotational principle.

Moreover, [M+] is an awkward blend of two very different approaches to the foundations of meaning; it is doubtful whether these can be combined to form a stable view. What should the proponent of [M+] say when the two clauses of the analysis come apart? Suppose, for example, that every member of the relevant community accepts *S* only when they believe *p*, and there's no other proposition that stands in this relation to *S*, but that the internalized semantic theories of the members of that community map *S* onto a formula which expresses the proposition *q*, and $q \neq p$. One doesn't want to say that *S* is meaningless, which seems to leave us two options: either *S* means *p*, or *S* means *q*. But on the former option, we're just giving up [M+] for [M], our original theory, leaving us without the needed explication of the direct/indirect distinction; and on the latter option we're using internalized semantic theories as a replacement rather than a supplement to [M], leaving us no explanation of the disquotational principle. So the proponent of [M+] must say that cases like this are impossible. But why should it be impossible for the beliefs and internalized semantic theories of members of a community to be aligned in this way?

It is important here not to be so focused on the counterexamples as to lose sight of the source of the objection. The basic point is that it seems to be in the nature of belief that facts about the beliefs of agents come, so to speak, in clumps: for many propositions, it is a necessary truth that any agent who believes such a proposition must also believe one or more distinct but related propositions. But facts about meaning just are not 'clumpy' in this way: there are no propositions which are such that it is a necessary truth that, if a sentence has it as its meaning, it must have one or more distinct propositions as its meaning as well. Because facts about meaning are, in this sense, more fine-grained than

facts about belief, it is difficult to see how we could give an account of the former in terms of the latter.²⁶

V Communitarianism Reconsidered

So much for the mentalist attempt to explain the necessity of the disquotational principle in terms of a theory of linguistic meaning which claims that sentences inherit their meanings from beliefs with which they are, in some way or other, correlated. Can a communitarian explanation, which explains the disquotational principle in terms of a theory of belief which claims that beliefs inherit their contents from the meanings of public language sentences which agents accept, fare any better?

At the outset I noted two fundamental problems with the communitarian direction of explanation that, I think, have convinced most contemporary philosophers of mind and language to favor either mentalism or the kind of 'interdependence' view discussed in II:

1. Beliefs and other mental states are possessed not only by adult language-using human beings, but also by non-linguistic animals and infants. So we cannot explain what it is to have a belief with a certain content in terms of relations to sentences of public languages.
2. The communitarian will attempt to explain the beliefs of an agent in terms of the sentences that agent accepts. But accepting a sentence is a kind of intentional action; and, since intentional actions are to be explained (roughly) as bodily movements caused in the right way by beliefs and desires, an explanation

26 It is worth mentioning briefly yet another mentalist response to the foregoing argument. Not all mentalists, of course, have sought to analyze meaning in terms of belief; many, following Grice, have sought to explain meaning in terms of the communicative intentions of speakers. Might such an analysis, if successful, explain the disquotational principle — especially if, as many have thought, there is a necessary connection of some sort between meaning and belief? I am skeptical about the extensional adequacy of Gricean analyses but, these worries aside, I think that it is clear that these analyses offer little prospect of explaining the disquotational principle. Griceans focus on the links between meaning something by an utterance and intentions to cause beliefs in *an audience*; and even if it is a necessary truth that if one intends to Φ one believes that one will Φ , it is hardly a necessary truth that if one intends to cause a belief that p one must, oneself, believe that p .

of the nature of belief in terms of a kind of intentional action is straightforwardly circular.

These are both serious challenges to communitarianism. My aim here will be to respond to them indirectly: by claiming that a version of each challenge can be raised against a mentalist as well as against a communitarian. This defensive point, along with the arguments given above against mentalist explanations of the disquotational principle, shows that the communitarian explanation of the necessary connections between belief and meaning is a viable alternative to mentalism.

Consider the first objection, based on the possibility of animal beliefs. If the communitarian resists the temptation to, implausibly, deny that non-linguistic creatures can have beliefs about their environments, this objection pushes her toward a disjunctive theory of belief: a theory which claims that what it is for an agent to have a belief is for that agent to be *either* disposed to accept a sentence *or* to satisfy some other, non-linguistic condition. One problem with this move is simply that of spelling out this non-linguistic condition. But there is a further problem of principle, which is of course that, absent special circumstances, disjunctive accounts look ad hoc and unconvincing.

But there are two points to be made here in the communitarian's defense: (i) there are reasons to think that a disjunctive account of belief might not be so unnatural; and (ii) as Mark Greenberg has argued, there is a sense in which standard theories of mental content are disjunctive as well.²⁷

Consider, first, two important differences between the beliefs of language-using adults and those of non-linguistic animals. First, language-using adults have a kind of privileged first-person access to the contents of their own beliefs; in standard cases, whenever an agent believes *p*, the agent is in a position to discover that she believes *p*. This is not true of non-linguistic creatures; there is no presumption that, just because Fido believes that there is a bone in the yard, Fido should be in a position to discover that he has this belief. A second difference is that the range of the contents of the beliefs of normal adults, though not of non-linguistic animals, extends beyond the deliverances of perceptual experience to beliefs about mathematics, morality, and the distant past and future. If a disjunctive account can provide explanations of such differences within the domain of facts being analyzed, then the disjunctive character of the account is a point for rather than against it.

27 See Greenberg ms.

Second, consider the form of a non-communitarian theory of belief which makes no use of facts about public languages. Initially, it looks as though there should be no problem in giving a unified account of the beliefs of language- and non-language-using creatures; a functionalist theory, for example, might try to account for both in terms of the relational properties of states of the relevant creatures' brains. But this is not what we find when we look at the writings of most theorists. The kinds of thought-experiments made prominent by Tyler Burge have convinced many that, within the theory of mental content, we have to give special treatment to cases in which mental states are due not to the agent satisfying some functional characterization but rather to the agent *deferring* to other members of his community.²⁸ But, as Greenberg has pointed out, this amounts to giving a disjunctive theory of belief (or of concept possession, or of mental content): the theorist is claiming that one can have a belief with a given content *either* by instantiating some functional property *or* by deferring in appropriate ways to members of his community.²⁹ So the mentalist complaint that a communitarian explanation of the disquotational principle leads to a disjunctive account of belief has little dialectical force.

The second objection to the communitarian direction of explanation, involving the use of facts about intentional action, in some ways cuts deeper. The thought that an account of what it is for an agent to act intentionally will involve causation by the kinds of propositional attitude states for which the communitarian is trying to give an account is a plausible one; and, if this idea is correct, it does imply that the communitarian's account of belief is circular. But, again, this point also poses a problem for the mentalist opponent of communitarianism.

The main alternative to a communitarian account of belief is some form of functionalism. Any plausible functionalist account of belief will include in the functional role constitutive of a state's being a belief with a certain content both relations between that state and perceptual 'input conditions' which cause an agent to be in that state, and behavioral 'output conditions' which are caused by the agent being in that state (along with, in most cases, relations to other inner states which stand in regular causal relations to the state in question). The salient question then is: is the behavioral output which figures in a functional-

28 See, e.g., Burge 1998, Burge 1982. For a clear discussion of this topic by a proponent of a theory of belief that is disjunctive in the present sense, see Peacocke 1992.

29 For an excellent critical discussion of the use of deference in the theory of content, see Greenberg ms. Greenberg does not endorse either of the two kinds of disjunctive theories discussed here.

ist account of belief a matter of 'non-intentional' bodily movements, or of intentional actions the agent is disposed to perform as a result of being in the belief state in question?

This is a difficult question; but I think that there is good reason to think that the behavioral output conditions will have to include facts about the intentional actions of agents. Consider, for example, the role that the output conditions play in the plausible form of functionalism about belief put forward by Robert Stalnaker in *Inquiry*. On that view, the output condition which a state must satisfy to be a belief state with content *p* is, roughly, that that state make the agent disposed to act so as to satisfy her desires in a world in which *p* and her other beliefs are true. Suppose, for example, that we are trying to determine whether an agent believes that there is beer in the refrigerator. Suppose that the agent desires a beer; then the output condition attached to the belief would be straightforwardly satisfied by the agent in question walking up to the refrigerator, opening it, and taking out a beer. Presumably, though, the output condition should not be satisfied by an agent tripping on the way past the refrigerator (or being disposed to trip on the way past the refrigerator) in such a way as to knock the door open and jar loose a bottle of beer — though this action will satisfy his desire in a world in which it is true that there is beer in the refrigerator as well as the first. The salient difference here seems to be that the first action, but not the second, is intentional. If so, then this form of functionalism employs facts about intentional actions as surely as does a communitarianism which makes use of dispositions to accept sentences of a language.

Of course, the claim that this result generalizes to other functionalist treatments of belief is not trivial; and to show that an objection applies to others as well as to one's own theory is not to reply to the objection.³⁰ Communitarianism of the kind being sketched will only ultimately be acceptable if it includes a satisfactory view of the relationship between belief and action. But the foregoing indicates that the main objections to communitarianism are also objections to its rivals, and that the mentalist explanation of the necessity of the disquotational principle faces some surprising problems. This is enough to show that the communitarian explanation of the disquotational principle is plausible; and this is also enough to show that the view that public language meaning has

30 One might not be worried by this sort of objection if one took 'intentional action' to be among the theoretical terms defined by platitudes connecting meaning and belief; this is related to the position mentioned in note 15.

a role to play in the constitution of thought has more to be said for it than most recent theorists have thought.³¹

References

- Burge, T. (1975). 'On Knowledge and Convention.' *Philosophical Review* **84**, 249–55.
- Burge, T. (1982). 'Two Thought Experiments Reviewed.' *Notre Dame Journal of Formal Logic* **23**, 284–293.
- _____. (1998). 'Individualism and the Mental.' In P. Ludlow and N. Martin, eds., *Externalism and Self-Knowledge*. Stanford, CA: CSLI Publications: 21–83.
- Dancy, J. (1995). 'Arguments from Illusion.' *Philosophical Quarterly* **45**, 421–38.
- Davidson, D. (1973a). 'Belief and the basis of meaning.' *Synthese* **27**, 313–28. (CHECK REF??)
- _____. (1973b). 'Radical Interpretation.' *Dialectica* **27**, 313–28. (CHECK REF??)
- _____. (2005). *Truth and Predication*. Cambridge, MA: Harvard University Press.
- Fine, K. (1994). 'Essence and Modality.' *Philosophical Perspectives* **8**, 1–16.
- Geach, P. (1956). 'Good and Evil.' *Analysis* **17**, 33–42.
- Greenberg, M. (ms.). 'Incomplete Understanding, Deference, and the Content of Thought.'
- Grice, P. (1957). 'Meaning.' *Philosophical Review* **66**, 177–88.
- _____. (1969). 'Utterer's Meaning and Intentions.' *Philosophical Review* **78**, 147–77.
- Hawthorne, J. (1990). 'A Note on "Languages and Language."' *Australasian Journal of Philosophy* **68**, 116–18.
- Johnston, M. (1988). 'The End of the Theory of Meaning.' *Mind & Language* **3**, 28–42.
- Kripke, S. (1979). 'A Puzzle About Belief.' In A. Margalit, ed., *Meaning and Use*. Boston: D. Reidel: 239–83.
- Lewis, D. (1969). *Convention*. Cambridge, MA: Harvard University Press.
- _____. (1970). 'How to Define Theoretical Terms.' *Journal of Philosophy* **67**, 427–46.
- _____. (1975). 'Languages and Language.' Reprinted in Lewis 1983: 163–88.
- _____. (1979). 'Counterfactual Dependence and Time's Arrow.' *Noûs* **13**, 455–76.
- _____. (1983). *Philosophical Papers*, Volume 1. Oxford: Oxford University Press.

31 Thanks to Scott Soames, Kieran Setiya, Mark Greenberg, and the participants in colloquia at the University of Southern California and the University of Maryland at College Park for helpful discussions of previous versions of this essay.

- _____. (1992). 'Meaning Without Use: Reply to Hawthorne.' *Australasian Journal of Philosophy* **70**, 106–10.
- Loar, B. (1981). *Mind and Meaning*. Cambridge: Cambridge University Press.
- Millikan, R. (2001). 'The Language-Thought Partnership: A Bird's Eye View.' *Language and Communication* **21**, 157–66.
- O'Leary-Hawthorne, J. (1993). 'Meaning and Evidence: A Reply to Lewis.' *Australasian Journal of Philosophy* **71**, 206–11.
- Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.
- Ramsey, F.P. (1927). 'Facts and propositions.' In his *Philosophical Papers*. Oxford: Basil Blackwell: 34–51.
- Salmon, N. (1990). 'A Millian Heir Rejects the Wages of Sinn.' In C.A. Anderson and J. Owens, eds., *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*. Stanford, CA: CSLI Press: 117–39.
- Schaffer, J. (2009). 'On What Grounds What.' In D.M. David Chalmers and R. Wasserman, eds., *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.
- Schiffer, S. (1972). *Meaning*. Oxford: Oxford University Press.
- _____. (1987). *Remnants of Meaning*. Cambridge, MA: MIT Press.
- _____. (2003). *The Things We Mean*. Oxford: Clarendon Press.
- Setiya, K. (2003). 'Explaining Action.' *Philosophical Review* **112**, 339–94.
- Shoemaker, S. (1994). 'Self-Knowledge and "Inner Sense": Lecture II: The Broad Perceptual Model.' *Philosophy and Phenomenological Research* **54**, 271–90.
- Stich, S. and T. Warfield, eds. (1994). *Mental Representation: A Reader*. Cambridge, MA: Basil Blackwell.
- Stroud, S. (2003). 'Weakness of Will and Practical Judgement.' In S. Stroud and C. Tapolet, eds., *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press: 121–46.