Published in *The Aristotelian Society* supplementary volume 77 (2003), pp. 249-93

BLIND REASONING

by Paul Boghossian and Timothy Williamson

II—Timothy Williamson

UNDERSTANDING AND INFERENCE

ABSTRACT The paper challenges the inferentialist account of concept possession that Paul Boghossian takes as a premise in his account of the transmission of justification by deductive reasoning in his paper 'Blind Reasoning'. Unorthodox speakers who reject the inferences in an alleged possession condition can still have the concept by understanding a word for it. In that sense, the inferences are not analytic. Inferentialist accounts of logical constants, theoretical terms (using the Ramsey-Carnap-Lewis method) and pejorative expressions such as 'Boche' are examined and rejected. It is suggested that epistemological questions cannot be reduced to questions in the theory of thought and meaning.

Introduction. In his paper 'Blind Reasoning', Paul Boghossian asks how deductive reasoning can transfer justification for belief from the premises of an argument to its conclusion. According to the view that he calls *Simple Inferential Externalism*, the key to the answer is just the validity of the inference itself, provided that the justification for believing the premises does not inappropriately depend on the conclusion itself. Although a deductively valid inference is perfectly reliably truth-preserving, Boghossian rejects Simple Inferential Externalism as subject to counterexamples similar to those that undermine other forms of reliabilism: cases in which someone is epistemically blameworthy for using a reliable process because he is completely in the dark about its reliability (Bonjour 1985: 41). At the opposite extreme, *Simple Inferential Internalism* tries to solve the problem by requiring the thinker to be able to know that the inference is deductively good. Boghossian rejects this alternative because, he argues, it is forced to invoke a mystifying faculty of rational insight and in any case generates circular justifications. Thus the condition that the inference is valid is insufficient for the transfer of justification, even granted the background conditions on the standing of the premises and conclusion; the condition that the thinker is able to know that the inference is valid is unnecessary. In Boghossian's Wittgensteinian terminology, the elusive condition must permit the inference to be blind but entitling. After dismissing some deflationary accounts of blind entitlement, Boghossian devotes the rest of his paper to refining the proposal that '[A] deductive pattern of inference P may be blamelessly employed, without any reflective appreciation of its epistemic status, just in case inferring according to P is a precondition for having one of the concepts ingredient in it'. [1] For example, if inferring according to modus ponens is a precondition for having the concept of the conditional, then modus ponens may be blamelessly employed without any reflective appreciation of its epistemic status. The thought is that inferring unreflectively according to P cannot be

blameworthy *per se*, for one can properly reflect on P only after one has its ingredient concepts, therefore only after one is already inferring according to P.

This paper concentrates on the concept constitution account of blameless blindness. It casts doubt on the existence of the requisite tight connections between what concepts one has and how one infers. At the meta-linguistic level, the rules of inference for a connective are in a sense not analytic: accepting them is not a precondition for understanding it. Thus doubt falls on inferentialist accounts of concept possession and linguistic understanding.

I

Logical Unorthodoxy and Concept Possession. As concept possession is usually conceived, a paradigmatic way to have a concept C is to understand a word that means C. For example, one can have the concept of the conditional, the concept *if*, by understanding the English word 'if' or a synonym in another language. Although understanding a word that means *if* may not be necessary for having the concept *if*, it is sufficient.[2] For to understand the word is to know what it means; since 'if' means *if*, to understand 'if' is to know that 'if' means *if*; how could one know that without having the concept *if*? To know that 'if' means *if* is not merely to know that the sentence '"if" means *if*' expresses a truth, for a monolingual speaker of Chinese could know that '"if" means *if*' expresses a truth on the reliable testimony of a bilingual speaker of Chinese and English without understanding a single word of English, in particular, without knowing that 'if' means *if*.

Is making inferences by modus ponens of the form 'If A then B; A; therefore B' a precondition for having the concept *if*? Vann McGee, a distinguished logician, has published

purported counterexamples to modus ponens.[3] Presumably, he refuses to make some inferences by modus ponens. Does McGee lack the concept *if*? Although I deny that his purported counterexamples to modus ponens are genuine, I admit that they have some initial plausibility on the ordinary understanding of 'if'. In conversation with McGee, he appears to understand the word 'if' quite well by ordinary standards. He certainly seems to know what we other English speakers mean when we use the word 'if'. Before he had theoretical doubts about modus ponens, he understood the word 'if' if anyone has ever understood it; surely his theoretical doubts did not make him cease to remember what it means. We may therefore assume that McGee has the concept *if*, just like everyone else.

Are only some inferences by modus ponens such that a willingness to make them is a precondition for having the concept *if*? Presumably, McGee will accept most inferences by modus ponens. However, any particular inference by modus ponens might be rejected by a student who, having just read McGee's article, was sufficiently impressed to try to put it into practice, sometimes unfortunately misapplying its doctrine through excess of zeal. Has the student ceased to understand the word 'if"? He may still use it competently in ordinary conversation; he may reject only very salient inferences by modus ponens, while making many others unawares. The student still has the concept *if*.

Is a willingness to make most inferences by modus ponens (no matter which ones) a precondition for having the concept *if*? Even this watered-down claim seems too weak for a concept constitution account of blameless blindness. For suppose that we wish to explain why a given inference that you made by modus ponens was blameless. The proposed explanation is that making the inference is a precondition for having the concept *if*. But we have just seen that making *that* particular inference, the relevant instance of modus ponens, is not a precondition for having the concept *if*, because the foolish student might have refused to

make the same inference yet still had the concept *if*. Although the student might be willing to make lots of other inferences by modus ponens, how does that help to explain the blamelessness of the inference that you actually made? You could have refused to make it yet still retained the concept *if*. The inference is not unavoidable in the way assumed by the motivating thought behind the concept constitution account of blameless blindness.

Could we invoke Putnam's division of linguistic labour, and say that making any given inference by modus ponens is a precondition only for *full* understanding of 'if', the kind of understanding characteristic of the expert rather than the layman?[4] The trouble is that McGee *is* an expert on conditionals. He publishes on them in the best journals. He does not defer in his use of 'if' to any higher authorities. Although the student is not an expert, any given inference by modus ponens might be rejected by some expert on conditionals, on the basis of a complex theoretical argument. The expert would be mistaken, but making a theoretical error about the logic of conditionals is quite compatible with fully possessing the concept *if*.

The problem is not specific to modus ponens, which is as good a candidate as any for an inferential precondition for concept possession. Consider any concept C, any word 'C' of a natural language that means C and any deductive pattern of inference P of which C is an ingredient. It is proposed that inferring according to P is a precondition for having C. Let Inst be any instance of P. An expert on C and P fully understands 'C'. She then becomes convinced by a complex theoretical argument that Inst is invalid, and therefore rejects Inst; her argument does not generalize in any obvious way to more than a small proportion of instances of P. By ordinary standards, the expert continues to understand 'C' fully. In conversation, she uses 'C' appropriately, and responds appropriately when others use it. She still has the concept C, despite her unorthodoxy over Inst. Thus willingness to make the

inference Inst is not a precondition for having C, even if willingness to infer in most instances according to P is such a precondition. Consequently, we cannot explain why one is blameless in making the inference Inst by saying that willingness to make it is a precondition for having C.[5]

The argument is independent of the validity of the pattern of inference P, and of the validity of its instance Inst. If Inst is valid, then the expert is mistaken in ruling it invalid, but she is nevertheless still an expert; complex theoretical arguments sometimes seduce even experts into accepting wildly false conclusions. If Inst is invalid, then its rejection by the expert may just manifest her expertise.

Might the person who has the concept C while unwilling to make the inference Inst retain the defeasible *disposition* to make the inference, overridden by another disposition, the latter induced by theoretical reflection? For something can be disposed to X in certain circumstances yet be prevented from Xing in those very circumstances by some defeating condition.[6] Someone generally disposed to accept inferences of the form 'A, B; therefore A and B' may assent to the premise 'Ticket *i* will not win' for each ticket in a lottery, yet dissent from the conjunctive conclusion. Boghossian might prefer not to put much weight on such an appeal to dispositions, for he has argued vigorously elsewhere against dispositional accounts of concept possession (1989: 528-40), although in the present paper he does say 'we come to grasp the logical constants by being disposed to engage in some inferences'. It is certainly unclear how having a concept could *consist in* having a disposition that is overridden by one's conscious reflection. Most of us have dispositions to commit various logical fallacies which we control by conscious reflection; presumably, the latter dispositions play no decisive role in determining which concepts we possess. One is not following a rule when one consciously refuses to act in accordance with that rule.[7]

What has not been argued is that one could have the concept *if* without any disposition at all to infer by modus ponens. The disposition will be defeasible by theoretical reflection. It may even be defeated pretheoretically by features of some kinds of examples; perhaps McGee's are like that, since they have some initial plausibility. [8] However, the mere disposition is not enough for Boghossian's account of justification, even when the disposition is undefeated. Perhaps someone can happily make the inference although justification is not transferred from premises to conclusion because he was surrounded by fallacious yet plausible theoretical considerations to which he had no answer; they left him unmoved only as a result of his pigheadedly obstinate character; he *ought* to have been moved to doubt. However, the main issue is whether, when justification is transferred, Boghossian can explain why it is transferred. For his explanation depends on the idea that one is blameless because one has no option: one cannot so much as entertain the inference unless one is willing to make it. If one can entertain the inference while unwilling to make it, provided that one has a defeasible disposition to make it, no such account of blamelessness is forthcoming. Suppose that there is a kind of cruelty that one cannot so much as imagine unless one has a defeasible disposition to commit cruel acts of that kind: it does not follow that one is blameless in committing such acts. Even if one compulsively imagines them, one can still be blamed for not resisting the temptation to live out one's fantasies, for not defeating one's disposition. [9]

As will emerge in the next section, Boghossian restricts the simple form of his explanation to concepts that he classifies as non-defective: but even the qualified explanation does not work if concept possession is a matter of defeasible dispositions. For it does not answer the question about an inference involving a non-defective concept: if a defeasible disposition to make the inference is necessary for the ability to entertain that entertain, how does that justify one in indulging the disposition?

II

Pejoratives and Conventional Implicature. Boghossian first considers an unqualified form of the concept constitution account, based on this principle about the meaning-entitlement connection:

(MEC) Any inferential transitions built into the possession conditions for a concept are *eo ipso* entitling.

He objects that there are clear counterexamples to (MEC).

Arthur Prior (1960) stipulated these disastrous rules for a connective 'tonk':

Tonk-Introduction	Tonk-Elimination
A	A tonk B
---	---
A tonk B	B

If inferring according to both Tonk-Introduction and Tonk-Elimination is a precondition for having the concept *tonk*, then by (MEC) such inferences are entitling, so anyone who has the concept can blamelessly infer anything from anything. However, Boghossian does not treat that as a clear counterexample to (MEC), because one might plausibly deny that 'tonk'

expresses any concept whatsoever. It is certainly hard to see how anyone could seriously make unrestricted use of both Tonk-Introduction and Tonk-Elimination. But in other cases no such defence of (MEC) seems plausible.

Boghossian gives the example of pejorative terms such as 'Boche'. Following Dummett's treatment (1973: 397 and 454), he suggests that to have the concept *Boche* is to be willing to infer according to these introduction and elimination rules:

Boche-Introduction	Boche-Elimination
x is German	*x* is Boche
---	---
x is Boche	*x* is cruel

Thus, according to (MEC), inferences by Boche-Introduction and Boche-Elimination are entitling, so anyone who has the concept can *eo ipso* blamelessly infer that Germans are cruel.[10] As Boghossian points out, it would be highly implausible to deny that the word 'Boche' expresses a concept; surely xenophobes use sentences in which it occurs to express complete thoughts, however bad those thoughts are. Since, in having the concept, one is not *eo ipso* in a position blamelessly to infer that Germans are cruel, Boghossian concludes that (MEC) requires modification.

The 'Boche' objection to (MEC) rests on the premise that one has the concept *Boche* only if one is willing to infer according to both Boche-Introduction and Boche-Elimination. Is that premise plausible? I think that I am one counterexample, and that Boghossian is another. Unlike someone who thinks that the word 'Boche' means *master*, we both fully understand

the word, for we understand the sentences in which it occurs that xenophobes utter; we know what 'Boche' means; we know that it means *Boche*. We find racist and xenophobic abuse offensive because we understand it, not because we fail to do so. Presumably, therefore, we have the concept *Boche*. Yet neither of us is willing to infer according to both Boche-Introduction and Boche-Elimination. Similarly, imagine a reformed xenophobe who once was willing to infer according to those rules but now has seen the error of his ways, while vividly recalling with shame what it was like to shout xenophobic abuse. He still remembers what 'Boche' meant. Since he knows that it meant *Boche*, he retains the concept *Boche*. But how can one have a concept while rejecting its constitutive rules?

Let us consider the example in more depth. On a widespread simple picture, the introduction and elimination rules for an atomic expression play a constitutive role in determining its reference. If exactly one assignment of reference makes the rules truth-preserving, then that assignment is correct (this is supposed to be the case for the connectives of standard first-order logic). If more than one assignment of reference makes the rules truth-preserving, then it is indeterminate which of those assignments is correct (a supervaluationist treatment is sometimes suggested). If no assignment of reference makes the rules truth-preserving, then every assignment is incorrect: the expression does not refer (although it may still have a sense, and express a concept). Since there are non-cruel Germans, no assignment of reference to 'Boche' makes both Boche-Introduction and Boche-Elimination truth-preserving. Therefore, on the picture just sketched, 'Boche' does not refer. We might therefore expect ascriptions of 'Boche' to lack a truth-value. On this picture, Boche-Introduction and Boche-Elimination carry equal weight in determining reference, or the lack of it, for 'Boche'. Let us test this claim by considering two candidate assignments of reference to 'Boche'.

(A) Suppose that 'Boche' refers to the property of being German. Thus 'Lessing was Boche' is true, because Lessing was German, but 'Nero was Boche' is false, because Nero was not German.[11] Boche-Elimination is not truth-preserving: every non-cruel German yields a counterexample. But Boche-Introduction is truth-preserving; in fact, being German is the strongest property that makes Boche-Introduction truth-preserving (when assigned as referent to 'Boche', given the intended interpretation of the other words). Equivalently, the converse of Boche-Introduction is also truth-preserving:

Boche-Elimination*

x is Boche

---

x is German

In Dummett's terms, the rules of Boche-Introduction and Boche-Elimination* are in harmony: Boche-Elimination* merely allows one to infer from '*x* is Boche' what Boche-Introduction allows one to infer '*x* is Boche' from, and Boche-Introduction merely allows one to infer '*x* is Boche' from what Boche-Elimination* allows one to infer from '*x* is Boche'. Consequently, if the two rules are added to a theory in a language without 'Boche', they yield a conservative extension: anything in the original language provable with them was already provable without them.[12] Boche-Elimination* is the elimination rule that naturally corresponds to Boche-Introduction. Together, Boche-Introduction and Boche-Elimination* would determine the property of being German as the referent of 'Boche'. This is the assignment of reference that naturally corresponds to Boche-Introduction.

(B) Suppose that 'Boche' refers to the property of being cruel. Thus 'Lessing was Boche' is false, because Lessing was not cruel, but 'Nero was Boche' is true, because Nero was cruel. Boche-Introduction is not truth-preserving: every non-cruel German yields a counterexample. But Boche-Elimination is truth-preserving; in fact, being cruel is the weakest property that makes Boche-Elimination truth-preserving. Equivalently, the converse of Boche-Elimination is also truth-preserving:

Boche-Introduction*

x is cruel

---

x is Boche

The rules of Boche-Introduction* and Boche-Elimination are in harmony: Boche-Introduction* merely allows one to infer '*x* is Boche' from what Boche-Elimination allows one to infer from '*x* is Boche', and Boche-Elimination merely allows one to infer from '*x* is Boche' what Boche-Introduction* allows one to infer '*x* is Boche' from. Consequently, if the two rules are added to a theory in a language without 'Boche', they yield a conservative extension: anything in the original language provable with them was already provable without them. Boche-Introduction* is the introduction rule that naturally corresponds to Boche-Elimination. Together, Boche-Introduction* and Boche-Elimination would determine the property of being cruel as the referent of 'Boche'. This is the assignment of reference that naturally corresponds to Boche-Elimination.

On the picture of Boche-Introduction and Boche-Elimination as carrying equal weight

in determining reference, or the lack of it, for 'Boche', cases (A) and (B) have exactly the same status; the advantages and disadvantages of the two putative assignments of reference are symmetrically related. But that result is wildly implausible. Intuitively, Boche-Elimination* is just as intimately linked to the meaning of 'Boche' as Boche-Introduction is. By contrast, Boche-Introduction* has no standing at all; even most xenophobes would reject it, for they think that Germans are not the only cruel foreigners.[13] Again, although both 'Lessing was Boche' and 'Nero was Boche' are regrettable utterances, the former seems to combine xenophobic abuse with literal truth while the latter combines it with literal falsity. The xenophobic abuse is preserved under negation, but 'Lessing was not Boche' seems to be literally false while 'Nero was not Boche' seems to be literally true. Thus Boche-Introduction trumps Boche-Elimination.

A further asymmetry between the two rules is observable in practice. Xenophobes typically treat Boche-Elimination as defeasible, because they allow that there are a few good Germans ('Some of my best friends are Boche'), whereas they treat both Boche-Introduction and Boche-Elimination* as indefeasible. Indeed, Boche-Elimination scarcely makes sense for many adjectival uses of 'Boche': xenophobes do not reason from 'He owns a Boche car' to 'He owns a cruel car'.

As a further check, we may briefly consider a third proposal about the reference of 'Boche'. It attempts a compromise between (A) and (B).

(C) Suppose that 'Boche' refers to the conjunctive property of being both cruel and German. Like (B), (C) makes Boche-Elimination truth-preserving and Boche-Introduction not truth-preserving; (C) has an advantage over (B) in making Boche-Elimination* truth-preserving and counting 'Nero was Boche' as literally false and 'Nero was not Boche' as literally true. But (C) is at a disadvantage with respect to (A), since it counts 'Lessing was

Boche' as literally false and 'Lessing was not Boche' as literally true. The new proposal also licenses the silly inference from 'He owns a Boche car' to 'He owns a cruel car'. Indeed, its classification of Boche-Elimination as exceptionlessly truth-preserving is a demerit rather than a merit, for that rule, we have noted, is defeasible even from the perspective of those who use 'Boche'. By contrast, Boche-Introduction is supposed to be indefeasible, so (A) has the advantage over (C) in classifying Boche-Introduction as exceptionlessly truth-preserving.

The natural conclusion is that, far from suffering reference-failure or massive indeterminacy of reference, 'Boche' has the same reference as 'German'. That is certainly the dictionary view of the matter. Under 'Boche', the *Concise Oxford Dictionary* gives the definition '(Contempt[uous] for) German'. Thus Boche-Introduction and Boche-Elimination* are exceptionlessly truth-preserving, while Boche-Elimination is very far from truth-preserving. Accounts that make Boche-Introduction and Boche-Elimination the basic rules of use for 'Boche' are therefore highly implausible.

In Fregean terminology, we may say that 'Boche' has the same reference as 'German' but a different tone. Indeed, since the differences between 'Boche' and 'German' apparently play no role in determining reference, and so make no difference to the way in which the terms contribute to the truth-conditions of sentences in which they occur, a Fregean might even count 'Boche' and 'German' as having the same sense. Frege himself gives just such an account of another pejorative term (1979: 140): 'cur' has the same sense and reference as 'dog' but a different tone. According to Frege, 'This dog howled the whole night' and 'This cur howled the whole night' express the same thought (the same sense). They differ in that only the latter conveys an attitude of contempt for the dog on the part of the speaker. Frege denies that 'This cur howled the whole night' expresses the further thought that the speaker has such an attitude, because the absence of the attitude would not falsify the sentence.[14]

It might be objected to Frege's account that if the words 'dog' and 'cur' have the same sense, then, on Frege's own account of propositional attitude ascriptions, the sentences 'Mary believes that every dog is a dog' and 'Mary believes that every dog is a cur' must have the same truth-value, however much Mary loves dogs. Mary, a fully competent speaker of English, assents to 'Every dog is a dog'; will she assent to 'Every dog is a cur'? If she agrees that 'Every dog is a cur' is true but misleading, we can surely agree that 'Mary believes that every dog is a cur' is also true but misleading. What if Mary, under the influence of too much inferentialist reading, claims that 'Every dog is a cur' lacks a truth-value in virtue of its badly matched introduction and elimination rules? If she believes that every dog is a cur, why does she not assent to 'Every dog is a cur', for surely she knows that it expresses her belief that every dog is a cur? Such problems undermine Frege's simple account of propositional attitude ascriptions, not his claim that pairs like 'cur' and 'dog' have the same truth-conditional meaning. For similar problems arise even for pairs of synonyms with the same tone. Kripke (1979) gives the example of the synonymous natural kind terms 'furze' and 'gorse'. A speaker might learn the two words on different occasions, from ostension of examples, and acquire normal competence with both without being sure that they refer to exactly the same kind of plant. Thus 'He believes that all furze is furze' and 'He believes that all furze is gorse' appear to differ in truth-value. It does not follow that 'furze' and 'gorse' are not synonyms after all. Such problems show nothing special about pejoratives. Whatever the right account of propositional attitude ascriptions, it is compatible with the Fregean view that 'cur' and 'dog' differ in tone but not sense (truth-conditional meaning).

Unfortunately, Frege's category of tone is too miscellaneous to take us very far in the analysis of the example. It includes stylistic differences, such as that between 'sweat' and 'perspire', which are significantly unlike the difference between 'Boche' and 'German'. Frege

discusses his examples by speaking unhelpfully of the images and feelings that the words evoke in hearers. Nevertheless, the classification of the difference between 'Boche' and 'German' as truth-conditionally irrelevant is at least a useful start, for it respects intuitive distinctions over which the original inferentialist account rides roughshod.

What needs explaining is this. Competent English speakers know, or are in a position to know, that 'German' and 'Boche' have the same reference, and therefore that 'Lessing was German' and 'Lessing was Boche' have the same truth-value. If educated, such speakers know, or are in a position to know, that both sentences are true. Nevertheless, although such speakers are willing to assert 'Lessing was German', they are not willing to assert 'Lessing was Boche', even on reflection, unless they are xenophobes. I know that 'Lessing was Boche' is true, but I refuse to assert 'Lessing was Boche'. Why?

The natural answer is that to assert 'Lessing was Boche' would be to *imply* that Germans are cruel, and I do not want to imply that, because the implication is both false and abusive. Since the false implication that Germans are cruel does not falsify 'Lessing was Boche', it is not a logical consequence of 'Lessing was Boche'. Rather, in Grice's terminology, 'Lessing was Boche' has the *conventional implicature* that Germans are cruel, in much the same way that 'Helen is polite but honest' has the conventional implicature that there is a contrast between Helen's being polite and her being honest.[15] Just as 'Lessing was Boche' and 'Lessing was German' differ in conventional implicatures while being truth-conditionally equivalent, so too 'Helen is polite but honest' and 'Helen is polite and honest' differ in conventional implicatures while being truth-conditionally equivalent. In Grice's terms, conventional implicatures are *detachable*, because they can differ between truth-conditionally equivalent sentences. But they are not easily *cancellable*, for someone who says 'Lessing was Boche, although I do not mean to imply that Germans are cruel' merely adds

hypocrisy to xenophobia; equally deviant would be an utterance of 'Helen is polite but honest, although I do not mean to imply that there is any contrast between her being polite and her being honest'. By contrast, Gricean conversational implicatures are easily cancellable but not detachable. Whether one says 'She is either in Paris or Rome' or something truth-conditionally equivalent to it, the maxims of conversation generate the implicature that one does not know which city she is in; nevertheless, one can easily cancel it by adding 'I know which city she is in, but I am not going to tell you'. The implicature that 'Boche' carries must be conventional rather than purely conversational, for if there were not already a significant difference between the words 'Boche' and 'German', the norms of conversation could not generate any difference in implicature between 'Lessing was Boche' and 'Lessing was German'.

The conventional implicatures of 'Boche' and 'but' are preserved under embedding in more complex sentences. For example, 'If Lessing was Boche then he was European' still conventionally implicates that Germans are cruel, and 'If Helen is polite but honest then she is honest' still conventionally implicates that there is a contrast between her being polite and her being honest. The implicatures are present just as strongly in non-indicative sentences, such as 'Is he Boche?', 'Translate this into Boche!', 'Is she polite but honest?' and 'Be polite but honest!'. In this respect, such conventional implicatures are like presuppositions rather than logical consequences. However, the relevant sense of 'presupposition' is closer to Stalnaker's than to Strawson's: although a presupposition modifies the context of utterance, its failure does not deprive the sentence of a truth-value.[16] In the case of 'Boche', the very use of the word generates the xenophobic implicature, irrespective of its position in the sentence.[17]

The false implicature of 'Lessing was Boche' is not merely that Lessing was cruel. For

otherwise the corresponding implicature of 'Hitler was Boche' would merely be that Hitler was cruel, which is commonly known to be true. Then I could decently assert 'Hitler was Boche': but I cannot. Thus even a singular ascription of 'Boche' carries the false general implicature that Germans are cruel. But since xenophobes treat Boche-Elimination as a defeasible rule, the implicature is not that all Germans without exception are cruel. The implicature is expressed by 'Germans are cruel' read as a generic sentence, meaning something like 'There is a tendency for Germans to be cruel'.

The implicature that 'Boche' carries is not merely about the speaker's psychological state. In particular, what is implicated is not merely that the speaker *believes* that Germans are cruel; such an implicature might well be true. When someone uses the word 'Boche' one can legitimately ask him to withdraw the anti-German implication; but if it is clear, as it may well be, that he does believe that Germans are cruel, then it is hardly legitimate to ask him to withdraw the implication that he has that belief. Although one might try to persuade him to abandon the belief, and even succeed, the view at issue makes the implicature of his original remark be that he had the belief at the time of utterance, not that he has it now, which leaves no false implicature to withdraw. The false implicature is that Germans are cruel, not that the speaker believes that they are cruel. Perhaps the use of 'Boche' does also carry the additional implicature that the speaker believes that Germans are cruel, since a linguistically competent speaker who uses 'Boche' without believing that Germans are cruel is being insincere; but such a belief condition would be a byproduct of the simple implicature that Germans are cruel, combined with the conversational norm of sincerity; it is not the source of what is most objectionable in the use of 'Boche'.

One might argue that the use of 'Boche' implies more than that Germans are cruel, namely, that the speaker knows that Germans are cruel, or even that it is common knowledge

that Germans are cruel. In asserting *p*, one implies in some sense that one knows *p*, although of course one does not thereby assert that one knows *p*, for one is in an epistemic position to assert *p* only if one knows *p*. Might implying *p* similarly generate the further implication that one is in an epistemic position to imply *p*? Such putative implications are not generated directly by the use of the pejorative word, but at best indirectly by the application of general principles of conversation to the direct implication; they will not be further discussed here.[18]

The conventional implicatures that words such as 'but' and 'Boche' generate are part of their meaning, in a broad sense of 'meaning'. If one is ignorant of them, one is at least partially ignorant of the meaning. An Englishman in Italy who thinks that 'ma' is synonymous with 'and' and 'e' with 'but' is mistaken, for 'e' is synonymous with 'and', not with 'but', and 'ma' is synonymous with 'but', not with 'and'. Fully to understand a word, one must have some awareness, however inexplicit, of the conventional implicatures that it generates. In the case of 'Boche', one might say, in Putnam's terminology, that cruelty is part of its associated *stereotype*; a stereotypical Boche is cruel. Putnam allows that stereotypes may be inaccurate; perhaps ferocity is part of the stereotype associated with the natural kind term 'gorilla', although really gorillas are gentle. On his view, the stereotype for a word plays no direct role in determining its reference, but to be competent with the word one must have the stereotype (1975: 247-52). Since a competent speaker may know that the stereotype is inaccurate, to have the stereotype is not to believe that it is accurate; what one must be aware of is that it is the stereotype. Someone who understands 'Boche' may know that cruelty is an inaccurate part of the associated stereotype. The exact relation between conventional implicatures and stereotypes deserves further investigation, but we have a clear enough view for present purposes. What is most crucial is the separation of those aspects of meaning that contribute to truth-conditions from those that do not.[19]

'Boche' and 'German' have the same reference, so 'The Boche are the Germans' is true, but in using 'Boche' one implies that Germans are cruel. One can fully understand the word 'Boche' and know all that without being committed to the claim that Germans are cruel, for one can refuse to use the word 'Boche'. One is not obliged to utter every sentence that one knows to be true. One can know that a rule of inference is truth-preserving without using it. The inferentialist accounts of pejoratives in Dummett, Brandom and Boghossian misconstrue the linguistic data. They lump together a deductive inference (Boche-Introduction) and a mere conventional implicature (Boche-Elimination) as if they played equal roles in the use of the term; if they predict anything about the reference of pejoratives, they falsely predict reference failure; by making willingness to use the roles a precondition for understanding they falsely imply that only those with the relevant prejudices can understand pejoratives.

Could one base a better inferentialist account of 'Boche' on the truth-preserving rules Boche-Introduction and Boche-Elimination*? On such an account, inferring according to those rules would be a precondition for understanding 'Boche'. But that too is false, just as the previous section would lead us to expect. Non-xenophobic speakers of English may acknowledge that the rules are truth-preserving yet still refuse to infer according to them. For one must *use* 'Boche' to infer according to those rules, whereas to classify them as truth-preserving is only to *mention* the word. Again, a philosopher of language may understand the word 'Boche' as well as the rest of us do, but misconstrue the nature of pejoratives and reach the false conclusion that no instance of Boche-Introduction is truth-preserving, on the mistaken grounds that 'Boche' suffers reference failure. His false philosophical theory does not cause him to forget the meaning of the word 'Boche'.

A further problem for any inferentialist account of 'Boche' that relies on its inferential links with 'German' is that someone might understand 'Boche' without understanding

'German' or any other non-pejorative word with the same reference. One might grow up in a narrow-minded community with only pejorative words for some things. One would articulate the relevant conventional implicature by saying 'Boche people are cruel'.

Pejoratives pose a quite general problem for use theories of understanding on which using a term in a given way is a precondition for understanding it. Unprejudiced speakers may understand a pejorative term but still refuse to use it in the specified way: in order to avoid commitment to a conventional implicature, they refuse to use it at all.

In one respect the foregoing result may be good news for Boghossian, since it removes a class of putative counterexamples to the unqualified meaning-entitlement connection (MEC). The consequent danger for (MEC) is not that it is false but that it is vacuously true, and therefore cannot help to explain our inferential entitlements.[20]

Pejoratives raise a further question about the nature of concepts. A vital feature of a pejorative *word* is that it carries a conventional implicature, but how could a pejorative *concept* do so? Conventional implicatures seem to arise from the communicative use of language. Although synonymous words in different languages carry the same conventional implicatures, how could they share those implicatures with a non-verbal concept? In the study of pejoratives, it is dangerous to give thought methodological priority over language by treating concepts as intrinsically non-linguistic yet still capable of doubling as linguistic meanings. For that treatment is appropriate only if the false pejorative implicature is a feature of the concept *Boche*; but then it is unclear how one could have the concept other than by understanding a word synonymous with 'Boche'. On such a view, the notion of having a concept seems to be parasitic on the notion of understanding a word, and language takes methodological priority over thought. If, on the other hand, the false implicature is not a feature of the concept *Boche*, then presumably *Boche* is the same concept as *German*: since

'Boche' is not synonymous with 'German', the meaning of a word is therefore not exhausted by the concept that it means, and the study of meanings cannot be subsumed under the study of concepts.[21] An adequate theory of concepts must resolve such questions.[22]

III

Stipulated Possession Conditions. One might wonder why Boghossian could not respond to the argument of the previous section by simply considering an artificial concept *Quoche*, stipulated to satisfy his account of *Boche*: one has it if and only if one is willing to infer according to rules like Boche-Introduction and Boche-Elimination (with *Quoche* in place of *Boche*). Boghossian makes a similar move when he considers modifications of the meaning-entitlement connection (MEC) on which the inferential transitions built into concept possession are required to be truth-preserving, or even knowable *a priori* to be truth-preserving on some assignment of reference. If the original version of (MEC) is vacuously true, then so *a fortiori* are these restricted versions. But Boghossian constructs putative counterexamples to them involving artificial concepts with introduction and elimination rules that together generate consequences to which merely having the concepts does not *eo ipso* entitle one. In effect, he stipulates that inferring according to the rules is a precondition for having the concepts.

How far can one go in stipulating possession conditions for concepts? Perhaps not all the way, according to Boghossian, for he allows that 'tonk' may express no concept. Consider this stipulation for an artificial concept *mansquare*:

MANSQUARE One has the concept *mansquare* if and only if one is male and has the concept *square*.

It seems clear that there is no such concept; MANSQUARE is as futile as the stipulation that one is immortal. Suppose that a man tries to introduce a word 'mansquare' to mean *mansquare*. He starts using 'mansquare' just as he uses 'square'. When the rest of us try to understand him, being a man gives one no special advantage, even if he insists that it does. Whatever is to be known about the meaning of 'mansquare', women can know it just as well as men.

Is the problem with MANSQUARE that being male is not an inferential condition? Consider this stipulation for an artificial concept *comsquare*:

COMSQUARE One has the concept *comsquare* if and only if one can prove the completeness of the propositional calculus and has the concept *square*.

Although the new conjunct is inferential, that does not help. COMSQUARE seems as futile as MANSQUARE. For although I already meet the condition that I can prove the completeness of the propositional calculus and have the concept *square*, that seems to give me no extra concept *comsquare*.

One might think that such doubts about the power of stipulation are out of place when one considers the possession conditions that Boghossian proposes, which simply involve willingness to make various inferences with the putative concept itself. However, the underlying difficulty from section I remains. Suppose, for reductio, that one has a concept C if and only if one is willing to infer according to a deductive pattern P in which C occurs, and

that P leads to inferences to which one is not *eo ipso* entitled. The community uses a word 'C' to mean C. One member of the community understands 'C' in the usual way, but then comes to realize that P leads to inferences to which one is not *eo ipso* entitled. She is therefore no longer willing to infer according to P. Before the change, she knew that 'C' means C. Her warranted rejection of P does not cause her to lose the knowledge that 'C' means C. She still knows that 'C' means C, so she still has the concept C. Thus the original supposition that one has C if and only if one is willing to infer according to P is false; it led to absurdity, given that P really does lead to inferences to which one is not *eo ipso* entitled.

The trouble is that having the concept is necessary for understanding the word (with its actual meaning), understanding the word is necessary for knowing what it means, and willingness to reason according to an objectionable pattern is not necessary for knowing what the word means. Thus willingness to reason according to the pattern is not necessary for having the concept. For all that has been said, someone could have a concept and reason with it according to the rules that Boghossian specifies. The point is that, if so, then someone else could have the same concept and not reason with it according to those rules. From section I we know that the problem generalizes even to cases in which the pattern of reasoning is not really objectionable but is taken to be so. At any rate, one cannot simply take an objectionable pattern of inference and define a concept as the one for having which it is necessary and sufficient to be willing to infer according to the pattern. There may be no such concept. That is why artificial stipulations are no substitute for accurate descriptions of concepts that we actually possess.

Could an inferentialist respond by pulling concept possession much further apart from linguistic understanding? The proposal might be that different thinkers can understand the same unambiguous word in virtue of using it to express different concepts, which are equated

with different inferential roles. For example, neo-Fregeans such as Gareth Evans (1982: 40) sometimes claim that different speakers can achieve linguistic competence with the same proper name by associating it with different concepts (modes of presentation) of the same object. The distinctions between inferential roles would thus cut finer than the distinctions between linguistic meanings. On this view, when our unorthodox thinker refuses to infer any longer according to P, she associates the word 'C' with a new concept, because she changes the associated inferential role, while retaining the same meaning for the same word. Hence we cannot properly give the meaning of 'C' by saying that it is used to express the concept C, for the association with C is not essential to that meaning. Thus the word 'C' can hardly be said to *express* the concept C even in the mouth of a speaker who does in fact associate 'C' with C.

Large methodological questions arise for the view that concepts are individuated much more finely than linguistic meanings. For example: to which concept does the phrase 'the concept *square*' refer if the word 'square', with its usual meaning in English, is associated with different concepts in the minds of different speakers of English at one time, or in the mind of the same speaker at different times? Presumably, in writing 'the concept *square*' one intends to refer to the concept that one associates with the word 'square' at the time of writing, but what happens if the reader associates a different concept with the word? Without the use of phrases like 'the concept *square*', accounts of concept possession would lose most of their examples. Since much of the discussion is anchored to examples presented in just that way, their loss would cast it adrift. An account full of phrases like 'my present concept *square*' would be disappointingly autobiographical.

The inferentialist might reply that the problem will be manageable if most speakers do in fact reliably associate the same concept with a given word. Since there are likely to be at

least minor differences between any two speakers in their willingness to accept inferences that involve the word, some principle is needed to distinguish those patterns of inference that are essential to the concept from those that are not. The meaning of the word supplies no such principle, on the envisaged view of concepts. It is deeply unclear whence the required principle is to come. Without such a principle, discussion of concepts becomes dangerously unconstrained.

This is no place to engage properly in the vast debate over the relative methodological priority of thought and language. Fortunately, Boghossian does not take refuge in the envisaged separation of the two sides. He assumes that words express concepts; his argument moves freely between concept possession and linguistic understanding. This paper will continue to follow that methodology.

IV

Conditional and Unconditional Concepts. Boghossian classifies concepts like *Boche* as defective. He introduces his positive account of the difference between defective and non-defective concepts with the example of the concept *neutrino*, as expressed by the word 'neutrino'. Let 'T(Neutrino)' abbreviate neutrino theory. In the tradition of Ramsey (1929), Carnap (1966: 269-72), Lewis (1970) and, more immediately, Horwich (2000), Boghossian divides T(Neutrino) into two components:

(S) $\exists F\, T(F)$

(M) $\exists F\, T(F) \rightarrow T(\text{Neutrino})$

T(Neutrino) is logically equivalent to the conjunction of the Ramsey sentence (S) and the Carnap sentence (M). According to Boghossian, our concept *neutrino* is *conditionalized*, in the sense that to have it one must assent to (M) but need not assent to (S), the latter being no logical consequence of the former.[23] Thus one can have our concept *neutrino* while still denying neutrino theory, because one denies (S): one agrees that if any things play the role that neutrino theory specifies for neutrinos then neutrinos do, but holds that no things play that role. Consequently, our concept *neutrino* leaves room for substantive scientific debate on whether any things do play the neutrino role; it is an epistemically non-defective concept. Boghossian allows that there is a correspondingly defective concept *neutrino+* which prejudges the scientific issue; it is unconditionalized in the sense that to have it one must accept both (S) and (M). The concept *neutrino+* is supposed to be the approximate analogue for physics of the concept *Boche*.

Boghossian relates entitlement, defectiveness and conditionalization by means of several theses. We may slightly amplify his account thus:

(Bold) If a conditionalized version of an unconditionalized concept is available, then the unconditionalized concept is defective.

(Sober) For some concepts no conditionalized version is available.

(Bold-C) If no conditionalized version of an unconditionalized concept is available, then the unconditionalized concept is non-defective.

(Cond) Any conditionalized concept is non-defective.

(MEC*) Any inferential transitions built into the possession conditions for a non-defective concept are *eo ipso* entitling.

Boghossian provides the obvious defence for (Bold): whenever possible, we should avoid prejudging factual questions; therefore, given the choice between an unconditionalized concept and a conditionalized version, we should use the latter. For (Sober), he argues that we cannot conditionalize away the very logical concepts (the conditional →, the quantifier ∃) that we must use in defining conditionalized concepts, otherwise, presumably, we should have nowhere to start. Boghossian states the principle here labelled '(MEC*)' as the natural modification of (MEC) within this framework to handle the objection from defective concepts. We should not expect the inferential transitions built into the possession conditions for a defective, unconditionalized concept to be *eo ipso* entitling, because they may prejudge issues that we do not have to prejudge. Boghossian does not explicitly state (Bold-C), which is in effect the converse of (Bold), but he seems to assume it when, having argued that no conditionalized version of the unconditionalized concept of the conditional is available, without further ado he applies (MEC*) to conclude that we are entitled to infer by modus ponens, which is built into the unconditionalized concept of the conditional. Thus he treats the move from the unavailability of a conditionalized version of the unconditionalized concept to the non-defectiveness of the unconditionalized concept as immediate; the principle on which he thereby seems to rely is explicit in (Bold-C). This is in effect to assume that the only relevant defect in a concept is unforced lack of conditionalization. Thus a conditionalized concept is never defective, so (Cond) also holds, although again Boghossian

does not make it explicit.

Is Boghossian correct in claiming that willingness to affirm the Carnap sentence (M) is necessary for having the concept *neutrino*? Doubts arise when one asks how neutrino theory, T(Neutrino), is to be demarcated. Presumably, not every small change in what physicists believe about the referent of 'neutrino' counts as their adopting a new concept *neutrino*; the possession condition for the concept associated with 'neutrino' remains constant through at least some such changes. Thus T(Neutrino) consists of some but not all of physicists' beliefs about neutrinos. Which are the privileged beliefs? Whatever the answer, an expert physicist might reject it on subtly mistaken theoretical grounds. The physicist might reason thus: 'It is crucial to the role of neutrinos that our current theory T*(Neutrino) hold. But T(Neutrino) is too weak to entail T*(Neutrino). If T(F) is true but T*(F) false, then the Fs are not neutrinos. Perhaps T(F) is true for some Fs while T*(F) is false for all Fs; then there are no neutrinos, so $\exists$F T(F) → T(Neutrino) is false. Therefore, I had better not commit myself to that conditional'. Thus the physicist explicitly refuses to affirm the Carnap sentence (M). Although his reasoning is by hypothesis unsound, it does not seem to show that the physicist lacks the concept *neutrino*, contrary to Boghossian's account. The physicist may be an acknowledged world authority on neutrinos, having played a leading role in the development of current neutrino theory, while being less than fully convinced of its truth. All his colleagues regard it as obvious that he understands the word 'neutrino'. His mistake about the Carnap sentence does not impinge on his work in physics. Perhaps circumstances never arise in which physicists agree that $\exists$F T(F) is true and $\exists$F T*(F) false, so they never have to decide whether there are neutrinos in such circumstances. Even if those circumstances do arise, and he thinks that it has turned out that there are no neutrinos while other physicists do not, he may still decide to assent to the sentence 'There are neutrinos' for purposes of

communication while insisting that he does so in a new sense of 'neutrino'. That still does not seem to show that he previously failed to understand the word 'neutrino'. His unorthodoxy at the meta-linguistic level does not impair the fruitfulness of his interactions with other physicists. They may all regard the meta-linguistic issue as a trivial matter of terminology. Thus considered refusal to affirm the Carnap sentence for neutrino theory seems to be quite compatible with full understanding of the word 'neutrino' as used by physicists, and with full possession of the concept *neutrino*.[24]

If one tries to avoid that problem by building more into the theory T(Neutrino) for the purposes of the Carnap sentence, new problems arise. For example, since the strengthened T(Neutrino) will be more controversial, some expert physicists may suspect that some neutrinos satisfy it in some states (making $\exists F\, T(F)$ true) but not in others (making T(Neutrino) false); on those grounds, they refuse to affirm the Carnap sentence. As before, it does not seem to follow that those physicists lack the concept *neutrino* or fail to understand the word.

Such problems suggest that concept constitution accounts of blameless inference rest on mistaken expectations of a theory of concept possession or linguistic understanding. They seek a conceptual shibboleth, an inference or principle acceptance of which is necessary and sufficient for knowing what a word means or possessing a given concept. Yet understanding words in a natural language has much to do with being able to use them in ways that facilitate smooth and fruitful interaction with other members of the community. That ability can be realized in indefinitely various forms. Speakers can compensate for their deviance on one point by their orthodoxy on others, their ability to predict the reactions of non-deviant speakers, their willingness in the long run to have their utterances evaluated by public standards. We have seen that such compensation is often possible when the deviance results

from localized interference in the normal practice of using a word by high-level theoretical concerns. On that picture, there is no litmus test for understanding. Whatever local test is proposed, someone could fail it and still do well enough elsewhere with the word to be counted as knowing what it means. If linguistic understanding is linked to concept possession in the standard way, so that fully understanding the word 'neutrino' as used in English is sufficient for having the concept *neutrino*, then there is also no litmus test for concept possession. Variety in use amongst those who understand 'neutrino' is *ipso facto* variety in use amongst those who have the concept *neutrino*.

Could an inferentialist reply that such objections trade on an everyday sense of 'understanding' that must be replaced by something more precise for theoretical purposes? It is far from clear that the inferentialist has a better alternative. The relevant features of the ordinary notion of understanding are not mere untheoretical sloppiness. Rather, they look like an appropriate response to an important constraint on a theory of concepts or linguistic meanings: that there is little point in talking about them unless they can be shared across differences in belief, between different individuals at the same time or the same individual at different times. They can survive factual learning and factual disagreement. Although inferentialist accounts respect the letter of that constraint, they violate its underlying spirit, by setting inflexible limits to the scope for genuine disagreement. The more holistic ordinary notion of understanding permits localized disagreement at virtually any point.

The cases just discussed hint at ways in which the failure of individualist accounts of meaning may go deeper than the immediate lessons of the famous anti-individualist arguments of Putnam (1975) and Burge (1979). Their cases are often analysed in terms of a distinction between experts with full understanding and lay-people with partial understanding who defer to the experts, in virtue of which one may correctly ascribe to them attitudes to the

contents that the experts determine.[25] But we have seen that experts themselves may make deviant uses of words as a result of theoretical errors and still count as fully understanding their words. Although they defer to nobody on the matters at issue, they are more than adequately integrated members of the speech community with respect to those very words. Their assignments of meaning to those words are not parasitic on the assignments that more privileged individuals make. Rather, each individual uses words as words of a public language; their meanings are determined not individually but socially, in virtue of the spectrum of linguistic activity across the community as a whole. The social determination of meaning requires nothing like exact match in use between different individuals; it requires only enough connection in use between them to form a social practice. Full participation in that practice constitutes full understanding. Consequently, there is no litmus test for understanding. Although this does not exclude the possibility of concept possession by an isolated individual, the link from linguistic understanding to concept possession precludes a litmus test for concept possession too.[26]

It is useful to come at these matters from several angles. The next two sections assess Boghossian's positive account in more detail.

V

Unique Characterizations and Unique Realizations. When Belnap (1962) discussed the desiderata for good rules for a new concept, he did not merely require the rules to yield a conservative extension of the old theory. He also required them to characterize the new concept uniquely, in the sense that any two expressions subject to the rules are provably

equivalent to each other. Inferentialists want the implicit definition to be satisfied by at least one concept; they also want it to be satisfied by at most one concept. Standard introduction and elimination rules for the usual logical connectives meet this unique characterization requirement (Harris (1982)). For example, if one introduces two connectives $\rightarrow_1$ and $\rightarrow_2$, both subject to conditional proof and modus ponens, one can easily use those rules to derive $p \rightarrow_1 q$ from $p \rightarrow_2 q$ and *vice versa*. If some pattern P of inferences does not provide unique characterization, then one can use both the expressions C_1 and C_2 according to P without being obliged to treat C_1 and C_2 as equivalent; thus P appears to single out no unique concept.[27]

For the defective, unconditionalized concept *neutrino+*, the governing rules are supposed to be tantamount to full neutrino theory, which is equivalent to the conjunction of (S) and (M). Thus the requirement of unique characterization is that when we introduce two predicates, 'Neutrino_1' and 'Neutrino_2', both subject to full neutrino theory, they should be provably equivalent. In other words, this argument should be valid:[28]

(!+A) $\quad \text{T}(\text{Neutrino}_1)$

$\quad \text{T}(\text{Neutrino}_2)$

$\quad$ ---

$\quad \forall x(\text{Neutrino}_1(x) \leftrightarrow \text{Neutrino}_2(x))$

The relevant notion of validity is broadly logical and independent of neutrino theory, which has instead been packed into the premises. Since the only constraints on the new predicates 'Neutrino_1' and 'Neutrino_2' are the premises of (!+A), it is valid if and only if this sentence is a truth of second-order logic:

(!+) $\forall G \forall H((T(G) \,\&\, T(H)) \rightarrow \forall x(Gx \leftrightarrow Hx))$

Roughly, (!+) says that at most one lot of things play the role specified for neutrinos: neutrino theory is uniquely realized. Let us provisionally assume that neutrino theory is indeed so strong that (!+) is logically true.

For our non-defective, conditionalized concept *neutrino*, the governing rules are tantamount just to (M). Thus the requirement of unique characterization is that, when we introduce two new predicates 'Neutrino$_1$' and Neutrino$_2$', this argument should be valid:

(!A) $\exists F\, T(F) \rightarrow T(\text{Neutrino}_1)$

$\exists F\, T(F) \rightarrow T(\text{Neutrino}_2)$

$\forall x(\text{Neutrino}_1(x) \leftrightarrow \text{Neutrino}_2(x))$

Just as (!+A) is valid if and only if (!+) is a truth of second-order logic, so (!A) is valid if and only if this sentence is a truth of second-order logic:

(!) $\forall G \forall H(((\exists F\, T(F) \rightarrow T(G)) \,\&\, (\exists F\, T(F) \rightarrow T(H))) \rightarrow \forall x(Gx \leftrightarrow Hx))$

But we can easily show that (!) is logically true only if the Ramsey sentence (S) of neutrino theory is logically true, which it is not: for all logic says, no things behave as neutrino theory says the neutrinos behave (otherwise the conditionalized concept would collapse into the unconditionalized one).[29] Thus Boghossian's conditionalized rule for 'neutrino' fails Belnap's unique characterization requirement. The problem is simple: the supposed possession

condition for the conditionalized concept *neutrino* involves only the Carnap sentence (M), which says nothing about which things are neutrinos if the Ramsey sentence is false. Each of the following is consistent with (M), and is therefore left open by Boghossian's possession condition for our concept *neutrino*: (S) is false and the neutrinos are the philosophers; (S) is false and the neutrinos are the non-philosophers; (S) is false and everything whatsoever is a neutrino. The problem obviously generalizes from 'neutrino' to other conditionalized terms: Boghossian's rules for them fail to achieve unique characterization.

David Lewis addresses related issues in his classic exposition (1970) of how to define theoretical terms using the Ramsey-Carnap method. Treating theoretical terms as names, he holds that they should name something only if the relevant theory has a realization. For present purposes, we can take that as the requirement that the predicate 'Neutrino' should be nonvacuous (apply to something) only if neutrino theory has a realization. Thus, if there are neutrinos, then some things behave as neutrino theory says the neutrinos behave:

(1) $\exists x\ \text{Neutrino}(x) \rightarrow \exists F\ T(F)$

Since the consequent of (1), the Ramsey sentence, is a logical consequence of T(Neutrino), so is (1) itself. If we make acceptance of (1) as well as Boghossian's (M) the possession condition for a modified conditionalized concept *neutrino-*, then unique characterization holds for the latter concept if and only if it holds for the unconditionalized concept, as it was assumed to do in the case of the concept *neutrino+*.[30] Together, (1) and (M) entail that if there are neutrinos, then they behave as neutrino theory says the neutrinos behave, so neutrino theory is true:

(2) $\exists x\ \text{Neutrino}(x) \rightarrow \text{T}(\text{Neutrino})$

If neutrino theory entails that there are neutrinos, as it presumably does, then the conditional in (2) can be strengthened to a biconditional:

(3) $\exists x\ \text{Neutrino}(x) \leftrightarrow \text{T}(\text{Neutrino})$

Neutrino theory is true if and only if there are neutrinos. The addition of (1) to (M) in the possession condition for our modified conditionalized concept *neutrino-* does not obviously violate the spirit of Boghossian's account. One might hope that the possession condition for a non-defective concept would uniquely determine its reference across all logically possible cases.

Further modifications of Boghossian's account are needed to align it fully with what Lewis says about theoretical terms. For Lewis also requires that there are neutrinos only if neutrino theory has at most one realization:

(4) $\exists x\ \text{Neutrino}(x) \rightarrow \forall \text{G} \forall \text{H}((\text{T}(\text{G}) \ \& \ \text{T}(\text{H})) \rightarrow \forall x(\text{G}x \leftrightarrow \text{H}x))$

For if neutrino theory has several realizations, which of them determines the extension of 'Neutrino'? Without something like (4), the Ramsey-Lewis method does not enable one to *define* theoretical terms. By contrast, the Carnap sentence (M) requires that if neutrino theory has multiple realizations, and therefore at least one, then neutrino theory is true, and therefore by (3) that there are neutrinos. Earlier, we assumed that (!+) is logically true; on that assumption, neutrino theory is logically incapable of multiple realizations, so the further

difference between Lewis and Boghossian does not arise. However, even if unique realization is plausible for neutrino theory, we can hardly expect it to hold in all cases; the relevant theory may simply be too weak. Applications of the Ramsey-Lewis method often fail to make unique realization plausible.[31] Since we are using 'neutrino' as a representative case, we had better drop the assumption that (!+) is logically true. Thus even the unconditionalized concept *neutrino+* may fail the unique characterization requirement. But then so too may the modified conditionalized concept *neutrino-*, for we noted above that the concept *neutrino-* satisfies unique characterization if and only if the concept *neutrino+* does. If we want a non-defective concept which satisfies unique characterization, we must in general modify the possession condition again. The natural move is to follow Lewis in weakening the Carnap sentence (M) to require only that neutrino theory is true if it has exactly one realization:

(M*) $(\exists F\, T(F)\ \&\ \forall G \forall H((T(G)\ \&\ T(H)) \rightarrow \forall x(Gx \leftrightarrow Hx))) \rightarrow T(\text{Neutrino})$

The principles (M*), (1) and (4) correspond to what Lewis calls his three 'meaning postulates' for theoretical terms. The corresponding possession condition for our concept *neutrino* would require acceptance of (M*), (1) and (4). This possession condition too secures unique characterization.[32]

The analogues of (1) and (M*) for other concepts are always logical consequences of the relevant theory, the analogue of T(Neutrino), but the analogue of (4) is often not, for the relevant theory may not entail that it has a unique realization. Even if it does in fact have a unique realization, that may not be guaranteed by the form of the theory itself. Nevertheless, the use of the theory in fixing the reference of the term in question might still be held to involve an implicit commitment to the analogue of (4).

A further complication is that, when Boghossian states introduction and elimination rules for a conditionalized concept, they are equivalent neither to the original Carnap sentence (M) nor to the Lewis-like meaning postulates (1), (4) and (M*). Boghossian's rules have this form:[33]

Neutrino-Introduction	Neutrino-Elimination
$\exists F(T(F) \,\&\, Fx)$	Neutrino(x)
---	---
Neutrino(x)	$\exists F(T(F) \,\&\, Fx)$

Evidently, these rules are equivalent to an explicit definition of Neutrino(x) as $\exists F(T(F) \,\&\, Fx)$. Automatically, therefore, the extension is conservative and the characterization unique. In effect, 'Neutrino' is defined as the disjunction of the realizers of neutrino theory. But, in cases of multiple realization, the disjunction of the realizers of a theory need not itself be a realizer of that theory. Consequently, there is no general guarantee independent of the specific structure of neutrino theory that the Carnap sentence (M) is derivable from Neutrino-Introduction and Neutrino-Elimination.[34] For other theories T, when we implicitly define a conditionalized concept by means of rules analogous to Neutrino-Introduction and Neutrino-Elimination, the corresponding Carnap sentence is false. Without recourse to the structure of T, we can derive only the weakening (M*) of (M) from Neutrino-Introduction and Neutrino-Elimination (in classical second-order logic).[35] Although we can regain (M) from (M*) given the auxiliary assumption that T has at most one realization, that assumption is not generally true. Conversely, without recourse to the structure of T, we cannot derive either Neutrino-

Introduction or Neutrino-Elimination from (M). For other theories T and terms C, the Carnap sentence $\exists F\, T(F) \rightarrow T(C)$ is true but the analogues of Neutrino-Introduction (from $\exists F(T(F) \,\&\, Fx)$ infer $C(x)$) and Neutrino-Elimination (from $C(x)$ infer $\exists F(T(F) \,\&\, Fx)$) are invalid.[36] The special auxiliary assumption that T has at most one realization would enable us to derive the introduction rule from the Carnap sentence; to derive the elimination rule from the Carnap sentence we need instead the Lewis postulate analogous to (1), that C applies to something only if it realizes T.

How are Boghossian's introduction and elimination rules related to the three Lewis meaning postulates (M*), (1) and (4)? As already noted, (M*) is always derivable from the introduction and elimination rules. Moreover, the elimination rule easily yields (1). However, the remaining postulate (4) cannot be derived from the introduction and elimination rules without recourse to the structure of the theory. For other theories T, when we implicitly define a conditionalized concept by means of rules analogous to Neutrino-Introduction and Neutrino-Elimination, the analogue of (4) is false.[37] In the converse direction, Boghossian's elimination rule is always derivable from the three Lewis postulates.[38] However, his introduction rule is not derivable from those postulates without recourse to the structure of the theory. For other theories T and terms C, the analogues of (M*), (1) and (4) are true but the analogue of Neutrino-Introduction is invalid.[39] These failures of equivalence are hardly surprising. The Lewis meaning postulates and Boghossian's introduction and elimination rules embody incompatible treatments of multiple realization, in case of which the former make the new concept apply nowhere while the latter make it apply wherever at least one of the realizations applies. The two approaches coincide given the special auxiliary assumption that the theory has at most one realization.

Despite their complicated differences, the various strategies considered in this section

all remain within the general spirit of Boghossian's approach. Nevertheless, the complexity itself reinforces earlier doubts about that approach in at least two ways.

First, it is far from obvious that expert physicists who understand 'neutrino' as well as anyone does will know what to make of complex, highly abstract principles such as (1), (4) and (M*) or Neutrino-Introduction and Neutrino-Elimination. Probably some physicists will assent to them once the symbols are explained while others will not. The reason for the failure to assent will not be failure to grasp their own concept *neutrino* (perhaps it is simply that they lack training in philosophy). Although one might maintain that the physicists are implicitly committed by their practice to (1), (4) and (M*) or Neutrino-Introduction and Neutrino-Elimination, that would not vindicate Boghossian's account, which formulates possession conditions in terms of positive affirmation rather than mere commitment.

The second problem arises with respect to people who are experts on both neutrinos and concept possession. Consider such a prodigy who, understanding neutrino theory and the various meaning postulates and rules as well as anyone does, nevertheless makes a subtle mistake in the theory of concept possession and adopts an account with deviant consequences for the case of multiple realization. Consequently, our expert consciously rejects some meaning postulate or inference rule assent to which is a precondition, on some account, for understanding 'neutrino' as used in the rest of the scientific community. Nevertheless, when talking physics he continues to use the word in a way that other physicists regard as normal, because they never discuss the possibility of multiple realization. The claim that he does not fully understand the word 'neutrino' as used by others is implausible, and quite at variance with ordinary standards for linguistic competence. The natural description of our expert is that he fully understands the word 'neutrino' in the normal way while holding false theoretical beliefs about its meaning. Therefore the account of what it takes to understand

'neutrino' is false.

Could Boghossian fall back on a set of minimal inference rules or meaning postulates that are neutral with respect to the case of multiple realization, and perhaps even with respect to the case of no realization? Such weakenings of content often involve increases of complexity in formulation, in order to express the qualifications, as in the retreat from the comparatively simple Carnap sentence (M) to the more complex Lewis postulate (M*). Moreover, even if some principles are in fact neutral with respect to a certain case, that does not prevent someone from mistakenly believing them on theoretical grounds not to be neutral, and rejecting them as a result. Thus the envisaged fallback position is still vulnerable to the problems raised in the two preceding paragraphs.

The complexities uncovered in this section have reinforced the conclusions of section IV. There is no litmus test for understanding a word or having a concept. In particular, willingness to infer according to a specified pattern is not a necessary condition.

VI

Logical Concepts. Boghossian suggests that our basic logical concepts are unconditionalized but non-defective, because no conditionalized version is available. For example, what understanding the conditional commits one to is that it obeys conditional proof and modus ponens, not just that it obeys them if anything does. The reason proposed for the unavailability of the conditionalized version is that the basic logical concepts already occur as auxiliaries in the rules for any conditionalized concept. More specifically, second-order quantification ($\exists$F) and the conditional ($\rightarrow$) occur in the Carnap sentence; second-order

quantification and conjunction occur in rules such as Neutrino-Introduction and Neutrino-Elimination; the Lewis meaning postulates deploy still further conceptual resources.

Boghossian's reasoning is somewhat elliptical. If basic logical concepts occur in all rules for conditionalized concepts, it does not immediately follow that one must have those basic logical concepts *before* one can acquire any conditionalized concepts. It is not straightforward to say why the occurrence of a concept in the rules for conditionalized concepts should imply that it is not itself a conditionalized concept. However, before we address that question we should ask whether it has really been shown that basic logical concepts must occur in the rules for all conditionalized concepts.

Although ∃ and → occur in the Carnap sentence ∃F T(F) → T(Neutrino), in place of that sentence Boghossian could have used the rule that allows one to infer T(Neutrino) directly from any premise of the form T(A). That rule is formulated without reference to any logical operators in the object-language, but is interderivable with the Carnap sentence once one has the standard rules for ∃ and →. Similarly, one could replace Neutrino-Introduction by the rule that permits one to infer Neutrino(x) from premises of the form T(A) and A(x), and Neutrino-Elimination by the rule that permits one to infer a conclusion B(x) from Neutrino(x) and auxiliary premises, given a deduction of B(x) from those auxiliary premises and T(F) and Fx (where the second-order variable F does not occur free in B(x) or the auxiliary premises). These rules too do not involve any logical operators in the object-language but are interderivable with Neutrino-Introduction and Neutrino-Elimination once one has the standard rules for ∃ and &. The three Lewis meaning postulates (M*), (1) and (4) could in principle be replaced by more elaborate inference rules that also do not involve any logical operators in the object-language. Thus the occurrence of ∃, → and & in Boghossian's rules for conditionalized concepts is an inessential artifact of their formulations.

Logical operators may of course occur in the theory T itself, although Boghossian does not appeal to that point. In any case, it seems insufficiently general for his argument, since for some less highly theoretical concepts than *neutrino* the analogue of the theory T for conditionalization may consist of some simple sentences free of logical operators.

What would the rules for a conditionalized logical concept be like? For conjunction, the standard (unconditionalized) introduction rule permits the deduction of the single conclusion A & B from the separate premises A and B, and the standard (unconditionalized) elimination rule permits the deduction of the separate conclusions A and B from the single premise A & B. Thus corresponding conditionalized rules for & might take this form: given deductions of a single sentence C from the separate premises A and B, and of the separate conclusions A and B from the single premise C, deduce the single conclusion A & B from the separate premises A and B, and deduce the separate conclusions A and B from the single premise A & B. In other words, if anything behaves like the conjunction of A and B, A & B does. Such rules are indeed intelligible, but of what use would they be? If one already had another sentence C with the desired inferential powers, such as ~(~A ∨ ~B), one could deduce that A & B also had the desired inferential powers, but from where is that other sentence C to come?

Suppose that one starts with a set of atomic sentences, each logically independent of all the others, and then adds to the language the operators &, ∨, →, ~, each subject to conditionalized versions of its standard introduction and elimination rules. The original language has the feature that any conclusion derivable from a set of premises is interderivable with one of those premises. Since the expanded language inherits that feature, it yields nothing like the usual inferences. For example, where *p* and *q* are distinct atomic sentences, we cannot derive *p* & *q* from *p* and *q* (because *p* & *q* is derivable neither from *p* nor from *q*);

we cannot derive $p \vee q$ from p or from q (because neither p nor q is derivable from $p \vee q$); we cannot derive q from $p \rightarrow q$ and p (because q is derivable from neither $p \rightarrow q$ nor p); we cannot derive q from p and $\sim p$ (because q is derivable from neither p nor $\sim p$). Thus the expanded language is deductively almost inert.[40]

Evidently, we need unconditionalized rules for at least some logical concepts. We may need them even for logical concepts that do not appear in Boghossian's conditionalized rules. For example, negation does not appear in those rules. If we start with just the standard introduction and elimination rules for &, $\vee$, $\rightarrow$, $\exists$ and $\forall$, and conditionalize the rules for ~ and everything else, then we cannot derive any classically valid inferences that are intuitionistically invalid. For the standard introduction and elimination rules for &, $\vee$, $\rightarrow$, $\exists$ and $\forall$ in classical logic are intuitionistically valid, and the conditionalized rules for the other operators never take one outside the realm of intuitionistic validity.[41] The intuitionist agrees that if any sentence B behaves inferentially as the classical logician takes ~A to behave, then ~A behaves in that way; but the intuitionist does not concede that any sentence does behave in that way. For consider the sentence $(p \rightarrow q) \vee (q \rightarrow p)$, which is classically valid but intuitionistically invalid. It is therefore not provable from the standard introduction and elimination rules for $\rightarrow$ and $\vee$, since they are intuitionistically valid, but it can be proved from those rules and standard classical rules for ~, by reductio ad absurdum of its negation. If the intuitionist conceded that any other sentences had the inferential powers attributed to sentences involving ~ in the classical proof, then those other sentences could be used in an intuitionistically acceptable derivation of $(p \rightarrow q) \vee (q \rightarrow p)$, for ~ does not occur in that conclusion itself; but the intuitionist denies that there is any such derivation.[42] Thus if one wants to recover the full power of classical logic, one had better have the unconditionalized classical rules for negation, even though it does not occur in Boghossian's rules for all

conditionalized concepts.

The problematic weakness of conditionalized rules is not limited to logical concepts. Consider a standard version of Zermelo-Fraenkel set theory as a theory ZF(∈) of the set membership relation ∈. The corresponding Carnap sentence is ∃R ZF(R) → ZF(∈). Suppose that the precondition for having the concept of set membership is affirming the Carnap sentence. Of what use would it be in reasoning about sets? Mathematically, the Ramsey sentence ∃R ZF(R) is almost as strong a commitment as ZF(∈) itself, for the consistency of ∃R ZF(R) implies the consistency of ZF(∈). What reason might one have for affirming ∃R ZF(R) other than ZF(∈) itself? Our mere inductive failure so far to find an inconsistency in ZF provides weaker evidence for ∃R ZF(R) than we seem to have for ZF(∈), and therefore for ∃R ZF(R). If we knew how to justify ∃R ZF(R), we should already have overcome the main obstacle to justifying set-theoretic reasoning. Although it might be claimed that not all the principles of Zermelo-Fraenkel set theory play a role in the constitution of the concept of set membership, so that ZF should be replaced by a weaker theory in the account of the conditionalized concept, it remains hard to see what the rules for that concept would contribute to set-theoretic reasoning.[43]

Conditionalized rules for logical and mathematical concepts are not unavailable: they are simply too weak to generate most of the interesting inferences. But that is not a justification of the unconditionalized rules of the kind for which Boghossian was looking. It is merely the crude pragmatic idea that we are justified in making inferences because we could not do without logic: an idea which does little to explain how any particular inference transmits justification from premises to conclusion.

VII

Conclusion. In 'Blind Reasoning', Boghossian does not aim to defend an inferentialist account of concept possession. Rather, he takes such an account as a premise, and makes ingenious use of it in attempting to explain how deductive inference transmits justification. Fine-tuning apart, the present paper has not provided a better strategy for inferentialism to pursue. It has concentrated instead on questioning the inferentialist premise. In doing so, it has relied on this schema, which connects linguistic understanding to concept possession:

(Have) If one understands the word 'C', one has the concept *C*.[44]

Unfortunately for inferentialism, the nature of language as a medium of communication between individuals who disagree with each other in indefinitely various ways undermines attempts to make accepting a given inference a necessary condition for understanding a word; therefore, by (Have), it undermines attempts to make accepting the inference a necessary condition for having the concept. The problem arose both for patterns of inference acceptance of which Boghossian regards as necessary for having defective concepts, such as pejoratives, and for patterns acceptance of which he regards as necessary for having non-defective concepts, such as conditionalized theoretical concepts and unconditionalized logical concepts. His attempt to find a special place for basic logical concepts as unconditionalized but non-defective led to problems of its own in the previous section.

One response to the failure of accepting given patterns of inference to be necessary for having a concept might be that the required conditions are normative rather than psychological. For instance, one understands → if one *ought* to use it in reasoning by

conditional proof and modus ponens, whether or not one actually so reasons. Deviant logicians are not counterexamples to that proposal. Nevertheless, it is apt to seem unsatisfying: if one ought to reason in some way, should not something deeper explain why one ought to reason in that way? In any case, normative inferentialism does not fit Boghossian's project. For he is trying to explain how deductive inference can transfer justification from premises to conclusion. An account of concept possession that simply helps itself to the idea of an inference that one ought to make seems far too close to what he is trying to explain to promise much illumination. If having a concept is a matter of being obliged to reason according to given rules, one might equally ask for an explanation of how one can have the concept, that is, of how one can be obliged to reason according to those rules. That seems no less reasonable than Boghossian's request for an explanation of how reasoning according to those rules can transmit justification.

In very broad terms, the strategy of Boghossian's paper is to reduce a question in the theory of knowledge to questions in the theory of thought and meaning. If the attempt fails, as it apparently does, that is more evidence for the autonomy of the theory of knowledge with respect to the theory of content.[45]

Notes

1 All quotations are from Boghossian (2003) unless otherwise specified. The substitution of blamelessness for justification can be challenged; one may be blameless in making an excusable mistake in a complex and difficult inference without being justified (Plantinga (1993: 39); Pryor (2001: 114-18); Wedgwood (2002: 351-2)).

2 '[A]ccording to Chrysippus, who shows special interest in irrational animals, the dog even shares in the far-famed "Dialectic." This person, at any rate, declares that the dog makes use of the fifth complex indemonstrable syllogism when, on arriving at a spot where three ways meet, after smelling at the two roads by which the quarry did not pass, he rushes off at once by the third without stopping to smell. For, says the old writer, the dog implicitly reasons thus: "The creature went either by this road, or by that, or by the other: but it did not go by this road or by that: therefore it went by the other"', Sextus Empiricus, *Outlines of Pyrrhonism*, i 69 (2000, 41-43). Does the dog have the concept *if*? Does he have a word in his language of thought that means *if*?

3 Here is one of McGee's cases; the others are similar:

> Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:

If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.

A Republican will win the race.

Yet they did not have reason to believe

If it's not Reagan who wins, it will be Anderson.

(1985: 462). Some people claim that such examples involve equivocation, and that on the reading on which they are invalid they are not genuine instances of modus ponens. That would not undermine the points in the text, for even if they are correct the inference still has a reading on which it is an instance of modus ponens, and someone may reject it on that very reading by illicitly shifting away from the reading and back again in the course of theoretical reflection. Note that McGee's putative counterexamples are directed at modus ponens for the 'if' of English, not at the truth-functional conditional $\rightarrow$ of formal logic. Nevertheless, a similar problem about concept possession arises concerning modus ponens for $\rightarrow$, which is equivalent to disjunctive syllogism (from p and $\sim p \vee q$ infer q), a rule that technically competent relevance logicians and dialetheists such as Graham Priest reject (Priest 1995: 5). According to Priest, the best account of paradoxes such as the Liar is that in special circumstances a sentence can be both true and false; when p is true and false while q is merely false, the premises of the disjunctive syllogism are true (for p is true; since p is also false, $\sim p$ is true, so $\sim p \vee q$ is true) but its conclusion is straightforwardly false. Whatever the errors underlying the rejection of modus ponens for $\rightarrow$, they do not arise from a lack of ordinary linguistic understanding of $\rightarrow$ on the part of relevance logicians and dialetheists.

4 Putnam (1975: 228):

HYPOTHESIS OF THE UNIVERSALITY OF THE DIVISION OF

> LINGUISTIC LABOR: Every linguistic community [...] possesses at least some terms whose associated 'criteria' are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers depends upon a structured cooperation between them and the speakers in the relevant subsets.

5 For a related argument about the understanding of words such as 'sofa' and congenial discussion see Burge (1986). Goldberg (2000) replies on behalf of Burge to Bach (1988) and Elugardo (1993).

6 See Martin (1994), Lewis (1997), Martin and Heil (1998) and Bird (1998) for discussion. Harman (1999: 213) relies on defeasible dispositions to infer in his conceptual role semantics.

7 Although there may be self-defeating rules such as 'Never follow a rule!', the rules associated with the usual logic constants are not of that kind.

8 Given that being disposed to use modus ponens for *if* is trivially sufficient for having the concept *if*, the disposition might be held to be both necessary and sufficient for having the concept. But *if* is not the only concept C such that being disposed to use modus ponens for C is necessary and sufficient for having C; the dispositionalist presumably thinks that the concepts of conjunction and of the biconditional satisfy that condition too. Modus ponens is an introduction rule; an elimination rule for *if* is also needed. The standard elimination rule for the conditional is conditional proof,

although that would make *if* equivalent to the truth-functional →. It might be held that *if* is the only concept C such that being disposed to use both modus ponens and conditional proof for C is necessary and sufficient for having C. That does not entail that *if* is the only concept for which one may be disposed to use both modus ponens and conditional proof. Someone who understands 'if and only if' in the normal way and thereby has the biconditional concept but convinces himself through fallacious theoretical reflection that it validates conditional proof may thereby acquire a disposition to use both modus ponens and conditional proof for the biconditional concept too; that disposition defeats but does not destroy his original disposition to make the standard inferences with the biconditional. A variant of that example suggests that being disposed to use modus ponens and conditional proof for 'if' is insufficient for understanding the word. Imagine an Italian learning English as a second language told by an incompetent teacher that 'if' translates 'se e solo se' in Italian; the pupil falsely believes that 'if' in English means *if and only if*; he does not understand the word. Later, he convinces himself through fallacious theoretical reflection that the biconditional concept validates conditional proof; he acquires a new disposition to use both modus ponens and conditional proof for 'if', but he still does not understand the word, for he still falsely believes that it means *if and only if*. It is doubtful that requiring the inferences to be immediate or primitive would help, for many standard rules seem to lack that status for many speakers (compare Harman (1986b)). Peacocke (1998) treats related issues by means of his theory of implicit conceptions, although he does not apply the theory to examples like the present ones.

9 Harman (1986a: 29-42) defends a form of epistemological conservativism on which

our beliefs are justified by default until some incoherence arises. Such a view might also be taken of our inferential dispositions. However, in 'Blind Reasoning', Boghossian rejects Harman's account of justification on the grounds that it depends on a notion of incoherence that is tantamount to what was to be explained.

10 It is not clear that 'cruel' exactly captures the pejorative connotation of 'Boche', but it will do for the sake of argument. Dummett has 'barbarous and more prone to cruelty than other Europeans' (1973: 454).

11 The change from present to past tense is immaterial.

12 Belnap (1962) emphasizes the failure of the rules for 'tonk' to yield a conservative extension; Dummett (1973: 397 and 454) extends the point to 'Boche'. Dummett (1991: 246-51) makes the connection with harmony. As Boghossian notes, Robert Brandom accepts much of Dummett's account of pejoratives but rejects his constraints of harmony and conservative extension (2000: 69-76).

13 A xenophobe might apply the term 'Boche' to someone he knew to be Asian whom he regarded as displaying a distinctively German kind of cruelty, but that is a recognizably metaphorical use. The xenophobe would not apply 'Boche' even metaphorically to someone he knew to be Asian whom he regarded as displaying a distinctively Asian kind of cruelty.

14 The example involves complications about the proper treatment of indexicals and of

descriptive elements in demonstratives, but Frege's general line is clear. Dummett (1973: 84-89) gives a nuanced account of Frege on tone.

15 See Grice (1989: 41) and (1961: section III). The example of the truth-conditionally irrelevant difference between 'and' and 'but' goes back to Frege (1879: §7), but he misdescribes the difference by requiring the contrast to consist in the unexpectedness of what follows 'but' (Dummett (1973: 86). Frege (1979: 140) obscures the conventional nature of the implicature by giving this analogy for someone who uses the word 'cur' without feeling the contempt that it implies: 'If a commander conceals his weakness from the enemy by making his troops keep changing their uniforms, he is not telling a lie; for he is not expressing any thoughts, although his actions are calculated to induce thoughts in others'. A question by Owen Greenhall about the relation between the pairs 'Boche'/'German' and 'but'/'and' in a class at Edinburgh first interested me in the present line of thought about pejoratives.

16 Contrast Stalnaker (1999: 38-40 and 47-62) with Strawson (1952: 175-79). For a related view of conventional implicature see Karttunen and Peters (1979).

17 If the mere use of the pejorative word is what generates the implicature, then the compositional properties of this kind of conventional implicature may contrast with those of the usual paradigms of presupposition. For example, a conditional does not automatically inherit the presuppositions of its consequent: whereas 'He has stopped beating his wife' presupposes 'He once beat his wife', the conditional 'If he once beat his wife then he has stopped beating his wife' lacks that presupposition (for a related

approach to compositional features of conventional implicature see Karttunen and Peters (1979, 33-48)). It is less clear that 'If Germans are cruel then he is a Boche' fails to inherit the conventional implicature 'Germans are cruel' from 'He is a Boche'. These are matters for a more detailed account of pejoratives.

18 On assertion see Williamson (2000: 238-69). Evidently, an infinite regress looms if implying *p* involves implying that one knows *p* (in the same sense of 'imply'), for then one also implies that one knows that one knows *p*, and so *ad infinitum*. Arguably, all but the first few of these implications are false, even if 'know' is weakened to 'is in a position to know' (2000: 114-30). But no such regress flows from the principle that when one conventionally implicates *p*, one conversationally implicates that one knows *p*.

19 The argument in the text does not require conventional implicatures to form a homogeneous category, and therefore withstands recent arguments that they do not (Bach (1999), criticized by Carston (2002: 174-77)). Much recent discussion of conventional implicatures focusses on their role in organizing discourse, as in the case of 'but'; pejoratives typically play no such role. See also note 17 above.

20 Vagueness is another case in which willingness to infer by rules that are not truth-preserving has been treated as a precondition for having some defective concepts. Thus Dummett (1975) argues that observational predicates in natural language are governed by rules that infect the language with inconsistency: for example, to understand 'looks red' one must be willing to apply a tolerance principle by which

one can infer from '*x* is visually indiscriminable from *y*' and '*x* looks red' to '*y* looks red', which generates sorites paradoxes because visual indiscriminability is non-transitive. More recently, Roy Sorensen (2001) has argued that linguistic competence with vague terms involves willingness to make inferences such as that from '*n* seconds after noon is noonish' to '*n*+1 seconds after noon is noonish', which commits us to inconsistent conclusions by sorites reasoning (given our other commitments, such as 'Noon is noonish' and 'Midnight is not noonish'); he combines that view with an epistemic account of vagueness on which vague expressions have non-trivial classical extensions. By the present arguments, such claims about linguistic competence and concept possession are mistaken. An ordinary speaker of English who understands 'looks red' and 'noonish' and has the concepts *looks red* and *noonish* in the normal way but then rejects the relevant tolerance principles in the light of the sorites paradoxes does not thereby cease to understand those expressions or to have those concepts (perhaps she treats the premises of the tolerance principles as providing excellent defeasible evidence for their conclusions, an attitude which is less than Dummett and Sorensen require for competence). Even if the whole linguistic community abandons its supposed commitment to the tolerance principles, without stipulating any cut-off points, that would not make 'looks red' or 'noonish' any more precise, so speakers' acceptance of tolerance principles is in any case quite inessential to vagueness.

21 If concepts are thought constituents, and *Boche* and *German* are the same concept, then in judging *Germans are German* one simultaneously judges *Germans are Boche*, however much one's reactions discriminate between the sentences 'Germans are

German' and 'Germans are Boche'. Compare: if *furze* and *gorse* are the same concept, then in judging *furze is furze* one simultaneously judges *furze is gorse*, however much one's reactions discriminate between the sentences 'Furze is furze' and 'Furze is gorse'.

22 Despite the remarks in the text, I share Boghossian's scepticism about the capacity of Wittgensteinian appeals to the community to solve (or dissolve) fundamental problems of philosophy.

23 Proof: (S) is a logical consequence of (M) only if (S) is a logical truth. For (M) is a logical consequence of ~(S), so (M) logically entails (S) only if ~(S) logically entails (S), in which case (S) is logically true. But (S) is not logically true; logic does not guarantee that some things play the role that neutrinos play according to the theory.

24 Like Boghossian, I treat 'neutrino' as a schematic example, without reference to any specific features of its actual use in physics.

25 See for example Peacocke's (1992: 29-33) discussion of deference-dependent propositional attitude ascriptions. Burge extends his earlier arguments and argues for such a deeper lesson in his (1986).

26 For the relevance of the model of full understanding as full induction into a practice to the theory of vagueness see Williamson (1994: 211-12).

27 For more discussion of the unique characterization requirement see Williamson (1987/88), McGee (2000) and references therein.

28 For simplicity, the comparatively weak condition of the provability of coextensiveness from the two theories is used as the criterion of equivalence. Since the provability of coextensiveness does not guarantee the provability of necessary coextensiveness, this leaves it open that equivalent terms do not stand for the same property. A fuller discussion would address such modal matters; they will be ignored here, as not central to the present issues.

29 Proof: Since the main antecedent in (!) follows by propositional logic from ~∃F T(F), (!) is logically true only if ∀G∀H(~∃F T(F) → ∀*x*(G*x* ↔ H*x*) is logically true; but it is logically equivalent to ~∃F T(F) → ∀G∀H∀*x*(G*x* ↔ H*x*) since the variables G and H do not occur in T(F); the consequent of that formula is logically false, since we may substitute ~G for H and assume that there is at least one individual; thus the formula as a whole is equivalent to ∃F T(F). That argument treats → in (M) as the material conditional, contrary to the stipulation of Horwich (2000: 157). However, Horwich specifies no alternative reading of →. Horwich's assumption that a theory is equivalent to the conjunction of its Ramsey sentence and its Carnap sentence also becomes questionable on a non-material reading of →, for why should T(Neutrino) entail ∃F T(F) → T(Neutrino)? The present paper follows the standard practice in discussion of these issues by using a material conditional.

30 Proof: Unique characterization for the modified conditionalized concept is equivalent

to the logical truth of this formula:

(!!) $\forall G\forall H(((\exists F\, T(F) \rightarrow T(G)) \,\&\, (\exists x\, Gx \rightarrow \exists F\, T(F)) \,\&\, (\exists F\, T(F) \rightarrow T(H)) \,\&\, (\exists x\, Hx \rightarrow \exists F\, T(F))) \rightarrow \forall x(Gx \leftrightarrow Hx))$

Now (!!) is easily seen to be logically equivalent to the following conjunction:

$((S) \rightarrow (!+)) \,\&\, (\sim(S) \rightarrow \forall G\forall H((\sim\exists x\, Gx \,\&\, \sim\exists x\, Hx) \rightarrow \forall x(Gx \leftrightarrow Hx))$

The latter conjunct is logically true, so the whole formula is logically equivalent to $(S) \rightarrow (!+)$. But since $\sim(S) \rightarrow (!+)$ is a vacuous logical truth, $(S) \rightarrow (!+)$ is logically equivalent to $(!+)$ itself. Thus (!!) is logically equivalent to $(!+)$. Since the modified conditionalized concept satisfies unique characterization if and only if (!!) is logically true, and the unconditionalized concept satisfies unique characterization if and only if $(!+)$ is logically true, the modified conditionalized concept satisfies unique characterization if and only if the unconditionalized concept does.

31 For the problem of unique realization in the case of Frank Jackson's (1998) attempt to apply the Ramsey-Lewis method to ethical terms see Williamson (2001: 629-30).

32 Proof: We can abbreviate (M*), (1) and (4) as $((S) \,\&\, (!+)) \rightarrow T(\text{Neutrino})$, $\exists x\, \text{Neutrino}(x) \rightarrow (S)$ and $\exists x\, \text{Neutrino}(x) \rightarrow (!+)$ respectively. In this case, therefore, unique characterization is equivalent to the logical truth of this formula:

$\forall G\forall H(((S) \,\&\, (!+)) \rightarrow T(G)) \,\&\, (\exists x\, Gx \rightarrow ((S) \,\&\, (!+))) \,\&\, ((S) \,\&\, (!+)) \rightarrow T(H)) \,\&\, (\exists x\, Hx \rightarrow ((S) \,\&\, (!+)))) \rightarrow \forall x(Gx \leftrightarrow Hx)$

That formula is easily seen to be equivalent to this one:

$(((S) \,\&\, (!+)) \rightarrow \forall G\forall H((T(G)) \,\&\, T(H)) \rightarrow \forall x(Gx \leftrightarrow Hx))) \,\&$

$(\sim((S) \,\&\, (!+)) \rightarrow \forall G\forall H((\sim\exists x\, Gx \,\&\, \sim\exists x\, Hx) \rightarrow \forall x(Gx \leftrightarrow Hx)))$

But the consequents of the two conjuncts are just (!+) and a logical truth respectively, so the whole formula is a logical truth.

33 Boghossian states his rules for the conditionalized version of one of his artificially stipulated concepts, but the form of his discussion is quite general. The rules are adapted to 'neutrino' here for ease of comparability with the other principles.

34 Example: Let T(F) be $\forall x(Fx \leftrightarrow \text{Square}(x)) \vee \forall x(Fx \leftrightarrow \sim\text{Square}(x))$, where Square applies to all and only square things. Thus T(Square) and T(~Square) are both true, so the Ramsey sentence $\exists F\ T(F)$ is also true. Define $C(x)$ as $\exists F(T(F) \,\&\, Fx)$. Since everything is either square or not square, C applies to everything. So T(C) is false, since some things are squares and others are not. Thus the Carnap sentence $\exists F\ T(F) \rightarrow T(C)$, the analogue of (M), is false. Here as elsewhere, more realistic cases could be provided, but trivial examples are logically the most perspicuous.

35 Proof: From Neutrino-Introduction we derive $\forall F(T(F) \rightarrow \forall x(Fx \rightarrow \text{Neutrino}(x)))$. From Neutrino-Elimination we derive $(!+) \rightarrow \forall F(T(F) \rightarrow \forall x(\text{Neutrino}(x) \rightarrow Fx))$. Putting the two together yields $(!+) \rightarrow \forall F(T(F) \rightarrow \forall x(\text{Neutrino}(x) \leftrightarrow Fx))$. We may assume that T is extensional in this sense: $\forall G\forall H(\forall x(Gx \leftrightarrow Hx) \rightarrow (T(G) \leftrightarrow T(H)))$. Thus we have $(!+) \rightarrow \forall F(T(F) \rightarrow T(\text{Neutrino}))$, which is equivalent to (M*).

36 Examples: Let C apply to all and only squares. For the introduction rule, let T be tautologous; thus $\exists F\ T(F) \rightarrow T(C)$ is logically true (because its consequent is) and everything satisfies $\exists F(T(F) \,\&\, Fx)$ (because everything satisfies $T(C) \,\&\, C(x)$ or $T(\sim C)$

& ~ C(x)), so every non-square yields a counter-instance to the inference from it to C(x). For the elimination rule, let T be inconsistent; thus ∃F T(F) → T(C) is logically true (because its antecedent is logically false) and nothing satisfies ∃F(T(F) & Fx), so every square yields a counter-instance to the inference to it from C(x).

37 Example: As in note 34.

38 Proof: (1), (4) and (M*) yield ∃x Neutrino(x) → T(Neutrino). Thus from Neutrino(x) we derive T(Neutrino) & Neutrino(x) and thence ∃F(T(F) & Fx).

39 Example: Let T be tautologous and C apply to nothing.

40 A few trivial inferences can still be made. For instance, p is interderivable with p & p, $p \vee (p \,\&\, p)$, and so on.

41 The reason is that even the intuitionistic introduction and elimination rules are strong enough for the proof of unique characterization; the full power of classical logic is not needed. Thus the intuitionistic negation rules for an operator imply the conditionalized classical rules for that operator. See Williamson (1987/88: 110-14) for discussion.

42 The example shows that the standard classical rules for negation yield a non-conservative extension of the standard classical rules for the other connectives, a point crucial to Dummett's (1991) meaning-theoretic case against classical logic. For an

alternative way of formulating classical logic in response to Dummett's critique see Rumfitt (2000).

43 We cannot expect set theory to characterize set membership uniquely, for if set theory has one realization, it has many. At most we can hope that our set theory will characterize set membership uniquely up to isomorphism; McGee (1997) provides a strengthening of Zermelo-Fraenkel set theory with individuals that does that. For set theory with individuals, we need a concept of set in addition to the concept of set membership, in order to distinguish the null set from the individuals; this is immaterial to the argument in the text. For simplicity, issues about unique characterization have been largely ignored in this section; they do not affect its main line of argument.

44 Understanding the word 'C' in (Have) must be read as understanding 'C' with its present meaning.

45 Thanks to Paul Boghossian and participants in classes at Oxford for discussion of some of the ideas in this paper, and to Nico Silins, Jason Stanley and Ralph Wedgwood for written comments on an earlier version.

REFERENCES

K. Bach. 1988. 'Burge's New Thought Experiment: Back to the Drawing Room', *Journal of Philosophy* 85, 88-97.

K. Bach. 1999. 'The Myth of Conventional Implicature', *Linguistics and Philosophy* 22, 327-66.

N. Belnap. 1962. 'Tonk, Plonk and Plink', *Analysis* 22, 130-4.

A. Bird. 1998. 'Dispositions and Antidotes', *Philosophical Quarterly* 48, 227-34.

P. Boghossian. 1989. 'The Rule-Following Considerations', *Mind* 98, 507-49.

P. Boghossian. 2003. 'Blind Reasoning', this volume.

L. Bonjour. 1985. *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard University Press.

R. Brandom. 2000. *Articulating Reasons: An Introduction to Inferentialism*, Cambridge, Mass.: Harvard University Press.

T. Burge. 1979. 'Individualism and the Mental', in P. French, T. Uehling and H. Wettstein, eds., *Midwest Studies in Philosophy 4: Studies in Metaphysics*, Minneapolis: University of Minnesota Press.

T. Burge. 1986. 'Intellectual Norms and Foundations of Mind', *Journal of Philosophy*, 83, 697-720.

R. Carnap. 1966. *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*, New York: Basic Books.

R. Carston. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Oxford: Blackwell.

M. A. E. Dummett. 1973. *Frege: Philosophy of Language*, London: Duckworth.

M. A. E. Dummett. 1975. 'Wang's Paradox', *Synthese*, 30, 301-24.

M. A. E. Dummett. 1991. *The Logical Basis of Metaphysics*, London: Duckworth.

R. Elugardo. 1993. 'Burge on Content', *Philosophy and Phenomenological Research*, 53, 367-84.

G. Evans. 1982. *The Varieties of Reference*, ed. J. McDowell, Oxford: Clarendon Press.

G. Frege. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert. Trans. S. Bauer-Mengelberg in J. van Heijenoort, ed., *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, Mass.: Harvard University Press, 1967.

G. Frege. 1979. *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, trans. P. Long and R. White, Oxford: Basil Blackwell.

S. Goldberg. 2000. 'Do Anti-Individualistic Construals of Propositional Attitudes Capture the Agent's Conceptions?', *Noûs* 36, 597-621.

H. P. Grice. 1961. 'The Causal Theory of Perception', *Proceedings of the Aristotelian Society*, sup. vol. 35, 121-52.

H. P. Grice. 1989. *Studies in the Way of Words*, Cambridge, Mass.: Harvard University Press.

G. Harman. 1986a. *Change in View: Principles of Reasoning*, Cambridge, Mass.: MIT Press.

G. Harman. 1986b. 'The Meaning of Logical Constants', in E. LePore, ed., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.

G. Harman. 1999. *Reasoning, Meaning, and Mind*, Oxford: Clarendon Press.

J. H. Harris. 1982. 'What's so Logical about the "Logical" Axioms?', *Studia Logica*, 41, 159-71.

P. Horwich. 2000. 'Stipulation, Meaning, and Apriority', in P. Boghossian and C. Peacocke,

eds., *New Essays on the A Priori*, Oxford: Clarendon Press.

F. Jackson. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

L. Karttunen and S. Peters. 1979. 'Conventional Implicature', in C.-K. Oh and D. A. Dinneen (eds.), *Syntax and Semantics, 11*, New York: Academic Press.

S. Kripke. 1979. 'A Puzzle about Belief', in A. Margalit (ed.), *Meaning and Use*, Dordrecht: Reidel.

D. K. Lewis. 1970. 'How to Define Theoretical Terms', *Journal of Philosophy*, 67, 427-46. Reprinted in his *Philosophical Papers*, vol. I, Oxford: Oxford University Press, 1983, to which page numbers refer.

D. K. Lewis. 1997. 'Finkish Dispositions', *Philosophical Quarterly*, 47, 143-58.

V. McGee. 1985. 'A Counterexample to Modus Ponens', *Journal of Philosophy*, 82, 462-71.

V. McGee. 1997. 'How We Learn Mathematical Language', *Philosophical Review*, 106, 35-68.

V. McGee. 2000. ''Everything'', in G. Sher and R. Tieszen, eds., *Between Logic and Intuition*, Cambridge: Cambridge University Press.

C. B. Martin. 1994. 'Dispositions and conditionals', *Philosophical Quarterly* 44, 1-8.

C. B. Martin and J. Heil. 1998. 'Rules and powers', in J. Tomberlin, ed., *Philosophical Perspectives, 12: Language, Mind and Ontology*, Oxford: Blackwell.

C. Peacocke. 1992. *A Study of Concepts*. Cambridge, Mass.: MIT Press.

C. Peacocke. 1998. 'Implicit Conceptions, Understanding and Rationality', in E. Villanueva ed., *Philosophical Issues, 8: Concepts*, Atascadero, Calif.: Ridgeview.

A. Plantinga. 1993. *Warrant: The Current Debate*, New York: Oxford University Press.

G. Priest. 1995. *Beyond the Limits of Thought*, Cambridge: Cambridge University Press.

A. N. Prior. 1960. 'The Runabout Inference-Ticket', *Analysis* 21, 38-39.

J. Pryor. 2001. 'Highlights of Recent Epistemology', *British Journal for the Philosophy of Science*, 52, 95-124.

H. Putnam. 1975. *Mind, Language and Reality: Philosophical Papers, Volume 2*, Cambridge: Cambridge University Press.

F. P. Ramsey. 1929. 'Theories'. Republished in his *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, ed. D. H. Mellor. London: Routledge & Kegan Paul, 1978.

I. Rumfitt. 2000. '"Yes" and "No"', *Mind*, 109, 781-823.

Sextus Empiricus. 2000. *Outlines of Pyrrhonism*, trans. R. G. Bury. Cambridge, Mass.: Harvard University Press.

R. Sorensen. 2001. *Vagueness and Contradiction*, Oxford: Clarendon Press.

R. Stalnaker. 1999. *Context and Content*, Oxford: Oxford University Press.

P. F. Strawson. 1952. *Introduction to Logical Theory*, London: Methuen.

R. Wedgwood. 2002. 'Internalism Explained', *Philosophy and Phenomenological Research* 64, 349-69.

T. Williamson. 1987/88. 'Equivocation and Existence', *Proceedings of the Aristotelian Society*, 87, 109-27.

T. Williamson. 1994. *Vagueness*, London: Routledge.

T. Williamson. 2000. *Knowledge and its Limits*, Oxford: Oxford University Press.

T. Williamson. 2001. 'Ethics, Supervenience and Ramsey Sentences', *Philosophy and Phenomenological Research*, 62, 625-30.