(forthcoming in *The Aristotelian Society* supplementary volume **80** (2006))

"Conceptual Truth"*

Timothy Williamson

ABSTRACT The paper criticizes epistemological conceptions of analytic or conceptual truth, on which assent to such truths is a necessary condition of understanding them. The critique involves no Quinean scepticism about meaning. Rather, even granted that a paradigmatic candidate for analyticity is synonymous with a logical truth, both the former and the latter can be intelligibly doubted by linguistically competent deviant logicians, who, although mistaken, still constitute counterexamples to the claim that assent is necessary for understanding. There are no analytic or conceptual truths in the epistemological sense. The critique is extended to purportedly analytic inference rules. An alternative account is sketched on which understanding a word is a matter of participation in a linguistic practice, while synonymy and concept identity consist in sameness of truth-conditional semantic properties. Although there are philosophical questions about concepts, the idea that philosophical questions in general are conceptual questions generates only an illusion of insight into philosophical methodology.

I

How do we know that vixens are female foxes? Such questions tend to receive short shrift. We are told that it is a *conceptual truth* that vixens are female foxes, or that it is *conceptually impossible* for something to be a vixen without being a female fox, or that being a vixen has *conceptual connections* to being female and being a fox. In unfashionable terminology, 'Vixens are female foxes' is said to be *analytic*. What, if anything, do such responses mean? How, if at all, do they answer the original question?

Since it is a boring triviality that vixens are female foxes, one might wonder how much those questions matter. Yet human reasoning is riddled with steps like those from 'vixen' to 'female' and 'fox'. Many are equally trivial, but more significant steps of reasoning have been assimilated to the trivial ones: for example, basic inferences in deductive logic, characteristic moves in philosophical argument, and fundamental inferences involving theoretical terms in natural science have all been treated as somehow built into the concepts or the meanings of the words at issue, and as backed by conceptual or analytic truths with a status not fundamentally different from that of 'Vixens are female foxes'. Until we are clear about the epistemology of the simple, unimportant cases, we are poorly placed to judge the aptness of assimilating the more complex, important cases to them, and whether doing so renders the latter epistemologically unproblematic. Some mathematicians use the rule of thumb that one should not try to solve a problem when one cannot yet solve a simpler problem of the same form. Applying the rule to questions of conceptual truth, let us start with 'Every vixen is a female fox'.[1]

A common view is that analytic or conceptual truths are epistemologically unproblematic because whatever cognitive work is necessary for understanding them is somehow already sufficient for knowing them to be true. Thus principles like this are implicitly or explicitly proposed:

UKt Necessarily, whoever grasps the thought that every vixen is a female fox knows that every vixen is a female fox.

Some clarifications are in order. To grasp a thought is to entertain it, irrespective of what specific attitude such as belief, disbelief or wonder one takes towards it. Since a thought in UKt is the candidate bearer of *conceptual* truth, we should presumably take it as composed of concepts in something like the way in which a sentence is composed of words. Thus a thought is not a set of possible worlds, which has no such structure. Nor is a thought a Russellian proposition, composed out of the objects, properties and relations which it is about; a thought about Vienna contains a concept of Vienna, not Vienna itself. If a proposition is a Fregean thought, the sense of a sentence, composed of the senses of its constituent expressions, then thoughts may be propositions. But if propositions are sets of possible worlds, Russellian propositions or something else of that worldly kind, then thoughts are not propositions, although they may express propositions. For example, if it is a conceptual truth that Hesperus is Hesperus, but no conceptual truth that Hesperus is Phosphorus, then the thought that Hesperus is Hesperus is not the thought that Hesperus is Phosphorus, even though the Russellian proposition that Hesperus is Hesperus is the Russellian proposition that Hesperus is Phosphorus, and the set of possible worlds in

which Hesperus is Hesperus is the set of possible worlds in which Hesperus is Phosphorus.[2,3] If propositions are not thoughts, questions of conceptual truth, conceptual necessity and conceptual connections do not arise at the level of propositions, but only at that of thoughts. Consequently, there is a terminological awkwardness if propositional attitude constructions in English (such as 'knows that every vixen is a female fox') ascribe relations to propositions rather than thoughts, for then UKt is not formulated at the level at which questions of conceptual truth arise. However, we can finesse the issue by reading UKt in that case as concerning only grasping and knowing which relate to the proposition that every vixen is a female fox under the guise of the thought which the sentence 'Every vixen is a female fox' expresses in the present context. For defenders of conceptual truth, such a reading should not be too artificial. Analogous qualifications will be in force throughout this paper. Thus the theorist of thought can allow that one may know that Hesperus is Hesperus without thereby knowing that Hesperus is Phosphorus, and that knowing that 0 = 0 does not automatically count as knowing that every vixen is a female fox.[4]

A linguistic analogue of UKt is:

UKl Necessarily, whoever understands the sentence 'Every vixen is a female fox' recognizes it as true.

Recognition as true here is assent to the sentence on the basis of knowledge of the truth which it expresses. For example, you recognize the sentence 'Every vixen is a female fox' as true because you know that it means that every vixen is a female fox and that

every vixen *is* a female fox. The explicit meta-linguistic thought that the sentence is true is not required. Thus recognition is factive, because knowledge is: one can recognize a sentence as true only if it is true. Of course, the sentence 'Every vixen is a female fox' in UKl must be taken with its actual present meaning in English, since UKl would trivially fail if we allowed understandings of 'Every vixen is a female fox' with a counterfactual meaning, such as that every viscount is a female fox. In charity too, we can and should read UKt and UKl so that a native speaker of English who has never explicitly addressed the question whether every vixen is a female fox does not thereby constitute a counter-example. Call principles such as UKt and UKl 'understanding/knowledge links'.

In effect, understanding/knowledge links say that the epistemological problem is automatically solved for the thoughts or sentences in question; but they do not say *how* it is solved. An account of that might start with some weak consequences of such links. For UKt and UKl entail UBt and UBl respectively:

UBt Necessarily, whoever grasps the thought that every vixen is a female fox believes that every vixen is a female fox.

UBl Necessarily, whoever understands the sentence 'Every vixen is a female fox' assents to it.

For, we may assume, knowledge entails belief and recognition as true entails assent, but not *vice versa*. Unlike knowledge and recognition, belief and assent are not factive. Call principles such as UBt and UBl 'understanding/belief' links. One might regard UBt and

UBl as more or less bedrock or constitutive constraints on, respectively, possession of the concepts which make up the thought that every vixen is a female fox and understanding of the words and modes of combination which make up the sentence 'Every vixen is a female fox'. On this view, belief is simply part of what it takes to grasp the thought, and assent is simply part of what it takes to understand the sentence. In the terminology of Carnap (1947: 222), the sentence is a meaning postulate. Similar conceptions remain in the background of much contemporary discussion.

Since belief does not entail knowledge and assent does not entail recognition as true, the respective paths back from UBt and UBl to UKt and UKl must be less direct. Nevertheless, attempts have been made to explain understanding/knowledge links by appeal to understanding/belief links. On such a view, an understanding/belief link for a thought or sentence somehow constrains it to have a content or meaning on which it is true: accordingly, UBt makes the thought that every vixen is a female fox true and UBl makes the sentence 'Every vixen is a female fox' true. Thus the understanding/belief link would somehow generate an understanding/truth link, so that the belief would be formed in a reliable way. That reliability might itself be regarded as epistemically transparent or otherwise special enough to generate, somehow or other, an understanding/knowledge link. In some such way, the constitutive conditions for grasping the thought or understanding the sentence are supposed to explain the knowledge in question. If such an epistemological schema is realized, the thought, for instance that every vixen is a female fox, will deserve the title of 'conceptual truth', and the corresponding sentence will deserve to be called 'analytic'.[5]

For philosophers of an internalist bent, the key epistemological notion is justification rather than knowledge. They will focus on understanding/justification links:

UJBt Necessarily, whoever grasps the thought that every vixen is a female fox is justified in believing that every vixen is a female fox.

UJBl Necessarily, whoever understands the sentence 'Every vixen is a female fox' is justified in assenting to it.

Some rationalists may regard such links as characteristic of *a priori* intuition. Whether they try to derive them from understanding/belief links, and whether they endorse the preceding story about understanding/truth and understanding/knowledge links, are further questions.

It is often allowed that understanding/belief links can fail to generate understanding/truth links, and thereby fail to generate understanding/knowledge links, if the beliefs built into understanding by the former links embody substantive commitments about the world, for the world may be unkind. Perhaps the understanding of some theoretical terms in science embodies theories not satisfied by any things in nature at all, as some say of 'phlogiston'. The beliefs built into understanding may even turn out to be inconsistent, as some say of 'true', adverting to the threat which the Liar paradox poses to disquotational principles about truth. Once the error has come to light, such beliefs cease to be justified, if they ever were, and the understanding/justification links fail, although versions watered down with a '*ceteris paribus*' clause might survive. We should reject

the beliefs together with the concepts and meanings in which they are embedded. Of course, where understanding/truth links fail there is no analytic or conceptual *truth*. Nevertheless, the idea of concepts or meanings which embody errors or confusions itself depends on something like understanding/belief links.

Such failures of the understanding/truth links may be treated as pathological cases. Thus understanding/belief links may still be held to generate understanding/truth links in non-pathological cases, and thereby indirectly to explain understanding/knowledge links. Others will treat the failure of understanding/belief links to generate understanding/truth links as more general, and reject the understanding/knowledge links altogether. For example, some naturalists hold that many assumptions built into ordinary concepts or the meanings of ordinary words are simply the prejudices of folk theory: even when the assumptions are in fact true, merely possessing the concepts or understanding the words does not enable one to know that they are true. But someone who rejects a folk concept or part of folk language on the grounds that it embeds a false theory still seems to presuppose something like an understanding/belief link, for without such links there is no embedding: one could keep the understanding and drop the beliefs.[6]

The foregoing remarks, however sketchy and programmatic, serve to indicate a familiar cluster of vague ideas. The cluster has been under a cloud since Quine argued in 'Two Dogmas of Empiricism' that 'a boundary between analytic and synthetic statements simply has not been drawn' (1953: 37). But without a strong independent desire to believe Quine's conclusions, it is hard to find his arguments compelling. They require verificationist or reductionist assumptions about meaning which we lack good reason to

accept. One of his targets, the idea of synonymy, is alive and well both in semantics as a branch of linguistics and in the philosophy of language, which suggests that his arguments, if they proved anything, would prove too much. Although the idea of analyticity has never quite regained the central position in analytic philosophy which it occupied before 'Two Dogmas', the reason for that may be less Quine's critique than Kripke's clarification of the differences between analyticity, apriority and metaphysical necessity. Kripke did not deny the existence of a boundary between the analytic and the synthetic; he merely distinguished it from other boundaries, between the a priori and the a posteriori and between the necessary and the contingent (1980: 39). In effect, he showed that the analytic/synthetic distinction could not do the work proper to those other distinctions, contrary to what had been expected of it in the heyday of the 'linguistic turn'.[7] Often under the less provocative guise of 'conceptual truth' or 'conceptual necessity', analyticity plays a reduced but still substantive role in contemporary philosophy.

There is little reason to doubt that if we try to sort statements as 'analytic' or 'synthetic' in the manner of chicken-sexers, we can generally achieve a rough consensus. Of course borderline cases will occur, but so they do for almost every distinction worth making: perfect precision is an unreasonable demand. The issue is what theoretical significance, if any, attaches to the rough boundary thus drawn. As indicated above, much of the putative significance of the analytic/synthetic distinction is epistemological. That epistemological significance depends on something like understanding/belief links.

This paper gives reasons to reject understanding/belief links. *A fortiori*, those are reasons to reject understanding/knowledge links too. Similar considerations defeat

understanding/justification links. Those putative links depend on a misapprehension of what it is to possess a concept or to understand a word. When we strip out the verificationism and scepticism about meaning from Quine's arguments, epistemological insights remain. The arguments of this paper involve no Quinean scepticism about meaning. Rather, they rely on our rough working conceptions of meaning and understanding — for example, when applying UBl — in order to reach the conclusion that no truths are analytic in the epistemological sense.

Section II argues that anything like an understanding/belief link fails for some elementary logical truths. Section III generalizes the argument to other truths which are often called 'analytic' or 'conceptual'. If analyticity or conceptual truth requires anything like an understanding/belief link, then there is no analyticity or conceptual truth.[8] The final section begins to sketch an alternative account of understanding, on which no such links are needed.

II

Here is an elementary logical truth:

(1) Every vixen is a vixen.

Few quantified logical truths are simpler than (1), in either syntactic complexity or the number of steps needed to derive them in a standard system of natural deduction rules.[9]

One might be tempted to endorse understanding/belief links for (1):

UBt* Necessarily, whoever grasps the thought that every vixen is a vixen believes that every vixen is a vixen.

UBl* Necessarily, whoever understands the sentence 'Every vixen is a vixen' assents to it.

On the pattern of section I, UBt* and UBl* might be exploited in an attempt to derive corresponding understanding/knowledge links for (1):

UKt* Necessarily, whoever grasps the thought that every vixen is a vixen knows that every vixen is a vixen.

UKl* Necessarily, whoever understands the sentence 'Every vixen is a vixen' recognizes it as true.

But are UBt* and UBl*, let alone UKt* and UKl*, even true? Consider two native speakers of English, Peter and Stephen.

Peter's first step in evaluating (1) is to notice that it seems to presuppose:

(2) There is at least one vixen.

On reflection, Peter comes to the considered view that the presupposition is a logical entailment. He regards the truth of 'There is at least one F' as a necessary condition for the truth of 'Every F is a G' quite generally, and the falsity of 'There is at least one F' as a sufficient condition for the falsity of 'Every F is a G'; he takes universal quantification to be existentially committing. More formally, he holds that 'Every F is a G' is true if and only if (i) there is a value of the variable '*x*' for which '*x* is an F' is true and (ii) there is no value of the variable '*x*' for which '*x* is an F' is true while '*x* is a G' is not, and that 'Every F is a G' is false if and only if it is not true. Peter also has the weird belief that (2) is false. For he spends far too much time surfing the internet, and once came across a site devoted to propagating the view that there are no foxes, and therefore no vixens, and never have been: all the apparent evidence to the contrary has been planted by a secret international agency; for sinister purposes best known to itself, it produces elaborate fox-hallucinations. Being a sucker for conspiracy theories, Peter accepted this one. Since he denies (2) and regards it as a logical consequence of (1), he also denies (1), and so does not assent to (1).[10]

Stephen has no time for Peter's pet theories. What worries him is vagueness. He believes that borderline cases for vague terms constitute truth-value gaps. Like many truth-value gap theorists (such as Soames 1999), he generalizes classical two-valued semantics by treating the gap as a third value ('undefined') and using Kleene's three-valued 'strong tables' (1952: 334). On Stephen's view, for 'Every F is a G' to be true is for the conditional '*x* is an F → *x* is a G' to be true for every value of the variable '*x*'; for 'Every F is a G' to be false is for '*x* is an F → *x* is a G' to be false for some value of '*x*'. On his semantics, for the conditional sentence with '→' to be true is for either its

antecedent to be false or its consequent to be true, and for it to be false is for its antecedent to be true and its consequent false. Stephen also believes that some clearly female evolutionary ancestors of foxes are borderline cases for 'fox' and therefore for 'vixen'. Consequently, for such an animal as the value of '*x*', '*x* is a vixen' is neither true nor false, so the conditional '*x* is a vixen → *x* is a vixen' is also neither true nor false, by the strong Kleene table for →. Hence 'Every vixen is a vixen' is not true; it is also not false, because the conditional is not false for any value of '*x*'. Thus Stephen treats (1) as a truth-value gap. Of course, his initial reaction when presented with (1) is not to go through this explicit metalinguistic reasoning; he just says 'What about borderline cases?' But his refusal to assent to (1) as true is firm.[11]

Let us assume that Peter and Stephen are wrong about (1), at least on its standard reading: it is in fact a logical truth. It is true however we interpret its only non-logical syntactically atomic constituent, 'vixen', given classical logic and two-valued semantics. If not, we can change the example, describing new characters who are deviant with respect to some sentence which really is a logical truth. Peter and Stephen do not assent to (1). Thus, according to UBl*, Peter and Stephen do not understand (1) (with its standard English meaning). If so, they presumably fail to understand at least one of its constituent words or modes of combination. Is that the impression which one would have in conversing with them?

Both Peter and Stephen treat 'vixen' as synonymous with 'female fox'. Stephen's popular but mistaken theory of vagueness does not prevent him from understanding 'female', 'fox' or their mode of combination. Even Peter's conspiracy theory, however silly, involves no semantic deviation, just as religious fanatics who assert that there were

never any dinosaurs do exactly that: they use the words ‘There were never any dinosaurs’ to assert that there were never any dinosaurs. Their problem is not that they fail to understand the word ‘dinosaur’, but that they have silly beliefs about prehistory. Peter, like Stephen, understands the word ‘vixen’.

The best candidate for a word or mode of composition in (1) which Peter and Stephen fail to understand is the word ‘every’. Is it a good enough candidate? Peter’s not uncommon conception of the existential commitments of universal quantification makes little difference in practice, for when sentences of the form ‘Every F is a G’ occur in conversation, ‘There is an F’ tends to be common ground amongst the participants anyway; it is a pragmatic presupposition in the sense of Stalnaker (1999). Pragmatically, Peter adjusts his conversation to a society which obstinately retains its belief in the existence of foxes much as members of many other small sects with unpopular beliefs have learned to adjust to an unenlightened world. Stephen’s deviation is less localized than Peter’s, because his Kleene-inspired semantics turns many universal generalizations with empirical predicates into truth-value gaps. In practice, however, he often manages to ignore the problem by focussing on a small domain of contextually relevant objects amongst which there are no borderline cases for the noun or complex phrase which complements ‘every’. Occasionally he cannot avoid the problem and sounds pedantic, as many academics too, but that hardly constitutes a failure to understand the words at issue. When Peter and Stephen are challenged on their logical deviations, they defend themselves fluently. In fact, both have published widely read articles on the issues in leading refereed journals of philosophy, in English. They appear to be like most

philosophers, thoroughly competent in their native language, a bit odd in some of their views.

Someone might insist that Peter and Stephen appear to be using the word ‘every’ in its standard sense because they are really using it in senses very similar to, but not exactly the same as, the standard one. Indeed, it may be argued, their non-standard senses were explained above, since in each case a semantics for the relevant fragment of English was sketched on which (1) is not true, whereas by hypothesis (1) is true on the standard semantics of English. However, matters are not so simple. Peter and Stephen are emphatic that they intend their words to be understood as words of our common language, with their standard English senses. They use ‘every’ and the other words in (1) as words of that public language. Each of them believes that his semantic theory is correct for English as spoken by others, not just by himself, and that if it turned out to be (heaven forbid!) incorrect for English as spoken by others, it would equally turn out to be incorrect for English as spoken by himself. Giving an incorrect theory of the meaning of a word is not the same as using the word with an idiosyncratic sense. Peter and Stephen’s semantic beliefs about their own uses of ‘every’ may be false, even if they sometimes rely on those beliefs in conscious processes of truth-evaluation. Indeed, we may assume that Peter and Stephen do not regard the elaborate articulations of truth-conditions and falsity-conditions for ‘Every F is a G’ above as capturing the way in which they or other English speakers conceptualize the meaning of ‘every’, which they regard as a semantically unstructured determiner for which a homophonic statement of meaning would be more faithful. For them, the more elaborate articulations are simply convenient records of important logical facts about ‘every’. We may further assume that only in

tricky cases do they resort to their non-standard semantic theories in evaluating non-meta-linguistic claims such as (1) expresses. Their non-meta-linguistic unorthodoxy as to when every F is a G is not ultimately derived by semantic descent from meta-linguistic unorthodoxy as to when 'Every F is a G' is true; rather, the meta-linguistic unorthodoxy is ultimately derived by semantic ascent from the non-meta-linguistic unorthodoxy.

Peter and Stephen learned English in the normal way. They acquired their non-standard views as adults. At least before that, nothing in their use of English suggested semantic deviation. Surely they understood (1) and its constituent words and modes of construction with their ordinary meanings then. But the process by which they acquired their eccentricities did not involve forgetting their previous semantic understanding. For example, on their present understanding of (1), they have no difficulty in remembering why they used to assent to it. They were young and foolish then, with a tendency to accept claims on the basis of insufficient reflection. By ordinary standards, Peter and Stephen understand (1) perfectly well. Although their rejection of (1) might on first acquaintance give an observer a defeasible reason to deny that they understood it, any such reason is defeated by closer observation of them. They genuinely doubt that every vixen is a vixen. Peter and Stephen are not marginal cases of understanding: their linguistic competence is far more secure than that of young children or native speakers of other languages who are in the process of learning English. If some participants in a debate have an imperfect linguistic understanding of one of the key words with which it is conducted, they need to have its meaning explained to them before the debate can properly continue. But to stop our logical debate with Peter and Stephen in order to explain to them what the word 'every' means in English would be irrelevant and

gratuitously patronising. The understanding which they lack is logical, is not semantic. Their attitudes to (1) manifest only some deviant patterns of belief. Since there clearly could have been, and perhaps are, people such as Peter and Stephen, we have counterexamples to UBl*.

It would be pointless to try to save UBl* by restricting it to rational agents. By ordinary standards, Peter and Stephen are rational agents. Although they fall short of some high standards of rationality, so do most humans. Understanding/belief links which do not apply to most humans will be of limited epistemological interest. The picture was that we can exclude those who appear to reject putatively conceptual truths from the discussion on the grounds that they lack the relevant concepts; but we cannot exclude humans who reject such truths from the discussion on those grounds if the connection between rejecting them and lacking the concepts holds only for super-humans, not for humans.

The foregoing considerations suggest a different possibility. Peter and Stephen once assented to (1). Perhaps they still have a disposition to assent to (1), masked by their later theorizing, and use 'every' and other words and modes of construction with the same senses as the rest of us because they have the same underlying inferential dispositions as the rest of us.[12] At some deep level, they have a disposition to accept (1) as true. That disposition is prevented from manifesting itself by conscious reflection at an overlying level of theory-construction, just as someone's pet views about grammar might interfere with their performance in speech while having no effect on the syntactic competence which they possess in virtue of their underlying language module. UBt* and UBl* might therefore be watered down as follows:

UDBt* Necessarily, whoever grasps the thought that every vixen is a vixen has a disposition to believe that every vixen is a vixen.

UDBl* Necessarily, whoever understands the sentence 'Every vixen is a vixen' has a disposition to assent to it.

Having a disposition to believe a proposition does not entail believing it; having a disposition to assent to a sentence does not entail assenting to it. Thus UDBt* and UDBl* are consistent with the denials of UBt* and UBl*. In particular, perhaps Peter and Stephen grasp the thought that every vixen is a vixen and understand the sentence 'Every vixen is a vixen', and have the dispositions to believe that every vixen is a vixen and to assent to 'Every vixen is a vixen', although as it happens they do not believe that every vixen is a vixen or assent to 'Every vixen is a vixen'.

The dispositional story might be used as the basis for an answer to the question 'How do we know that every vixen is a vixen?' similar to that sketched in section I. On such a view, a link between understanding and a disposition to believe a thought somehow constrains the belief to have a content on which it is true, and a link between understanding and a disposition to assent to a sentence likewise somehow constrains the sentence to have a meaning on which it is true: accordingly, UDBt* makes the thought that every vixen is a vixen true and UDBl* makes the sentence 'Every vixen is a vixen' true. Thus even the understanding/disposition-to-believe link would generate an understanding/truth link, so that the disposition to believe would be reliably truth-

directed. That reliability might itself be regarded as epistemically transparent or otherwise special enough to generate, somehow or other, a corresponding understanding/disposition-to-know link:

UDKt* Necessarily, whoever grasps the thought that every vixen is a vixen has a disposition to know that every vixen is a vixen.

UDKl* Necessarily, whoever understands the sentence 'Every vixen is a vixen' has a disposition to recognize it as true.

But are UDBt* and UDBl*, let alone UDKt* and UDKl*, even true?

There are two salient ways of filling out the dispositional story: the *personal level* account and the *sub-personal level* account. On the personal level account, the postulated dispositions require counterfactual conditionals to the effect that the person would be brought round to appropriate propositional attitudes by sufficient conscious reflection, exposure to further arguments or the like. On this view, Peter and Stephen would believe that every vixen is a vixen and accept 'Every vixen is a vixen' as true if only they thought about it more or talked to more experts. By contrast, on the sub-personal level account, the postulated dispositions can be grounded in the structure of the person's unconscious logic module or the like, even if the personal-level counterfactual conditionals are false. On this view, the default outcome of Peter and Stephen's underlying competence is belief that every vixen is a vixen and assent to 'Every vixen is a vixen', even if that default is irreversibly overridden by stable dispositions to the contrary from other sources.

An analogous contrast arises in the case of syntax. To take a standard example, native speakers of English tend to reject (3) at first sight as ill-formed:

(3) The horse raced past the barn fell.

They want to insert 'and' between 'barn' and 'fell'. But they tend to change their minds about (3) when asked to consider the result of inserting 'which was' between 'horse' and 'raced' instead: they realize that the original string was well-formed after all; 'the horse' can be the object rather than the subject of 'raced'. Conversely, native speakers often unreflectively accept ill-formed strings as well-formed, for example when a verb in the plural is separated from its subject in the singular by a long intervening string which includes a noun in the plural, but can be brought to acknowledge their mistake, for example when a draft of a paper is corrected. On a personal level account, such conscious reflective judgments, actual or counterfactual, are necessary for well-formedness. On the contrasting sub-personal level account, such judgments play a merely evidential role: what constitutes well-formedness is the structure of the syntactic component of the unconscious language module, even if the person's conscious reflective judgment is irreversibly contrary as a result of extraneous factors, such as their dogmatic commitment to a pet theory of syntax.

The personal level account fails to shield UBtd* and UBld* from the counterexamples of Peter and Stephen. For, by hypothesis, their refusal to accept (1) as true is stable under conscious reflection, exposure to further arguments and the like. We may assume that as they became comfortable with their deviant theories they gradually

ceased to feel even an initial inclination to assent to (1), although they still remember what it was like to feel such an inclination. They assimilate the change to one in which education gradually eradicates the tendency to make a particular false assumption. On the personal level account, they are *not* disposed to accept (1) as true. Perhaps that makes them irrationally obstinate, but not more so than most philosophers and many other people. Such a degree of obstinacy in defence of a favourite view is frequently combined with possession of the concepts and understanding of the words and modes of construction at issue.

The sub-personal level account has more room for manoeuvre in defence of UDBt* and UDBl*. For it can insist that even though Peter and Stephen's personal refusal to accept (1) as true is stable under conscious reflection, exposure to further arguments and so on, they still retain a disposition to accept (1) as true in virtue of features of their unconscious logic module. This view requires the postulated module to have a structurally distinct existence, for if it consisted only in acquired habits of reasoning, Peter and Stephen's earlier habits would eventually be erased by their later ones, and the disposition to accept (1) as true would disappear. Moreover, the module must include a component for deductive reasoning, since that is the kind of reasoning relevant to (1). After all, if the grounds for accepting (1) as true were merely inductive — that we have never observed a vixen which was not a vixen — then some people who understood (1) could reasonably refuse to accept it as true on the grounds that they had observed too few vixens to be in a position to judge. A *prima facie* attractive conjecture is that the deductive component of the reasoning module would comprise basic rules for natural language connectives similar to the introduction and elimination rules in a

Gentzen-style system of natural deduction. But do humans have an unconscious logic module of the required sort?

One might suppose the primary adaptive value of a cognitive module to be its capacity to perform a specific type of useful information-processing quickly and reliably enough for the purposes of action in a changing environment. Its design can exploit special features of the type of task to which it is dedicated, in order to achieve efficiencies which would be impossible for a general purpose central processing unit. A diversion through higher mental processes, in particular through consciousness, would be slower and less reliable. Thus one might expect unconscious modular deductive reasoning to pay its way by the speed and reliability of its results, just as modules for vision and natural language processing appear to do. Naturally, performance would tail off as the complexity of problems increased, but there should be good performance over a worthwhile range of non-trivial problems. Is that prediction borne out?

Evidence from empirical psychology, amassed over several decades, suggests that most humans are strikingly bad at even elementary deductive reasoning, a finding which should come as no surprise to those who have taught introductory logic. For example, in the combined results of over 65 large-scale experiments by different researchers on simple conditional reasoning, although 97% (not 100%!) of subjects endorsed modus ponens (if *p* then *q*; *p*; therefore *q*), only 72% endorsed modus tollens (if *p* then *q*; not *q*; therefore not *p*), while as many as 63% endorsed the fallacy of affirming the consequent (if *p* then *q*; *q*; therefore *p*) and 55% endorsed the fallacy of denying the antecedent (if *p* then *q*; not *p*; therefore not *q*). When the antecedent is negative, affirming the consequent overtakes modus tollens in popularity.[13] In some cases, when a further

premise of the form 'If *r* then *q*' is added to modus ponens only a minority endorses the inference (Byrne 1989). Similar phenomena arise for elementary syllogistic reasoning.

Performance greatly improves when the conditional premise in a reasoning task has a realistic deontic content, such as 'If you use a second class stamp, then you must leave the envelope unsealed' (Wason and Shapiro 1971, Manktelow and Over 1987). In general, the real-life credibility or otherwise of premises and conclusion strongly influences judgements of validity and invalidity.

For simple problems in formal deductive reasoning, when the specific subject matter provides no helpful clues, success is significantly correlated with intelligence, in whatever sense it is measured by IQ tests, SAT scores or the like (Stanovich and West 2000). For some simple tasks, success is rare except amongst those with the intelligence of able undergraduates (Newstead *et al.* 2004). Contrast this with the efficient success which humans typically show in judging whether short strings of words constitute well-formed sentences of their native language, for example. There is little sign of anything like a module for formal deductive reasoning.

A widespread although not universal view amongst psychologists of reasoning is that humans have two reasoning systems. In the terminology of Stanovich and West, System 1 is associative, holistic, automatic, relatively undemanding of cognitive capacity, relatively fast, and acquired through biology, exposure and personal experience; its construal of reasoning tasks is highly sensitive to personal, conversational and social context. System 2 is rule-based, analytic, controlled, demanding of cognitive capacity, relatively slow, and acquired by cultural and formal tuition; its construal of reasoning tasks is rather insensitive to personal, conversational and social context.[14] System 1 lacks

the formal rules which enable deductive reasoning to succeed in the absence of helpful clues from the content of premises and conclusion. It is defeasible and only moderately reliable, but it performs an important role in tasks of the kind for which it presumably evolved, such as integrating new information from perception or testimony with our standing beliefs. System 1 is not a system for formal deductive reasoning. A suitably educated, highly intelligent person can achieve success in formal deductive reasoning by means of System 2, but its structure is not that of an unconscious, sealed-off module.

We can apply this picture to Peter and Stephen. With respect to System 1, they do not deviate from the human norm. They are slightly unusual with respect to System 2, but it is in any case much more sensitive than System 1 to specific features of the individual's intelligence and education. But neither high intelligence nor a good education is needed for understanding simple sentences such as (1). Any putative System 2 differences at issue between Peter or Stephen and average speakers of English are entirely consistent with Peter and Stephen's competence in their native language. If Peter and Stephen do have any underlying disposition to accept (1) as true, it concerns their System 1. But System 1 is only distantly related to the truth-conditions of sentences, for its overall performance in assessing whether arguments are truth-preserving is very poor. Many System 1 judgments of such matters have to be overruled by System 2. An underlying System 1 disposition to accept (1) as true would therefore be quite insufficient to generate an understanding/truth link of the sort needed to advance from an understanding/disposition-to-believe link such as UDBl* to an understanding/disposition-to-know link such as UDKl*. Thus the sub-personal level account fails to deliver a disposition to recognize (1) as true.

The two systems picture has not been conclusively established; it may well need revision. Nevertheless, it throws into relief the empirical assumptions on which the sub-personal level account depends, and their clash with much current thinking in the psychology of reasoning. If the two systems picture is right to even a first approximation, the sub-personal level account is in trouble.

One might wonder how System 1 or any other system could evaluate deductive arguments without use of something like formal rules for reasoning with logical constants in natural language, even if their effect is almost swamped by associations, heuristics and other pragmatic factors (perhaps as in Braine and O'Brien 1991). But there are alternatives. In particular, one of the main current psychological theories of deductive reasoning is the *mental models* approach. Two of its leading proponents write:

> The evidence suggests that it [the reasoning mechanism] is *not* equipped with logical rules of inference, which it sometimes uses correctly and sometimes misuses, misapplies or forgets. This analogy with grammar, which has seduced so many theorists, is a mistake. The reasoning mechanism constructs a mental model of the premises, formulates a putative conclusion, and tests its validity by searching for alternative models in which it is false. The search is constrained by the meta-principle that the conclusion is valid only if there are no such models, but it is not governed by any systematic or comprehensive principles.
> (Johnson-Laird and Byrne 1993: 178)

Thus subjects may erroneously classify an invalid argument as valid, because the unrepresentative sample of models which they have examined includes no counter-model, and they wrongly treat the sample as representative. They may erroneously classify a valid argument as invalid, because, although the representative sample of models which they have examined includes no counter-model, they do not treat the sample as representative. Background beliefs about the specific subject matter of an argument influence its classification because they focus attention on some mental models rather than others. Johnson-Laird and Byrne argue that their theory gives the best fit to the empirical data.

On the mental models approach, the nearest one normally comes to employing deductive rules of inference is in the procedures for evaluating sentences (premises or conclusions) with respect to a given model, itself conceived as a mental representation.[15] But that process does not involve deductive reasoning in a natural language. Nor need it involve anything like natural deduction rules, in or out of a natural language. For example, in the case of conditionals, it need not involve modus ponens or conditional proof (standard proofs of formalizations of (1) use the latter rule). More generally, it might involve something closer to an imaginative analogue of the processes which issue in complex perceptual judgements such as 'Everybody over there is wearing a hat'. Not all such universally quantified conclusions are reached by deduction from further premises. One might employ this argument:

A is wearing a hat.

B is wearing a hat.

C is wearing a hat.

Everybody over there is A, B or C.

Everybody over there is wearing a hat.

But of course the final premise is itself a universally quantified perceptual judgment. To suppose that it too was reached as the conclusion of a deductive argument is merely to embark on a futile regress.

Although the mental models theory does not apply to all human reasoning — for example, to the kind which some humans learn to carry out in logic classes — it may apply to a considerable proportion of it. In any case, the theory is a salutary reminder that reasoning with the usual logical constants need not be formal deductive reasoning, and that the empirical evidence suggests that in humans it usually is not.

To summarize: The case for treating rejection of (1) as a failure to manifest linguistic competence depends on the status of (1) as an elementary truth of deductive logic. But human deductive competence is far more sensitive than linguistic competence to high intelligence and advanced education. Deductive competence is a reflective skill, often painfully acquired and under one's personal control. It is not insulated from one's conscious theorizing. Thus deductive proficiency is not a precondition of linguistic competence. One can understand (1) as a sentence of English without assenting or being disposed to assent to it.

So far, the argument of this section has been mainly at the level of language rather than of thought. It has been directed primarily against UBl* and UDBl*, and therefore against UKl* and UDKl*, rather than against UBt* and UDBt*, and therefore against UKt* and UDKt*. Could a theorist of thought admit that Peter and Stephen understand (1) as a sentence of English without assenting or being disposed to assent to it, but nevertheless insist that they do so by means of different concepts from those which ordinary speakers of English employ? The suggestion might be that Peter and Stephen do not grasp the thought that every vixen is a vixen, that is, the thought which *we* express with (1); instead, they grasp only some other thought which *they* express with (1). Thus they would not be counterexamples to UBt* and UDBt*, even though they are acknowledged as counterexamples to UBl* and UDBt*. Might there be consequent hope for UKt* or UDKt*, although not for UKl* or UDKl*, after all?

In effect, the envisaged theorist of thought treats (1) as indexical, expressing different propositions (thoughts) in different contexts. Many semanticists would accept that (1) is context-sensitive, for example because 'every' ranges over different domains of quantification in different contexts, but that form of context-sensitivity is only marginally relevant to the present concerns.[16] Our theorist of thought is proposing that (1) is context-sensitive in an unexpected way, because what proposition it expresses depends on the identity of the speaker, more specifically, on the speaker's theoretical proclivities. For the sake of argument, let us allow for such context-sensitivity. Consider UBt* and UDBt* as uttered in Peter or Stephen's context. Thus we may think of the following argument as uttered by Peter or Stephen:

> Given what has already been conceded, I understand the English sentence (1). As uttered in this context, it expresses the thought that every vixen is a vixen. Indeed, I know that (1) expresses the thought that every vixen is a vixen. Consequently, in the sense relevant to UBt* and UDBt*, I grasp the thought that every vixen is a vixen; I entertain it in doubting that every vixen is a vixen. Thus I satisfy the antecedent of UBt* and UDBt*. But I do not satisfy the consequents; I neither believe that every vixen is a vixen nor in any relevant sense am disposed to believe it. For I know that (1) expresses that thought, and I firmly reject (1), nor am I inclined to accept any other sentence which expresses that thought. Thus I counterexemplify UBt* and UDBt*, just as much as I do UBl* and UDBl*.

We should accept the argument. It relies on none of Peter or Stephen's idiosyncratic views as premises; it merely requires to be uttered by a speaker with those views in order to express the relevant thoughts, according to the form of context-sensitivity which the envisaged theorist of thought postulates. In particular, we should not object that Peter believes that every vixen is a vixen under the guise of the sentence 'If there is at least one vixen then every vixen is a vixen', for that sentence expresses a thought more complex than the thought (1) expresses; similarly for other candidate guises significantly different from (1). Thus the postulation does not save UBt* or UDBt* as schemas, since both have false instances in contexts of utterances such as Peter or Stephen's.

The theorist of thought's postulation of a special form of context-sensitivity in (1) is in any case implausible. Why should the meanings of the everyday words and modes of construction in (1) make special provision for Peter or Stephen's abstruse theoretical

speculations? For the theorist of thought's claim is *not* that they use (1) with an idiosyncratic meaning other than the one it usually has in English; it has already been conceded that they use (1) with its ordinary meaning in English. Rather, the claim is that, in virtue of its ordinary meaning in English, when (1) is used by a theoretically unorthodox speaker such as Peter or Stephen it expresses a thought other than the one it expresses as used by a theoretically orthodox speaker. That is not just unlikely in itself; it distorts the way in which Peter and Stephen hold themselves responsible to the meaning of (1) in the public language. They do not try to defend their rejection of (1) by appealing to the special context in which they are speaking; each of them acknowledges that if other speakers are correct in accepting (1) then he is incorrect in rejecting it. Part of the point of a natural language is to articulate and negotiate or resolve divergences of viewpoint. That point would be undermined by a mechanism of context-sensitivity which automatically made speakers talk past each other in case of such divergence. Peter and Stephen doubt that every vixen is a vixen. That is, they doubt exactly what we affirm. The hypothesis of a special form of context-sensitivity in (1) is to be rejected.

A theorist of thought might reply that although the sentence (1) does not *express* different thoughts as used by different speakers of English, it is nevertheless *associated with* different thoughts in the minds of different speakers. Of course, in a loose sense any normal speaker of English associates (1) with many different thoughts, for example those corresponding to all the inferences they are disposed to make from (1). 'Associated' in the proposal should therefore be read as something like 'associated most directly'. The suggestion is that Peter and Stephen do not associate (1) with the thought that every vixen is a vixen, the thought we associate (1) with; they associate (1) with some other thought,

even if the two thoughts determine the same Russellian proposition. One could simply use the word 'thought' subject to the stipulation that the inferential differences between Peter, Stephen and us *constitute* differences between the thoughts we associate (1) with. But what would be the point of such a stipulation? As seen above, the linguistic understanding of (1) we share with Peter and Stephen already suffices for them and us to articulate our disagreements in rational discourse; we are not merely talking past one another. In its small way, (1) determines a piece of the common intellectual heritage of mankind, something we share with Peter and Stephen in our very capacity to disagree with respect to it. To insist that the thought we associate (1) with nevertheless differs from the thoughts Peter and Stephen associate (1) with is to undermine Frege's requirement of the publicity of senses, and in particular thoughts. Moreover, if Peter and Stephen associate (1) with different thoughts from ours, should we not understand them better by translating their idiolects non-homophonically into ours? Presumably we should seek sentences other than (1) which we associate with the very thoughts they associate (1) with, or at least sentences we associate with thoughts more similar to the thoughts they associate (1) with than is the thought we associate (1) with, and translate their dissent from (1) in their mouths as dissent from those other sentences in our mouths. But the use of such a translation scheme would be intellectually disreputable, just because it would involve a refusal to acknowledge the full challenge which Peter and Stephen have issued to (1) in our mouths, not just in theirs. However mistaken their challenge, it is real. They are quite explicit that they are challenging our thought that every vixen is a vixen, and that we should not apply any non-homophonic translation scheme when interpreting their dissent from (1). To insist on applying such a non-homophonic translation scheme to

them in the teeth of their protests would be to treat them less than fully seriously as human beings, like patients in need of old-fashioned psychiatric treatment, whose words are merely symptoms. The claim that Peter and Stephen associate (1) with different thoughts from ours repackages our disagreement with them in a way which makes it sound less threatening than it really is. It misleadingly bundles together logical and semantic differences, without any genuine unification of the two categories. To call the logical disagreement a difference in associated 'thoughts' is an advertising trick. Since a homophonic reading of (1) in the mouths of Peter and Stephen is more faithful to their intentions than is any non-homophonic reading, they associate (1) with the same thought as we do in any relevant sense of 'thought'.

Naturally, when Peter dissents from 'Every F is a G', we may decide in the light of his logical unorthodoxy to store only the information that either not every F is a G or there are no Fs. But this is not a non-homophonic *translation*, any more than it is when someone notorious for exaggeration says 'At least six thousand people went on the march' and we decide to store only the information that at least one thousand people went on the march. By 'six thousand' the speaker did not mean what we mean by 'one thousand'. If exactly one thousand people went on the march he spoke falsely, not truly, for he was speaking English. Since we do not fully trust him, when he asserted one thing we stored only something weaker. Similarly, since we do not fully trust Peter, we store something weaker than he asserts. If there were no Fs, he spoke falsely, not truly, for he was speaking English. Our lack of trust in Peter and Stephen's logical proclivities is entirely consistent with reading their utterances homophonically.

Peter and Stephen are counterexamples to UBt* and UDBt* as uttered by us, just as they are to UBt* and UDBt* as uttered by them. *A fortiori*, UKt* and UDKt* fail in both contexts too. At both the level of thought and the level of language, one can understand (1) without recognizing it as true or even having a disposition to do so.

A further watering-down of the epistemological claims might be proposed. For one might regard Peter and Stephen as wilfully and perversely turning their backs on knowledge which is available to them: although they do not know and are not even disposed to know, their understanding puts them *in a position* to know. It offers them the knowledge, but they may refuse to accept it. On this view, UKt* and UDKt* should be replaced by UPKt*, and UKl* and UDKl* by UPKl*:

UPKt* Necessarily, whoever grasps the thought that every vixen is a vixen is in a position to know that every vixen is a vixen.

UPKl* Necessarily, whoever understands the sentence 'Every vixen is a vixen' is in a position to recognize it as true.

The intended point of the phrase 'in a position to' here is that the knowledge is readily available, not merely that some non-observational psychological route of unspecified complexity leads ultimately to the knowledge in question, for the latter reading makes UPKt* and UPKl* too weak to be of much interest.

But do Peter and Stephen really satisfy UPKt* and UPKl*? The mere existence of an a priori argument for (1) in their language does not put *them* in a position to recognize

(1) as true, for they cannot appreciate the argument's force. From the theoretical positions into which they have got themselves, they cannot give (1) more than insincere or superficial verbal assent. Psychologically, they are now incapable of simply walking away from their theoretical commitments. Although sustained psychological and social pressure might cause a change of view, it does not follow that they are *presently* in a position to recognize (1) as true. Such pressure might in any case convince Peter and Stephen that every vixen is a vixen by a causal route too insensitive to truth for the belief to constitute knowledge. They have blinded themselves, but the upshot is still that they cannot see. Although they can still grasp the content of (1), they have lost their insight into its truth. Even UPKt* and UPKl* miss this point. One can grasp the thought that every vixen is a vixen without being in a position to know that every vixen is a vixen, and one can understand the sentence 'Every vixen is a vixen' without being in a position to recognize it as true.

Do at least the understanding/justification links survive for (1)? Consider the analogues of UJBt and UJBl:

UJBt* Necessarily, whoever grasps the thought that every vixen is a vixen is justified in believing that every vixen is a vixen.

UJBl* Necessarily, whoever understands the sentence 'Every vixen is a vixen' is justified in assenting to it.

Do Peter and Stephen satisfy UJBt* and UJBl*? That is, are they justified in believing that every vixen is a vixen (even though they do not in fact believe that every vixen is a vixen), and are they justified in assenting to (1) (even though they do not in fact assent to it)? They are justified at most in a weak sense, for, as just argued, they are not in a position to know that every vixen is a vixen, nor are they in a position to recognize (1) as true. Moreover, they have reflected on the matter as carefully as they can, and come to a considered rejection of (1). Of course, their problem is not that they lack empirical evidence for (1). Nevertheless, if they were to accept (1) as true, and believe that every vixen is a vixen, they would be acting in a way which looks deeply irrational to them. From an external perspective, we may note that (1) has probability one on their evidence, simply because it is a logical truth, and therefore deserves to be believed. But justification rather than knowledge is the central epistemological question only for internalist theories; they try to concentrate on factors which are available from the subject's own point of view. From the point of view of Peter or Stephen, the logical truth of (1) is not available, given their deep theoretical orientation. To exclude such matters of theoretical orientation from the subject's point of view by insisting that logical truths are always available from that point of view is to give up much of the spirit of internalism, if not the letter, and in a way which threatens to undermine the motivation for what is left. Thus UJBt* and UJBl* are unlikely to survive in a form which makes them a satisfying substitute for UKt* and UKl* or even for UPKt* and UPKl*.[17] In any case, the concern of this paper is primarily with knowledge. For unless a radical form of scepticism holds, *we* know that every vixen is a vixen, even though Peter and Stephen do not.

III

It is time to apply the lessons of section II to a supposed paradigm of analyticity, such as the sentence used in section I:

(4) Every vixen is a female fox.

Given that 'vixen' is synonymous with 'female fox', (4) results from substituting synonyms for synonyms in the logical truth (1). Thus (4) is what is sometimes called 'Frege-analytic', without itself being a logical truth.[18] In section II we saw that the understanding/belief links fail for (1), even in watered-down versions, and therefore do not explain our knowledge that every vixen is a vixen. If the understanding/belief links do no better for (4), then our knowledge that every vixen is a female fox also cannot be explained along the lines of the programme sketched in section I.[19]

One might argue that the understanding/belief links for (4) reduce to those for (1), because (1) and (4) express the very same thought; thus UBt and UBl are equivalent to UBt* and UBl* respectively. Consequently, since Peter and Stephen are counterexamples to UBt* and UBl*, they are automatically counterexamples to UBt and UBl too.

At the level of thought, the argument is simply that, necessarily, one grasps the thought that every vixen is a vixen if and only if one grasps the thought that every vixen is a female fox (because it is the very same thought) and one believes that every vixen is a vixen if and only if one believes that every vixen is a female fox (because it is the very

same belief). Thus the antecedent and consequent of UBt are necessarily equivalent to the antecedent and consequent of UBt* respectively, so UBt is equivalent to UBt*.

At the level of language, the argument must be slightly more complicated, because (1) and (4) are distinct sentences even if they express the same thought. More specifically, the antecedent of UBl* does not entail the antecedent of UBl. Someone can understand (1) without understanding (4): consider, for instance, a native speaker of another language who is learning English; she has mastered the construction 'Every … is a ---', and understands the word 'vixen' through being taught it as a synonym for a word of her native language, but has not yet encountered the words 'female' and 'fox'. Nor do the consequents of UBl* and UBl entail each other. Someone who understands neither (1) nor (4) can assent to one of them without assenting to the other, on the testimony of someone else who tells her that the former is true without telling her that the latter is true. Nevertheless, one might try arguing for the equivalence of UBl and UBl* themselves, as follows.

Suppose that UBl* holds. Consider a possible situation in which someone understands (4); thus she understands the expressions 'vixen' and 'female fox', because they are constituents of (4); therefore, since those expressions are synonymous, she treats them as intersubstitutable *salva veritate* in non-quotational contexts which she understands; since 'female fox' occurs in (4) in a non-quotational context which she understands in virtue of understanding (4), and (1) is the result of substituting 'vixen' for 'female fox' in (4), she assents to (1) if and only if she assents to (4); but she also understands (1), because it is composed entirely out of words ('vixen') and modes of construction ('every … is a ---') which she understands in virtue of understanding (4);

therefore, by UBl*, she assents to (1); consequently, she assents to (4). Therefore UBl holds. Thus UBl* entails UBl. Conversely, suppose that UBl holds. Consider a possible situation S in which someone understands (1). Thus she understands the word 'vixen' and the construction 'every … is a ---'. She may not understand (4), because she may not understand 'female' or 'fox'. However, we can consider another situation Sf like S except that, in Sf, she understands 'female', 'fox' and 'female fox'. In Sf she still understands the word 'vixen' and the construction 'every … is a ---'. Moreover, since the words 'female' and 'fox' do not occur in (1), in Sf she assents to (1) if and only if in S she assents to (1). In Sf, since 'vixen' and 'female fox' are synonymous expressions which she understands, she treats them as intersubstitutable *salva veritate* in non-quotational contexts which she understands; since 'female fox' occurs in (4) in a non-quotational context which she understands, and (1) is the result of substituting 'vixen' for 'female fox' in (4), she assents to (1) if and only if she assents to (4); but she understands (4), because it is composed entirely out of words and modes of construction which she understands; therefore, by UBl, she assents to (4); consequently, she assents to (1) — in Sf. But, as already noted, she assents to (1) in Sf if and only if she assents to (1) in S, the original situation. Hence, in S, she assents to (1). Therefore UBl* holds. Thus UBl entails UBl*. So the understanding/belief links for (1) and (4) are equivalent at the level of language too.

The preceding arguments for the equivalence of UBt and UBl with UBt* and UBl* rely on controversial assumptions. Burge (1978) has built on a point of Mates (1952) to argue that synonyms cannot always be substituted for synonyms *salva veritate* in belief ascriptions. Thus someone (not Peter or Stephen) under the misapprehension

that the term ‘vixen’ also applies to immature male foxes may believe that every vixen is a vixen without believing that every vixen is a female fox. Burge argues powerfully against attempts to reconstrue such beliefs as meta-linguistic. If his interpretation of such examples is right, the argument above for the equivalence of UBt and UBt* rests on a false premise.

Can one reply that whoever believes that every vixen is a vixen *ipso facto* believes that every vixen is a female fox, even if they understand (4) without assenting to it? But if they understand (4), they know that it means that every vixen is a female fox, and presumably therefore that it is true if and only if every vixen is a female fox; if they believe that every vixen is a female fox, why do they not conclude that (4) is true, and accept it as such? The question is not unanswerable. Even good logicians may be unable to draw valid inferences when the same proposition is presented to them under different guises, for example by the sentences (1) and (4).[20]

Whether or not one accepts Burge’s view of attitude ascriptions, one may argue that a subject can understand two synonymous expressions without knowing them to be synonymous.[21] A subject may assent to (1) while dissenting from (4) and still count as having attained at least a minimal level of understanding of all the relevant words and modes of construction, enough to use them to express non-metalinguistic propositional attitudes. If so, the argument above for the equivalence of UBl and UBl* rests on a false premise.

Fortunately, we need not resolve those disputes for present purposes; we can remain agnostic over the cogency of the objections to the arguments for the equivalence of UBt and UBl with UBt* and UBl* respectively. For we can deploy Peter and Stephen

directly as counterexamples to UBt and UBl themselves, as well as to UBt* and UBl*, without settling the general question of equivalence. They *do* use 'vixen' and 'female fox' interchangeably, at least in non-quotational contexts which they understand. Even if 'vixen' and 'female fox' were equivalent by explicit stipulative definition, that would be no objection to the present argument. Peter and Stephen's unorthodoxy concerns general issues of logic, not what it is to be a vixen. By the argument of section II, they understand (1); but their understanding of (4) is as good as their understanding of (1); therefore, they understand (4). They manifestly refuse assent to (4), just as they manifestly refuse assent to (1). Thus they are counterexamples to UBl. Similarly, by the argument of section II, they grasp the thought that every vixen is a vixen; but their grasp of the thought that every vixen is a female fox is as good as their grasp of the thought that every vixen is a vixen (whether or not it is strictly the same thought); therefore, they grasp the thought that every vixen is a female fox. By the argument of section II again, they do not believe that every vixen is a vixen; but they are no more inclined to believe that every vixen is a female fox than they are to believe that every vixen is a vixen (whether or not is strictly the same belief); therefore, they do not believe that every vixen is a female fox. Thus Peter and Stephen are counterexamples to UBt too. Crudely: (1) is at least as obvious a truth as (4). So if the understanding/belief links fail for (1), they will not hold for (4).

Since Peter and Stephen are counterexamples to the understanding/belief links for (4), *a fortiori* they are counterexamples to the corresponding understanding/knowledge links, UKt and UKl. Moreover, the attempt to replace the understanding/knowledge links for (4) by understanding/disposition-to-know links analogous to UDKt* and UDKl*, or by understanding/position-to-know links analogous to UPKt* and UPKl*, faces obstacles

exactly analogous to those considered in section II. A fallback to understanding/justification links for (4) is equally unhelpful, just as it was for (1).

The underlying style of argument is quite general; it does not depend on any special feature of (1) or (4) as a candidate for links of the kinds at issue. Of course, which forms of logical unorthodoxy are relevant depends on the form of the given candidate for analytic or conceptual truth. But with a little ingenuity one can always find or invent some relevant forms of logical unorthodoxy. Often, when the understanding/belief links are shown to fail for some candidate conceptual truth by reference to some form of logical unorthodoxy, there is a natural alternative candidate conceptual truth, a watered-down sentence for which the understanding/belief links cannot be shown to fail by reference to that form of logical unorthodoxy. A different form of logical unorthodoxy is needed to show that the new candidate fails: but there will be such a different form.

The attempt to base the epistemology of obvious truths such as (1) and (4) on preconditions for understanding them, as envisaged in section I, rests on a false conception of understanding. For even the simplest candidates for analyticity or conceptual truth, understanding is consistent with considered rejection.[22]

Someone might accept the conclusion so far with respect to analytic or conceptual *truth* but resist its generalization to the analytic *validity* of inference rules ('conceptual connections'). For in treating 'vixen' and 'female fox' as intersubstitutable *salva veritate* in non-quotational contexts, do not Peter and Stephen accept the inference from '*a* is a vixen' to '*a* is a female fox' (and its converse)? Thus they do not counterexemplify a putative link between understanding the words and accepting the inference. However, the

preceding considerations destabilize such a conception of analytic inference rules without analytic truths in at least two ways.

First, the preceding considerations show that if Peter and Stephen accept the inference from '*a* is a vixen' to '*a* is a female fox', then they must reject other inference rules more central to logic. Stephen must reject the rule of conditional proof, the standard introduction rule for the material conditional →, since it would take him to the conclusion '*a* is a vixen → *a* is a female fox', which he rejects as a truth-value gap when '*a*' denotes a borderline case for 'vixen' and 'female fox'. Consequently, Stephen counterexemplifies the putative understanding/acceptance link for the rule of conditional proof. Unlike Stephen, Peter accepts that singular conditional, but he must reject the rule of universal generalization, the standard introduction rule for the universal quantifier, since it would take him on to the conclusion (4) itself, which he rejects. Consequently, Peter counterexemplifies the putative understanding/acceptance link for the rule of universal generalization. Thus the proposed conception of analytic inference rules without analytic truths omits logically central inference rules.

Second, once Stephen applies his logical unorthodoxy to his meta-language, his original acceptance of the inference from '*a* is a vixen' to '*a* is a female fox' is undermined. For in order to accept it as valid once he reflects on the question, he must accept that it is truth-preserving. More specifically, he must assent to the conditional '"*a* is a vixen" is true → "*a* is a female fox" is true'. But just as both '*a* is a vixen' and '*a* is a female fox' are borderline in some cases, so both '"*a* is a vixen" is true' and '"*a* is a female fox" is true' are borderline in some cases: the latter cases are the same as the former if a disquotational principle holds for true, but even if Stephen doubts the

disquotational principle, he acknowledges that there are cases of the latter kind, in view of higher-order vagueness. Thus he cannot reflectively endorse the inference from '*a* is a vixen' to '*a* is a female fox', even though in practice he uses 'vixen' and 'female fox' interchangeably. This result may be the basis for a philosophical objection to Stephen's views, but for reasons already seen it does not show that he lacks understanding of the relevant words or grasp of the relevant concepts. Thus Stephen counterexemplifies the understanding/acceptance link even for the inference for which it was originally proposed. Thus the envisaged conception of analytically valid inference rules ultimately fares no better than did the conception of analytic truths.[23,24]

IV

Old theories tend to survive refutation in the absence of new theories to take their place. Despite all the evidence against the existence of understanding/belief links, it can be hard to resist the idea that there *must* be such links, otherwise the distinction between understanding and not understanding would dissolve: speakers who all understood the same term might have nothing substantive in common to constitute its shared meaning. For example, in the case of moral vocabulary, which he treats as representative, Frank Jackson writes:

> Genuine moral disagreement, as opposed to mere talking past one another, requires a background of shared moral opinion to fix a common, or near enough

> common, set of meanings for our moral terms. We can think of the rather general principles that we share as the commonplaces or platitudes or constitutive principles that make up the core we need to share in order to count as speaking a common moral language. (1998: 132)[25]

Jackson's only argument for these claims is failure to see an alternative. But there is an alternative.

What binds together uses of a word by different agents or at different times into a common practice of using that word with a given meaning? This is an instance of a more general type of question: what binds together different events into the history of a single complex object, whether it be a stone, a tree, a table, a person, a society, a tradition, or a word? In brief, what makes a unity out of diversity? Rarely is the answer to such questions the mutual similarity of the constituents. Almost never is it some invariant feature, shared by all the constituents and somehow prior to the complex whole itself — an indivisible soul or bare particular. Rather, it is the complex interrelations of the constituents, above all, their causal interrelations. Although we should not expect a precise non-circular statement of necessary and sufficient conditions for the unity in terms of those complex interrelations, we have at least a vague idea of what it takes. The similarity of the constituents is neither necessary nor sufficient; different constituents can play different but complementary roles in constituting the unity: both events in the head and events in the heart help constitute the life of a person. The claim that a shared understanding of a word requires a shared stock of platitudes depends on the assumption that uses of a word by different agents or at different times can be bound together into a

common practice of using that word with a given meaning only by an invariant core of beliefs. But that assumption amounts to one of the crudest and least plausible answers to the question of what makes a unity out of diversity. In effect, it assumes that what animates a word is a soul of doctrine.

That different speakers can make asymmetric contributions to binding together different uses of a word into a common practice of using it with a given meaning was one of Putnam's best insights. After describing the role of scientific experts in fixing the reference of natural kind terms, he makes this conjecture:

> HYPOTHESIS OF THE UNIVERSALITY OF THE DIVISION OF LINGUISTIC LABOR: Every linguistic community exemplifies the sort of division of linguistic labor just described: that is, 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. (1975: 228)

Even if Putnam oversimplified the relation between natural kind terms in natural language and scientific theory, a more refined account will still respect the division of linguistic labour, for distinctions between levels of expertise are observable even within the pre-scientific use of natural kind terms.

Putnam's insight is relevant far beyond the class of natural kind terms, as Burge observed (1986). Even where we cannot sensibly divide the linguistic community into experts and non-experts, the picture of a natural language as a cluster of causally

interrelated but constitutively independent idiolects is still wrong, because it ignores the way in which individual speakers defer to the linguistic community as a whole. They use a word as a word of a public language, allowing its reference in their mouths to be fixed by its use over the whole community.[26] Such verbal interactions between speakers can hold a linguistic practice together even in the absence of a common creed which they are all required to endorse.

Evidently, much of the practical value of a language consists in its capacity to facilitate communication between agents in epistemically asymmetric positions, when the speaker or writer knows about things about which the hearer or reader is ignorant, perhaps mistaken. Although disagreement is naturally easier to negotiate and usually more fruitful against a background of extensive agreement, it does not follow that any particular agreement is needed for disagreement to be expressed in given words. A practical constraint on useful communication should not be confused with a necessary condition for literal understanding. Moreover, the practical constraint is holistic; agreement on any given point can be traded in for agreement on others. The same applies to principles of charity as putatively constitutive conditions on correct interpretation: imputed disagreement on any given point can be compensated for by imputed agreement on others.[27]

It is far easier and more rewarding to discuss the existence of true contradictions with a dialetheist such as Graham Priest than intelligent design with a Christian fundamentalist or Holocaust denial with a neo-Nazi.[28] The difficulty of engaging in fruitful debate with fundamentalists or neo-Nazis cannot plausibly be attributed to some failure of linguistic understanding on their part (or ours); it arises from their wilful

disrespect for the evidence. Such difficulty as there is in engaging in fruitful debate with dialetheists provides no significant reason to attribute to them (or us) a failure of linguistic understanding. Competence with the English language no more requires acceptance of some law of non-contradiction or any other logical law than it requires acceptance of the theory of evolution or the historical reality of the Holocaust.

We cannot anticipate all our disagreements in advance. What strike us today as the best candidates for analytic or conceptual truth some innovative thinker may call into question tomorrow for intelligible reasons. Even when we hold fast to our original belief, we can usually find ways of engaging rationally with the doubter. If a language imposes conditions of understanding which exclude such a doubt in advance, as it were in ignorance of its grounds, it needlessly limits its speakers' capacity to articulate and benefit from critical reflection on their ways of thinking. Such conditions are dysfunctional, and natural languages do not impose them.[29]

There is of course a distinction between understanding a word and not understanding it. One can lack understanding of a word through lack of causal interaction with the social practice of using that word, or through interaction too superficial to permit sufficiently fluent engagement in the practice. But sufficiently fluent engagement in the practice can take many forms, which have no single core of agreement.[30]

If we accept such an account of linguistic understanding, what should we say about grasp of concepts? There is no quick generalization from the former to the latter. Different uses of the same word must be causally related, at least indirectly.[31] Creatures who are causally unrelated to us cannot have a word numerically identical with our word 'set'; at best they can have a word exactly like our word in its general syntactic, semantic

and phonetic properties. By contrast, on the usual view, they can have a concept numerically identical with our concept *set*. If different uses of the very same concept need not be bound together by causal ties, what does bind them together? Are links between the possession of concepts and belief needed here?

Similar questions arise at the linguistic level once we consider the relation of synonymy, since uses of distinct but synonymous words need not be causally related, even indirectly. Fortunately, the tradition of truth-conditional semantics is richly provided with resources for an account of synonymy, if we take it seriously as a branch of linguistics and put aside Quinean reservations. Synonymous expressions do not simply have the same extension in the actual world. With respect to each context of utterance, they have the same intension, that is, the same extension with respect to each circumstance of evaluation; equivalently, they have the same character in the sense of Kaplan (1989). Adapting Carnap's notion of intensional isomorphism (1947: 56), we may further insist that synonymous complex expressions can be mapped onto each other in ways that preserve both constituent structure and the character of corresponding constituents; thus '0 = 0' and '1 = 1' come out non-synonymous. We can even distinguish pairs of simple expressions such as 'and' and 'but', which make the same contribution to the truth-conditions of sentences in which they occur, by requiring synonymous expressions to have their conventional implicatures satisfied with respect to the same contexts of utterances and circumstances of evaluation. Such a fine-grained conception of synonymy makes no appeal to analyticity in the form of supposed understanding/belief links.

A criterion of synonymy along the envisaged lines does not distinguish between rigid non-indexical semantically simple terms without conventional implicatures. For instance, it does not distinguish 'Cicero' from 'Tully' or 'furze' from 'gorse' in meaning. But that is the right result, for 'Cicero' is synonymous with 'Tully' and 'furze' is synonymous with 'gorse', even if not all rational speakers with a minimal understanding of the relevant words recognize the synonymies. Some such speakers assent to 'All furze is furze' while refusing assent to 'All furze is gorse'. To insist that 'furze' and 'gorse' must therefore differ in meaning is, implausibly, to make the individuation of meanings as fine-grained as the individuation of words, and thereby in the end to lose the point that causally unrelated words can be synonymous.[32] We do better to stick with a criterion of synonymy along the lines of the previous paragraph.

Given that linguistic meanings can be individuated without appeal to supposed understanding/belief links, there is no reason to expect the individuation of concepts to appeal to such links. Attempts to argue for the existence of analytic or conceptual truths in the epistemological sense from the need to individuate linguistic meanings or concepts are hopeless.

Although we can make some sense of the analytic/synthetic distinction in epistemological terms, nothing falls on the analytic side: in this sense, there are no conceptual truths, conceptual necessities or conceptual connections. Thus philosophical questions had better not be conceptual questions, if that means questions whose answers are conceptual truths. Of course, philosophers can legitimately ask questions *about* concepts, as was done here, but with equal legitimacy they can ask questions about minds or bodies, space or time, numbers or sets, properties or relations. Reflection on how we

can know the answers to philosophical questions is liable to induce epistemological panic, and make us run to conceptual truth. The refuge is illusory. Don't panic.

Notes

* Some of the material in this paper was presented in Oxford and some derives from the 2005 Jack Smart Lecture at the ANU and the 2005 Blackwell/Brown Lectures at Brown University; I am grateful to the audiences on all these occasions for discussion, and to Brian Leftow, Ofra Magidor and Oliver Pooley for written comments.

1 Of course, it is not always clear in advance which problem is simpler. Sometimes it is easier to find a proof by mathematical induction of a stronger hypothesis than of a weaker one.

2 Add the condition 'If Hesperus exists' if needed.

3 Attempts are sometimes made to distinguish the Russellian proposition that Hesperus is Hesperus from the Russellian proposition that Hesperus is Phosphorus by treating the proper names as abbreviations of (rigidified) definite descriptions. Those who favour such attempts may wish to identify thoughts with Russellian propositions. Attempts are also sometimes made to distinguish the set of possible worlds in which Hesperus is Hesperus from the set of possible worlds in which Hesperus is Phosphorus by considering the worlds as live epistemic ('indicative') possibilities rather than as metaphysical ('counterfactual') possibilities, so that the sets are 'primary intensions' rather than 'secondary intensions' (see Chalmers 2002). Those who favour the latter

attempts can make finer-grained distinctions by identifying thoughts with complexes composed quasi-sententially out of primary intensions as the atomic constituents rather than simply with primary intensions themselves.

4 For accounts of propositional attitude ascriptions on which UKt needs such fine-tuning see Stalnaker 1984 and 1999 and Salmon 1986.

5 For a recent example of a programme of something like the envisaged kind see Boghossian 1997 and 2003. The careful qualifications there make little difference to the arguments below. Williamson 2003 replies to Boghossian 2003; the present paper takes its arguments further. The theory of concept possession developed in Peacocke 1992 and modified in many subsequent publications is related to understanding/belief links. More generally, programmes which go under titles such as 'conceptual role semantics', 'inferentialism', 'use theories of meaning' and the like tend to rely on assumptions of the kind at issue. The focus of this paper is not on some few thinkers in particular; it aims to make explicit and criticize a conception on which many contemporary philosophers still rely, often tacitly, at various points in their work.

6 In effect, Horwich 1998: 131-53 allows understanding/belief links for which the understanding/truth links fail.

7 For some criticisms of the linguistic turn which complement the considerations of this paper see Williamson 2004a.

8 This paper is not directly concerned with conceptions of analyticity as truth in virtue of meaning, although it does show that such conceptions lack the epistemological payoff which might be hoped from them. They are in any case less influential in contemporary philosophy. It is widely acknowledged that 'Vixens are female foxes' is true not simply because it means that vixens are female foxes but because it means that vixens are female foxes and vixens *are* female foxes, just as 'Vixens are hunted' is true not simply because it means that vixens are hunted but because it means that vixens are hunted and vixens *are* hunted. Of course, it is necessary that every sentence which means that vixens are female foxes is true, and merely contingent whether every sentence which means that vixens are hunted is true, but that does not show that 'Vixens are female foxes' is true in virtue of meaning in any interesting sense, even given that the sentence means that vixen are female foxes, in the absence of an independently established connection between meaning and necessity — just as the corresponding point with 'always the case' in place of 'necessary' does not show that 'Vixens are female foxes' is true in virtue of meaning, since there is no independently established connection of the right kind between meaning and time. It is necessary that every sentence which means that 97 is prime is true, but that does not show that '97 is prime' is true in virtue of meaning, even given that it means that 97 is prime. Similarly, that a sentence is synonymous with a logical truth does not show that it is true in virtue of meaning in the absence of an independent argument that logical truths are true in virtue of meaning.

9 Parenthetical numerals such as '(1)' are taken throughout to refer to sentences rather than to the thoughts which those sentences express. On a standard formalization of (1) as $\forall x(Vx \rightarrow Vx)$, one proves it by starting from an instance of the rule of assumption, $Vx \vdash Vx$, applying the standard introduction rule for $\rightarrow$, conditional proof, to discharge the premise, giving $\vdash Vx \rightarrow Vx$, followed by the standard introduction rule for $\forall$, universal generalization, to reach $\vdash \forall x(Vx \rightarrow Vx)$ (no logical truth can be derived by the usual quantifier and structural rules alone, since none of them permits the discharge of all assumptions). A formalization of (1) closer to the English original uses a binary quantifier: $\vdash$ (EVERYx(Vx; Vx)) is derivable from $Vx \vdash Vx$ in a single step by an appropriate introduction rule for EVERY.

10 Alternatively, one can imagine that Peter thinks that foxes were only recently hunted to extinction, but that his presentist conception of time implies that (2) is true only if there is now at least one vixen. Yet another alternative is that Peter is a metaphysician who denies (2) on the grounds that putative macroscopic objects such as foxes do not exist, because if they did they would have vague boundaries, and nothing can have vague boundaries (compare Horgan 1998).

11 Note that while Peter assents to the conditional 'If there are vixens, then every vixen is a vixen', Stephen does not, because it has a true antecedent and an undefined consequent, and is therefore itself undefined on the Kleene semantics. Given the qualifications in Boghossian 2003, this makes Stephen more problematic than Peter for Boghossian's programme.

12 See Martin 1994 and Martin and Heil 1998 for relevant discussion.

13 See Schroyens and Schaeken 2003; the percentages are as summarized by Oaksford 2005: 427.

14 See Stanovich and West 2000: 659, where a list is also provided of earlier authors who have proposed similar views.

15 Mental models need not be visualized (*ibid.*: 182). Johnson-Laird and Byrne also claim that human reasoning is a semantic rather than a syntactic process (*ibid.*: 180), but the significance of this claim is not entirely clear, since they treat reasoning as a manipulation of representations.

16 Some restricted domains of quantifications contain no vixens, so Peter's concern becomes less far-fetched.

17 See Williamson forthcoming for a critique of an internalist conception of justification.

18 The term 'Frege-analytic' is taken from Boghossian 1999; in this connection he refers to §3 of Frege 1950, without insisting on the historical accuracy of attributing

exactly this notion of analyticity to Frege. Quine considers this notion of analyticity (amongst others) in 'Two Dogmas'.

19 Quine 1966: 111 notes that so-called truths by definition (compare (4)) depend on prior logical truths (compare (1)).

20 See Kripke 1979 and Horwich 1998: 100-1.

21 This contradicts Dummett's claim that 'It is an undeniable feature of the notion of meaning—obscure as that notion is—that meaning is *transparent* in the sense that, if someone attaches a meaning to each of two words, he must know whether these meanings are the same' (1978: 131). A direct argument against Dummett's claim is from pairs of synonymous natural kind terms such as 'furze' and 'gorse' (Kripke 1979). For more theoretical considerations see Williamson 2000: 94-106.

22 For a similar conclusion concerning lexical competence in a shared language see Marconi 1997: 56

23 Strictly speaking, Stephen cannot even accept the circular argument from '*a* is a vixen' to itself as valid, since the same argument shows that he must reject the conditional '"*a* is a vixen" is true → "*a* is a vixen" is true' in some cases.

24 Whether a similar metalogical problem affects Peter depends on the exact form of the generalization. On his view, 'Every argument of the form "*a* is a vixen, therefore *a* is a female fox" with a true premise has a true conclusion' has a false existential commitment (because there is no argument of the form "*a* is a vixen, therefore *a* is a female fox" with a true premise) but 'Every argument of the form "*a* is a vixen, therefore *a* is a female fox" is truth-preserving' does not (because there is an argument of the form "*a* is a vixen, therefore *a* is a female fox").

25 Jackson's application of the Ramsey-Carnap-Lewis method for defining theoretical terms to moral vocabulary (and more generally in his programme of conceptual analysis) requires not merely some agreed role for moral terms but an agreed role specific enough to be uniquely instantiated: this further assumption is criticized at Williamson 2001: 629-30. Jackson's reply on this point (2001: 656) reiterates something like the assumption in the quoted passage. He further misunderstands the objection by falsely supposing that the claim that we can mean the same by a word and disagree radically about its application restricts the disagreement to what occupies the roles, rather than the roles themselves, however one imagines the latter as demarcated. For criticism of the Ramsey-Carnap-Lewis method as applied in Boghossian 2003 see Williamson 2003.

26 If the term is indexical, what is fixed by use over the whole community is not the content but the character in the sense of Kaplan 1989. For the bearing of this on communication in a vague language see Williamson 1999: 512-14.

27 Davidson famously endorses a holistic principle of charity while rejecting the analytic-synthetic distinction (e.g. 2001: 144-9). For an argument that charity should maximize imputed knowledge rather than imputed truth see Williamson 2004b: 131-47.

28 For examples of rational debate for and against a law of non-contradiction see Priest, Beall and Armour-Garb 2004.

29 W.B. Gallie's intriguing account of the positive function of 'essentially contested concepts' is relevant here; his examples are 'the concepts of a religion, of art, of science, of democracy and of social justice' (1964: 168).

30 Someone who understands a word without being disposed to utter it (perhaps because they find it obscene or unpronounceable) can still count as sufficiently engaged in the practice of using it.

31 On the metaphysics of words see Kaplan 1990.

32 Consider a community C which uses a word W with a meaning M and a hitherto causally unrelated community C* which uses a word W* with the very same meaning M. If C and C* encounter each other, each may come to understand the other's words without realizing that W and W* are synonymous. Kripke 1979 highlights such cases.

References

Boghossian, P.A. 1997. 'Analyticity', in R. Hale and C. Wright, eds., *A Companion to the Philosophy of Language*. Oxford: Blackwell.

Boghossian, P.A. 2003. 'Blind reasoning', *The Aristotelian Society*, sup. 77: 225-248.

Braine, M.D., and O'Brien, D.P. 1991. 'A theory of *if*: a lexical entry, reasoning program, and pragmatic principles', *Psychological Review*, 98: 182-203.

Burge, T. 1978. 'Belief and synonymy', *Journal of Philosophy*, 75: 119-138.

Burge, T. 1986. 'Intellectual norms and foundations of mind', *Journal of Philosophy*, 83: 697-720.

Byrne, R.M.J. 1989. 'Suppressing valid inferences with conditionals', *Cognition*, 31: 1-21.

Carnap, R. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: The University of Chicago Press.

Chalmers, D.J. 2002. 'Does conceivability entail possibility?', in T. Szabó Gendler and J. Hawthorne, eds., *Conceivability and Possibility*. Oxford: Clarendon Press.

Davidson, D. 2001. *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.

Dummett, M.A.E. 1978. *Truth and Other Enigmas*. London: Duckworth.

Frege, G. 1950. *The Foundations of Arithmetic*, trans. J.L. Austin. Oxford: Blackwell.

Gallie, W.B. 1964. *Philosophy and the Historical Understanding*. London: Chatto & Windus.

Horgan, T. 1998. 'The transvaluationist conception of vagueness', *The Monist* 81: 313 -330.

Horwich, P. 1998. *Meaning*. Oxford: Clarendon Press.

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

Jackson, F. 2001. 'Responses', *Philosophy and Phenomenological Research*, 62: 653 -664.

Johnson-Laird, P.N, and Byrne, R.M.J. 1993. 'Models and deductive rationality', in K.Manktelow and D. Over, eds., *Rationality: Psychological and Philosophical Perspectives*. London: Routledge.

Kaplan, D. 1989. 'Demonstratives: an essay on the semantics, logic metaphysics, and epistemology of demonstratives and other indexicals', in J. Almog, J. Perry and H. Wettstein, eds., *Themes from Kaplan*. Oxford: Oxford University Press.

Kaplan, D. 1990. 'Words', *Aristotelian Society* sup. 64: 93-119.

Kleene, S.C. 1952. *Introduction to Metamathematics*. Amsterdam: North-Holland.

Kripke, S. 1979. 'A puzzle about belief', in A. Margalit, ed., *Meaning and Use*. Dordrecht: Reidel.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.

Manktelow, K.I., and Over, D.E. 1987. 'Reasoning and rationality', *Mind and Language*, 2: 199-219.

Marconi, D. 1997. *Lexical Competence*. Cambridge, Mass.: MIT Press.

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

Martin, C.B., and Heil, J. 1998. 'Rules and powers', *Philosophical Perspectives*, 12: 283 -312.

Mates, B. 1952. 'Synonymity', in L. Linsky, ed., *Semantics and the Philosophy of*

Language. Urbana: University of Illinois Press.

Newstead, S.E., Handley, S.J., Harley, C., Wright, H., and Farrelly, D. 2004. 'Individual differences in deductive reasoning', *Quarterly Journal of Experimental Psychology*, 57A: 33-60.

Oaksford, M. 2005. 'Reasoning', in N. Braisby and M. Gellatly, eds., *Cognitive Psychology*. Oxford: Oxford University Press.

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

Priest, G., Beall, J.C., and Armour-Garb, B., eds. 2004. *The Law of Non-Contradiction: New Philosophical Essays*. Oxford: Clarendon Press.

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

Quine, W.V.O. 1953. *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.

Quine, W.V.O. 1966. *The Ways of Paradox and Other Essays*. New York: Random House.

Salmon, N. *Frege's Puzzle*. Cambridge, Mass.: MIT Press.

Schroyens, W., and Schaeken, W. 2003. 'A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning', *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29: 140-149.

Soames, S. 1999. *Understanding Truth*. Oxford: Oxford University Press.

Stalnaker, R.C. 1984. *Inquiry*. Cambridge, Mass.: MIT Press.

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

Stanovich, K.E., and West, R.F. 2000. 'Individual differences in reasoning: implications

for the rationality debate?', *Behavioral and Brain Sciences*, 23: 645-665.

Wason, P.C., and Shapiro, D. 1971. 'Natural and contrived experience in a reasoning problem', *Quarterly Journal of Experimental Psychology*, 23: 63-71.

Williamson, T. 1999. 'Schiffer on the epistemic theory of vagueness'. *Philosophical Perspectives*, 13: 505-517.

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

Williamson, T. 2001. 'Ethics, supervenience and Ramsey sentences'. *Philosophy and Phenomenological Research*, 62: 625-630.

Williamson, T. 2003. 'Understanding and inference'. *Aristotelian Society*, sup. 77: 249 -293.

Williamson, T. 2004a. 'Past the linguistic turn?', in B. Leiter, ed., *The Future for Philosophy*. Oxford: Clarendon Press.

Williamson, T. 2004b. 'Philosophical "intuitions" and scepticism about judgement'. *Dialectica*, 58: 109-153.

Williamson, T. Forthcoming. 'On being justified in one's head', in J. Greco, A. Mele and M. Timmons, eds., *Rationality and the Good*, Oxford: Oxford University Press.