

4

Epistemological Conceptions of Analyticity

1

As observed in the previous chapter, metaphysical conceptions of analyticity do not themselves imply that linguistic or conceptual competence constrains one's attitudes to analytic sentences or thoughts. If our interest is in such constraints, we had best consider them directly. We can then assess what role, if any, they play in explaining the armchair methodology of philosophy.

If someone is unwilling to assent to the sentence "Every vixen is a female fox," the obvious hypothesis is that they do not understand it, perhaps because they do not understand the word "vixen." The central idea behind epistemological conceptions of analyticity is that, in such cases, failure to assent is not merely *good evidence* of failure to understand; it is *constitutive* of such failure. Of course, it is not by itself constitutive of failure to understand the word "vixen", since someone who understands that word may nevertheless not assent to the sentence, for example because they do not understand the word "fox"; a monolingual speaker of another language may understand "vixen" through the testimony of a bilingual without understanding any other word of English. Rather, failure to assent to the sentence can by itself only be constitutive of failure to understand the whole sentence. An unqualified link from understanding to assent is this:

(UAI) Necessarily, whoever understands the sentence "Every vixen is a female fox" assents to it.

One proposal is to generalize UAI to define an epistemological notion of analyticity: a sentence *s* is analytic just in case, necessarily, whoever

understands s assents to s . We could go further, by articulating an explicitly constitutive and not merely modal connection, but for present purposes the question is whether even this proposed necessary connection holds.

Three obvious glosses on UAI must be taken as read throughout. First, it concerns “Every vixen is a female fox” with its current meaning, for of course if the phonetically individuated sentence had meant something different, someone might easily have understood it and refused to assent. Second, assent is dispositional, for of course we are not actively assenting to any sentence whenever we understand it. Third, assent is a mental attitude, not a merely verbal one, for someone might easily understand “Every vixen is a female fox” while refusing to give it overt assent, for example because overt assent to a triviality looks uncool. We could speak of belief rather than assent, but the latter term sounds more natural in relation to inference rules, to which the notion of analyticity will be generalized.

A corresponding notion of analyticity can be defined for thoughts: a thought t is analytic just in case necessarily, whoever grasps t assents to t . If the thought *every vixen is a female fox* is analytic in this sense, then:

(UAt) Necessarily, whoever grasps the thought *every vixen is a female fox* assents to it.

On the simplest view, thinking a thought with any attitude towards it suffices for grasping it. Friends of principles like UAt should beware of straying too far from that simple view, by claiming that “full grasp” of a thought requires much more than the ability to think it (Peacocke 1992: 29–33, Bealer 1998: 221–2). For such a defence of UAt risks trivializing it, by in effect writing the consequent into the antecedent by hand. At any rate, grasp of a thought should be a matter of normal conceptual competence, just as understanding of a sentence is a matter of normal linguistic competence. We shall return to these issues below.

Call UAI and UAt “understanding-assent links” for language and thought respectively. The picture is that grasping a thought consists of grasping its constituent concepts and the way in which they have been put together just as understanding a sentence consists of understanding its constituent expressions and syntax.

Assent is no metalinguistic or metaconceptual attitude: normally, in actively assenting to “Grass is green,” one is saying or thinking that grass is green, not that the sentence “Grass is green” or the thought *grass is green* is true. However, thinking *grass is green* cannot be uncontentiously equated with thinking that grass is green. For thinking that grass is green presumably has as its object the proposition that grass is green. On a Russellian view, that proposition is made up of grass and greenness themselves, not of the concepts *grass* and *green*. Thus the thought *grass is green*, which is composed of concepts, must be distinguished from the proposition that grass is green. The thought is something like a mental vehicle for the proposition. Moreover, the same proposition can have different vehicles. For example, on this Russellian view, the proposition that Hesperus, if it exists, appears in the evening is the proposition that Phosphorus, if it exists, appears in the evening. The friend of conceptual connections is still likely to distinguish the concept *Hesperus* from the concept *Phosphorus*, and the thought *Hesperus, if it exists, appears in the evening* from the thought *Phosphorus, if it exists, appears in the evening*, on the grounds that the former embodies a conceptual connection while the latter does not. Thus understanding-assent links for thought must be articulated in terms of thoughts rather than propositions, in case there is a difference (for Fregeans, the proposition is the thought). Assenting to the thought *grass is green* is something like judging that grass is green under the guise of that thought. Similarly, assenting to the sentence “Grass is green,” for someone who understands it, is something like believing that grass is green under the guise of that sentence. More generally, in a context in which the sentence *s* expresses the proposition *p*, assenting to *s*, for someone who understands it, is something like believing *p* under the guise of *s*. For you, assenting to “I am hungry” is something like believing that you are hungry under the guise of the sentence “I am hungry,” since in your context that sentence expresses the proposition that you are hungry, not the proposition that I am hungry. Similarly, in a context in which the thought *t* expresses the proposition *p*, assenting to *t* is something like believing *p* under the guise of *t*.

The notion of an understanding-assent link can be generalized from individual sentences or thoughts to arguments at the level of language or thought. For example, if someone is unwilling to assent to the inference from “This is red and round” to “This is red,” the

obvious hypothesis is that they do not understand one of the sentences, most probably because they do not understand the word “and.” For epistemological conceptions of analyticity, failure to assent in such cases is again not merely good evidence of failure to understand but constitutive of such failure. Gerhard Gentzen introduced the idea that some rules of his natural deduction systems of logic have definitional status. Following him, a tradition which includes Dag Prawitz, Michael Dummett, Per Martin-Löf, Christopher Peacocke, Robert Brandom, Paul Boghossian and many others has developed in various ways the conception of acceptance of such inference rules as playing a constitutive role in understanding the logical constants, and therefore in understanding the sentences in which they occur. For many of these thinkers, this is one step towards a quite general “inferentialist” account of meaning and understanding for expressions in terms of their conceptual roles.¹

Understanding-assent links, or something like them, are also commonly thought to play a leading role in the understanding of theoretical terms in science: if you don’t assent to some core sentences of electron theory, in which the word “electron” occurs, you don’t understand the word, and therefore don’t understand those sentences.

A natural project is therefore to try to explain the armchair methodology of philosophy as based on something like understanding-assent links: our sheer linguistic and conceptual competence mandates assent to some sentences or thoughts and inferences, which form the starting-point for philosophical inquiry. This chapter assesses the prospects for such a project.

The envisaged method cannot accurately be characterized as “reflection on our own concepts.” For that description specifies the method only as “reflection,” which applies to virtually all forms of philosophy. Moreover, it specifies the subject matter as “our own concepts,” whereas the envisaged method involves reflection *with* our own concepts, and is therefore reflection *on* whatever our concepts

¹ The case of deductive logic is a useful reminder that many short, trivial steps of no apparent philosophical significance can be chained together into a long, non-trivial argument of obvious philosophical significance. The short steps were not really philosophically insignificant after all: no apologies for concentrating on them here.

happen to refer to – in most cases, not concepts. The idea is rather to exploit whatever epistemic assets we have simply in virtue of our linguistic and conceptual competence. Suppose that a philosopher arrives at a theory about understanding, reference, and concepts by employing a battery of general armchair techniques that rely on far more than mere linguistic and conceptual competence. Say, for definiteness, that the theory gives a crude “best fit” account of reference, and entails that justice is whatever best fits our beliefs about justice. Pretend that the theory is true. Even so, it does not follow that “Justice is whatever best fits our beliefs about justice” is epistemologically analytic. For it was not reached on the basis just of linguistic and conceptual competence. Similarly, a definition of “conceptual truth” as “truth of the theory of concepts” is unhelpful for present purposes, since it merely raises the question how the truths of the theory of concepts are known (“metaconceptual truth” would be less misleading terminology).

In what follows, we will consider more rigorously what is epistemically available simply on the basis of linguistic and conceptual competence. To a first approximation, the answer is: nothing.

2

We start with a provisional sketch of some obstacles to extracting epistemological consequences from understanding-assent links and of some attempts to overcome them. Then we turn in Section 3 to the main argument: that understanding-assent links simply do not hold.

Our concern is knowledge or justification, not just belief or assent. On the most optimistic view, understanding-assent links generate understanding-knowledge links like these:

- (UKI) Necessarily, whoever understands the sentence “Every vixen is a female fox” knows “Every vixen is a female fox.”
- (UKt) Necessarily, whoever grasps the thought *Every vixen is a female fox* knows *Every vixen is a female fox*.

Here, knowing “Every vixen is a female fox” amounts to knowing that every vixen is a female fox under the guise of the sentence “Every

vixen is a female fox,” and knowing *every vixen is a female fox* amounts to knowing that every vixen is a female fox under the guise of the thought *every vixen is a female fox*. Since knowing something entails assenting to it (we may assume), UKl and UKt entail UAl and UA_t respectively. But since assenting to something does not entail knowing it, how are understanding-knowledge links to be extracted from understanding-assent links? UAl and UA_t do not entail UKl and UKt in any obvious way.

An even more elementary problem arises. Knowledge is factive. Thus understanding-knowledge links entail corresponding understanding-truth links:

- (UTl) Necessarily, someone understands the sentence “Every vixen is a female fox” only if it is true.
- (UTt) Necessarily, someone grasps the thought *Every vixen is a female fox* only if it is true.

Thus if understanding-assent links somehow imply the corresponding understanding-knowledge links, *a fortiori* they also imply the understanding-truth links. Perhaps UTl and UTt hold because the sentence “Every vixen is a female fox” and the thought *every vixen is a female fox* are necessarily true. But in other cases the question of truth becomes more urgent.

Consider theoretical terms from discredited theories. If an understanding-assent link holds for “phlogiston,” and understanding “phlogiston” necessitates assent to a core of phlogiston theory, how could it follow that someone understands sentences of phlogiston theory only if a core of it is true? Didn’t proponents of phlogiston theory understand their own theory, despite its untruth? The example is not completely straightforward, for at least two reasons. First, it requires the untruth of the core of phlogiston theory in the understanding-assent links, not just of phlogiston theory as a whole. Some will treat a universal generalization of the form “All phlogiston is . . .” as vacuously true if phlogiston does not exist. Second, if there is nothing for “phlogiston” to refer to, one might alternatively treat sentences in which it occurs as failing to express propositions, in which case it is unclear that genuine understanding of phlogiston theory is possible. For the sake of the example, however, we may suppose that a core claim of phlogiston theory is of the form “Phlo-

giston plays role R,” that a necessary condition of understanding the term “phlogiston” is assenting to that claim, and that the claim is untrue, because nothing plays role R. Suppositions of this kind will be questioned later.

We are sometimes advised to drop various ordinary terms, on the grounds that obsolete and false folk theories are built into them. Those who offer such advice may be assuming that understanding-truth links fail for some critical sentences of the folk theory in which those terms occur while the corresponding understanding-assent links hold (if so, they presumably do not count themselves as fully understanding the folk theory). For if we can understand the critical sentences of the folk theory without assenting to them, in what sense is the theory built into the key terms? For example, we could use them to assert the negations of central principles of the theory.²

Some understanding-assent links might even be to logically inconsistent sentences or thoughts. For example, the ordinary notion of truth is sometimes held to be incoherent, on the grounds that a necessary condition for understanding “true,” and so for understanding sentences in which it occurs, is assent to a disquotational principle for “true” which the Liar paradox shows to be inconsistent. Tarski’s description of natural languages as “inconsistent” in virtue of the paradox (1983a: 164–5) may involve such a view, for if we can understand “true” in English without assenting to the troublesome instances of the disquotational principle, what prevents us from using English consistently?³ Similarly, Prior’s connective “tonk” has mismatched introduction and elimination rules; the introduction rule licenses the inference from “P” to “P tonk Q,” while the elimination rule licenses the inference from “P tonk Q” to “Q” (Prior 1960). By putting these rules together, one can derive any conclusion “Q” from any premise “P.” If assent to instances of those rules is necessary for understanding them, because necessary for understanding “tonk,” it hardly follows that the rules are truth-preserving (in the context of someone who understands “tonk”); they are so only if either every

² In effect, Horwich (1998: 131–53) allows understanding-belief links for which the understanding-truth links fail.

³ See Eklund (2002) for a defense of the idea of inconsistent languages.

sentence or no sentence of the language is true (including atomic sentences, in which “tonk” does not occur).⁴

Such examples can be interpreted in diverse ways. Nevertheless, they show at least that to advance from understanding-assent links to understanding-truth links, let alone to understanding-knowledge links, is no trivial task.

One response to the examples is to stop trying to link understanding to knowledge and truth in this way, and try only to establish links to justification, conceived as non-factive. The hope would be to reach understanding-justification links like these:

- (UJl) Necessarily, whoever understands the sentence “Every vixen is a female fox” is justified in assenting to it.
- (UJt) Necessarily, whoever grasps the thought *Every vixen is a female fox* is justified in assenting to it.

But this retreat from knowledge and truth to justification does less than full justice to the examples. Imagine a dogmatic proponent of phlogiston theory, who continues to accept it long after the accumulating negative evidence has made this unjustifiable. Suppose that “phlogiston” does indeed provide a counterexample to the putative entailment from the understanding-assent link to the understanding-truth links. Thus although understanding a core of phlogiston theory necessitates assent to that core, because understanding the core necessitates understanding the term “phlogiston” and understanding “phlogiston” necessitates assent to the core of phlogiston theory, someone can understand the core despite its untruth. But if anyone can understand the core of phlogiston theory, its proponents can. Moreover, they do not stop understanding it when they unjustifiably refuse to take seriously the mounting negative evidence. Thus our last-ditch defender of phlogiston theory understands its core but is unjustified in assenting to it: the understanding-justification links fail too. For more blatantly defective concepts, the assent mandated by understanding-assent links may be unjustifiable from the start, as with “tonk.” In

⁴ An example in which understanding is more clearly possible: Dummett (1973: 397, 454) claims that the rules for pejorative terms such as “Boche” suffer from a related kind of incoherence; Brandom (1994: 126; 2000: 69–70) and Boghossian (2003: 241–2), among others, have relied on his description of the practice of using such terms. I argue that it is mistaken in Williamson (2003a and 2008b), and suggest an alternative.

such cases too, an understanding-assent link which lacks the understanding-truth link also lacks the understanding-justification link.

Could one defend versions of UJt and UJl by qualifying the justification as *prima facie*? Consider someone who is introduced to a long list of mutually inconsistent theories of combustion, including phlogiston theory. Their content is explained without any assurance that there was ever any serious evidence for any of them. Irrationally, this person plumps for phlogiston theory and assents to its principles (unbeknownst to him, he is being influenced by happy associations from early childhood of the sound of the word “phlogiston”). By ordinary standards, he is linguistically competent with the sentences of phlogiston theory and grasps the corresponding thoughts, but he is not even *prima facie* justified in assenting to them, since he has no evidence, even by testimony, of their truth.

The examples do not motivate a retreat from knowledge and truth to non-factive justification. Rather, if they work, they show that some understanding-assent links have no positive epistemological upshot at all.

A different response to the examples is that they do not work: either the understanding-assent link fails or the understanding-truth link holds.

Since the relevant sentences or thoughts in the examples are clearly untrue, the understanding-truth link can hold in them only vacuously. That is, in such pathological cases, understanding is impossible: no meaning or concept is there to be grasped.⁵ This response seems plausible for “tonk,” for any serious attempt to apply the “tonk” rules would lead to almost immediate disaster. The envisaged response also makes the links from understanding to truth and any positive epistemic status hold vacuously. Where there is no understanding, we can hardly expect much of a positive epistemological upshot from a constraint on understanding. A trickier question is whether such possibilities of an illusion of understanding have negative epistemological repercussions for cases of genuine understanding, since a skeptical doubt can arise for the subject in the latter cases too as to whether the understanding is not an illusion. If it could avoid such repercussions, this response might maintain a general entailment from understanding-assent links to understanding-knowledge links and the rest.

⁵ See Peacocke (1992: 21) and Boghossian (2002).

However, the response is less plausible for “phlogiston” and some of the other examples than for “tonk,” since communities used the rules for “phlogiston” and “true” for years before running into any trouble.⁶ There does seem to be some sort of difference between understanding the word “phlogiston” and not understanding it. Although speakers cannot know the reference of a term if it has none, they can attain some sort of ordinary linguistic competence with it, and in that attenuated sense understand it. If such understanding of theoretical terms requires understanding-assent links in general, it is unclear why it should fail to do so for the term “phlogiston” in particular. Similarly, even if sentences with “phlogiston” fail to express propositions, because “phlogiston” fails to refer, there is still an attenuated sense in which some speakers have the empty concept *phlogiston*, an empty mental vehicle, while others do not. If such possession of theoretical concepts requires understanding-assent links in general, it is unclear why it should fail to do so for the concept *phlogiston* in particular.

Alternatively, someone might maintain that the understanding-assent links in these examples fail, but that understanding-assent links for other sentences or thoughts hold; the examples involve genuine understanding. On this view, understanding-assent links may still be held to entail the corresponding understanding-knowledge links. It claims that the examples picked the wrong candidates for understanding-assent links. Either such links hold only for non-defective words or concepts or for those defective cases they hold only for cautiously circumscribed sentences or thoughts. For instance, rather than the core of phlogiston theory itself, we might have the conditional “If phlogiston exists then . . .,” with that core filling in the dots. Arguably, however, since “phlogiston” fails to refer, that conditional too fails to express a proposition, so even this more cautious sentence is not true, although it is also not false. A more general objection is that this response treats our practices as though they are bound to have anticipated from the start all problems that could subsequently arise for them. Presumably, if understanding-assent links hold, they do so because they are built into the linguistic or conceptual practices at issue. Consider, for instance, the

⁶ Boghossian (2003: 242–3), which represents a change of view from Boghossian (2002).

hypothesis that understanding “true” necessitates assent to a disquotational principle carefully and ingeniously modified to avoid all the semantic paradoxes. Since they scarcely ever arise in ordinary life, why was our ordinary practice with the word “true” tailored in advance to avoid them? Indeed, the puzzlement they cause suggests quite the opposite. That such precautions are part of every possible linguistic or conceptual practice is even less likely. If understanding-assent links hold for some other reason than that they are built into the linguistic or conceptual practices at issue, what is that other reason? Even if one moderates the approach by substituting understanding-justification links for understanding-knowledge links, a version of the objection still applies. If our linguistic or conceptual practices can make assent to inference rules a precondition of understanding, nothing seems to stop bad practices from requiring assent to rules, like those for “tonk,” that generate consequences not involving the original word or concept at issue. Such consequences may include arbitrary pernicious dogmas (such as racist ones) for which no justification is provided. More cautious fallbacks need not even implicitly have been provided; the practice simply breaks down once the dogma is abandoned. So this alternative way of maintaining a general entailment from understanding-assent links to understanding-justification links, let alone understanding-knowledge links, is unpromising. The objections tell equally against the putative understanding-knowledge or understanding-justification links, even if no attempt is made to *derive* them from understanding-assent links.

A more moderate response concedes that defective practices give rise to understanding-assent links without corresponding links to truth or any positive epistemological status, but maintains that understanding-assent links for non-defective practices do yield such links. For instance, one might try to tell a story on which understanding-assent links for non-defective practices constrain the reference of the relevant words or concepts so that the sentences or thoughts in the links come out true (for some defective practices, this constraint cannot be met). Under such conditions, understanding-assent links generate understanding-truth links. Thus assent to those sentences or thoughts (while understanding or grasping them) is, completely reliably, assent to truths. One might hope to squeeze understanding-knowledge links out of such reliability considerations, perhaps when

enhanced by an argument that the reliability is not completely hidden from the subject. Clearly, much work would be needed to vindicate such a programme.⁷

A lazy alternative simply postulates understanding-knowledge or understanding-justification links for non-defective practices without attempting to derive them from understanding-assent links. But this has little explanatory value. I understand “Every vixen is a female fox,” and it has some positive epistemic status for me. How does it get that status? How do I know “Every vixen is a female fox”? Why am I justified in assenting to it? The lazy theorist may try to dismiss the question, saying that it is simply part of our linguistic practice that “Every vixen is a female fox” has that positive epistemic status for whoever understands it. But the examples of defective practices show that it is not simply up to linguistic practices to distribute positive epistemic status as they please. That the practice is to treat a given sentence as having some positive epistemic status for competent speakers of the language does not imply that it really has that epistemic status for them. Their belief may be untrue and unjustified, however much the practice deems otherwise. Thus the only plausible way to make the relevant practice guarantee the putative link from understanding to the positive epistemic status is by making absence of the epistemic status constitute absence of understanding, just as absence of assent was supposed to do. On this account, whoever does not know “Every vixen is a female fox” or is not justified in assenting to it *thereby* fails to understand it. But this direction of explanation does not trivialize the positive epistemic status, to which it assigns the role of constituter, not constituted. Thus the lazy theorist cannot simply dismiss the question: how does “Every vixen is a female fox” get its positive epistemic for whoever understands it? Positing direct links from understanding to knowledge or justification does not remove the need for substantive epistemology here. Indeed, it makes the armchair nature of understanding problematic. Even when the relevant sentence or thought has the positive epistemic status at issue, the reason is not simply that the linguistic or conceptual practice deems it to be so – which of course is not to say that the practice is

⁷ The treatment of the issue in Boghossian (2003) is of this general kind. For detailed criticism see Williamson (2003a).

irrelevant to its epistemic status. In any case, if understanding-assent links fail, as is argued below, then *a fortiori* so do understanding-knowledge links, and understanding-justification links turn out to fail for similar reasons.

Let us consider understanding-assent links in more depth. If they hold, with or without normative consequences, they should cast some light on the actual practice of philosophy. For if an understanding-assent link holds for a philosophically significant sentence, and we do understand it, then we do assent to it, whether or not we are justified in doing so. But the next sections argue that understanding-assent links fail even for paradigms of “analyticity.” The main focus will be on the simplest cases, since those are the ones for which understanding-assent links have the best chance: if they fail there, they fail everywhere. We will start by examining unqualified understanding-assent links, beginning at the level of language. They fail. We then consider various ways of loosening them.

3

In their classic response to Quine’s critique of the analytic-synthetic distinction, Grice and Strawson give the sentence “My neighbor’s three-year-old child is an adult” as an example of a sentence that we could not understand someone using with its ordinary literal meaning to make an assertion (1956: 150–1). That suggests an understanding-assent link for the sentence “No three-year-old child is an adult”: necessarily, whoever understands it assents to it. But the link fails. Someone may believe that normal human beings attain physical and psychological maturity at the age of three, explaining away all the evidence to the contrary by *ad hoc* hypotheses or conspiracy theories (many three-year-olds pretend to be eighteen-year-olds in order to vote, the abnormally polluted local water slows development, and so on). However foolish those beliefs, they do not constitute linguistic incompetence. Friends of analyticity will reply that the example was badly chosen. It is therefore best to start with the most elementary examples possible.

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.⁸

One may be tempted to endorse understanding-assent links for (1):

- (UAI') Necessarily, whoever understands the sentence "Every vixen is a vixen" assents to it.
 (UAt') Necessarily, whoever grasps the thought *every vixen is a vixen* assents to it.

Are UAI' and UAt' true? Consider two native speakers of English, Peter and Stephen.

Peter's first reaction to (1) is 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. Of course, Peter does not always think in such theoretical, metalinguistic terms, but he resorts to them in rationalizing and defending his

⁸ Parenthetical numerals such as "(1)" are taken throughout to refer to sentences rather than to thoughts. 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: $\vdash (\text{EVERY}x(Vx; Vx))$ is derivable from $Vx \vdash Vx$ in a single step by an appropriate introduction rule for EVERY.

pattern of assent and dissent to individual sentences. 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 MI6, which even organizes widespread fox-hallucinations, so that people will protest about fox-hunting rather than the war in Iraq. 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 it.⁹

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 ("indefinite") and using Kleene's three-valued "strong tables" (1952: 334), along the lines explained in Chapter 2. On Stephen's view, for "Every F is a G" to be true is for the conditional " x is an F \rightarrow 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 \rightarrow x is a G" to be false for some value of " x ." On his semantics, for a conditional sentence with " \rightarrow " 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 \rightarrow x is a vixen" is also neither true nor false, by the strong Kleene table for \rightarrow . 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

⁹ 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, for if they did they would have vague boundaries, which are metaphysically impossible (compare Horgan (1998)).

about borderline cases?” But his refusal to assent to (1) as true is firm.¹⁰

We may 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 that really is an elementary logical truth. Peter and Stephen do not assent to (1). Thus, according to UAI, Peter and Stephen do not understand (1) (with its standard English meaning). If so, they presumably misunderstand at least one of its constituent words or modes of combination. Is that the impression 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 “vixen,” “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, thereby expressing their belief that there were never any dinosaurs. Their problem is not that they misunderstand the word “dinosaur,” but that they have silly beliefs about evolution. Peter, like Stephen, understands the word “vixen.”

The best candidate for a word or mode of composition in (1) that Peter and Stephen misunderstand is “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 at least one F” tends to be common ground among the participants anyway. It is (usually, not always) a pragmatic presupposition in the sense of Stalnaker (1999). Pragmatically, Peter adjusts his conversation to a society that obstinately retains its belief in the existence of foxes much as members of many other small

¹⁰ 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 indefinite consequent, and is therefore itself indefinite on the Kleene semantics. Given the qualifications in Boghossian (2003), this makes Stephen more problematic than Peter for Boghossian’s program.

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 focusing on a small domain of contextually relevant objects among 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 seem 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 truth-conditional 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. But 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 are not making unilateral declarations of linguistic independence. They use "every" and the other words in (1) as words of the 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 – linguists who work on the semantics of natural languages often do the former without doing the latter. 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: even for us “Every F is a G” is not strictly synonymous with “There is no F that is not a G,” since the former does not contain negation. For Peter and Stephen, the more elaborate articulations are simply convenient records of important logical facts about “every.” Only in tricky cases do they resort to their non-standard semantic theories in evaluating non-metalinguistic claims such as (1) expresses. Their non-metalinguistic unorthodoxy as to when every F is a G is not ultimately derived by semantic descent from metalinguistic unorthodoxy as to when “Every F is a G” is true; rather, their metalinguistic unorthodoxy is ultimately derived by semantic ascent from their non-metalinguistic unorthodoxy.

Of course, the intention to use words with their normal public meanings does not guarantee success: it can fail in cases of sufficiently gross and extensive error. But that does not suggest that the intention is *irrelevant* to whether someone is using the words with those meanings. The intention is normally successful, in the absence of special defeating circumstances, just as the intention to use a proper name with the same reference as it has in the rest of the community is normally successful. The question is whether Peter and Stephen’s eccentricities are sufficiently gross and extensive to constitute defeating circumstances. By ordinary standards, they are not. Although they look gross enough when seen in isolation, they are compensated for by Peter and Stephen’s normality in other respects.

Peter and Stephen are native speakers who 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 under-

stood it, any such reason is defeated by closer observation of them. They genuinely doubt that every vixen is a vixen. Nor are Peter and Stephen 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. They joined the club of “every”-users; since they haven’t resigned or been expelled, they are still members.

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 patronizing. We cannot understand them better if we translate their word “every” by some non-homophonic expression, or treat it as untranslatable. The understanding 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 UAI’.

The argument that Peter and Stephen mean what we mean by their words exemplifies two interlocking themes: Quine’s epistemological holism, on which the epistemological status of a belief constitutively depends on its position in the believer’s whole system of beliefs, and Putnam and Burge’s semantic externalism (discussed in more detail below), on which the content of a belief constitutively depends on the believer’s position in a society of believers. Epistemological holism explains how unorthodoxy on one point can be compensated for by orthodoxy on many others, so that overall Peter and Stephen’s usage of the key terms is not beyond the pale of social acceptability; since they remain participants in the relevant linguistic practice, semantic externalism then explains how they can still use the terms with their normal public senses. But neither epistemological holism nor semantic externalism figured as *premises* of the argument. Rather, the argument appealed to features of the relevant systems of belief that make epistemological holism plausible, and to features of our ascription of beliefs that make semantic externalism plausible.

To try to save UAI’ by restricting it to rational agents would be pointless. By ordinary standards, Peter and Stephen are rational agents. Although they fall short of some high standards of rationality,

so do most humans. Understanding-assent links that do not apply to most humans would be of limited epistemological interest. The picture was that those who appear to reject analytic sentences can be excluded from the discussion because they lack the linguistic competence to engage in it; but we cannot exclude humans who reject such sentences on those grounds if the connection between rejecting them and lacking competence holds only for super-humans, not for humans.

The problem for UAI' is clearly not specific to sentences of the form "Every F is an F" Let us see how it generalizes to rules of inference.

It is often claimed that assent to arguments by modus ponens of the form "If A then B; A; therefore B" is a precondition for understanding the word "if" (Boghossian 2003, for instance). Indeed, this is a standard example in the literature. However, Vann McGee, a distinguished logician, has published purported counterexamples to modus ponens for the indicative conditional in English. Here is one of them; the others are similar:

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

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

A Republican will win the race.

Yet they did not have reason to believe:

If it's not Reagan who wins, it will be Anderson. (McGee 1985: 462)

With reasonable confidence, they combined assent to both premises of an argument by modus ponens with dissent from the conclusion, so they rejected the argument.¹¹ If McGee's examples are counterexamples to modus ponens, they are also counterexamples to the claim that assent to instances of modus ponens is necessary for understanding "if." But let us assume, with the majority, that modus ponens is

¹¹ The formulation in the text is intended to distinguish the case from examples in which speakers' confidence in each premise of a modus ponens argument is just above a probabilistic threshold which their confidence in the conclusion is just below. In McGee's case, speakers are sufficiently confident of the conjunction of the two premises.

valid, so McGee's examples are not in fact counterexamples.¹² Perhaps the conclusion was true, because Reagan won; although the poll was not misleading, our usual methods for evaluating conditionals lead us astray in this case. A currently popular objection to the examples is that they depend on an illicit shift of context, perhaps in the treatment of "If it's not Reagan who wins, it will be Anderson" between the consequent of the first premise and the conclusion.¹³ But even if some such confusion *causes* the pattern of assent and dissent to the premises and conclusion, the *effect* is that McGee and his envisaged speakers end up accepting the premises and rejecting the conclusion in a single context, when they look back on all three sentences.¹⁴ They genuinely reject a genuine instance of modus ponens.¹⁵ Such reactions do not manifest the superimposition of a perverse semantic or logical theory on native speaker intuitions; they flow from native speaker intuitions themselves in a fairly natural way, despite being mistaken.

¹² For early critical reactions to McGee's examples see Sinnott-Armstrong, Moor, and Fogelin (1986), Lowe (1987) and Over (1987). But some authors have accepted the examples (Lycan 2001: 66–7).

¹³ Recent examples of context-shifting charges include Nolan (2003: 264) and Gauker (2005: 86).

¹⁴ Contrast McGee's example with instances of modus ponens such as "I know that I have hands; if I know that I have hands then I know that I'm not a brain in a vat; therefore, I know that I'm not a brain in a vat." Many people accept the premises and reject the conclusion when they encounter them in that order. However, once they have rejected the conclusion, they are typically inclined to retract their acceptance of the first premise, not out of concern for modus ponens but because it no longer looks plausible to them in its own right, in the new context that arises once the skeptical possibility becomes relevant. For contextualists in epistemology, this is a paradigm case of context-shifting (Stine (1976), Cohen (1988), DeRose (1995), Lewis (1996); see Hawthorne (2004), Stanley (2005) and Williamson (2005b) for some critical discussion and more references). By contrast, the premises of McGee's argument continue to look plausible to those who reject the conclusion.

¹⁵ Edgington (2001: 408) suggests that McGee's example is not a genuine instance of modus ponens on the grounds that the first premise has a misleading surface form; on her view, conditionals do not express propositions, so what look like conditionals with conditional antecedents or consequents must be reinterpreted. It is doubtful that such a view is consistent with a systematic account of the structure of English sentences, which permits a wide variety of such embeddings, for example, "If it is the case that if it's not Reagan who wins it will be Anderson, then a Republican will win the race."

Does McGee not understand the English word “if”? In conversation, he appears to understand it perfectly well. By ordinary standards, he *does* understand it. Before he had theoretical doubts about modus ponens, he understood the word “if” if anyone has ever understood it. Surely his theoretical doubts did not make him cease to remember what it means. Moreover, his doubts derive from taking at face value a natural pattern of native speaker reactions to an ingeniously chosen case. If he counts as not understanding “if,” so do millions of other native speakers of English.

Could we invoke the division of linguistic labor (Putnam 1975: 228), and say that making any given inference by modus ponens is a precondition only for *full* understanding of “if,” the kind of understanding characteristic of the expert rather than the layman? The trouble is that McGee *is* an expert on conditionals. He publishes on them in the best journals. He does not defer in his use of “if” to any higher authorities. He may lack some theoretical understanding of conditionals, just as experts on neutrinos may lack some theoretical understanding of neutrinos, but none of that amounts to any lack of linguistic competence with “if” or “neutrino” at all.

Are only some arguments by modus ponens such that assent to them is a precondition for understanding “if”? Presumably, McGee will accept most arguments by modus ponens. However, any particular such argument might be rejected by another expert on conditionals, on the basis of a subtle theoretical argument. By hypothesis, the expert would be mistaken, but making a subtle theoretical error does not constitute linguistic incompetence.

The problem is not just the vagueness of natural languages. Similar problems arise for carefully constructed formal languages. Consider modus ponens for the material conditional \rightarrow , explained by the standard truth-table. It is equivalent to disjunctive syllogism: from A and $\neg A \vee B$ derive B . Technically competent relevance logicians and dialetheists such as Graham Priest reject disjunctive syllogism (Priest 1995: 5). According to him, the best account of paradoxes such as the Liar is that in special circumstances a sentence can be both true and false; one can be on different lines of the truth-table simultaneously. When A is true and false while B is merely false, the premises of disjunctive syllogism are true (for A is true; since A is also false, $\neg A$ is true, so $\neg A \vee B$ is true), while its conclusion is straightforwardly false. Whatever the errors underlying the rejection of modus

ponens for \rightarrow , they do not arise from a lack of linguistic competence with \rightarrow on the part of relevance logicians and dialetheists.

As a final example, consider the natural deduction rules for conjunction. Instances of the introduction rule are arguments of the form “A; B; therefore A and B.” Instances of the elimination rule are arguments of the converse forms “A and B; therefore A” and “A and B; therefore B.” These are just about the simplest rules for a non-trivial binary connective. One must formulate what acceptance of the introduction rule requires with particular care, since the probability of a conjunction may be less than the probability of either conjunct. Iterations of the introduction rule yield the Lottery and Preface paradoxes. Given a lottery known to have at most a million tickets and only one winner, each premise of the form “Ticket *i* will lose” is overwhelmingly probable, even though their conjunction is known to be false. The author of a book may endorse each individual statement in it, yet admit in the preface that, despite all her efforts, it is bound to contain errors, and on those grounds reject the conjunction of the individual premises. Of course, these paradoxes do not show that the introduction rule fails to preserve truth, although they might be used as grounds for rejecting the rule by a theorist who (mistakenly) used a probabilistic criterion for acceptance. The elimination rule does not suffer from these problems, since the probability of a conjunction is never higher than the probability of any given conjunct.

Let us therefore concentrate on the elimination rule for conjunction, as having the best chance of being non-discretionary for competent speakers.¹⁶ Consider Simon, whose view of vagueness resembles Stephen’s, except that Simon’s practice conforms to a semantics with Kleene’s weak three-valued tables rather than his strong ones. On these tables, a conjunction is indefinite (neither true nor false) if at least one conjunct is, irrespective of the value of the other conjunct; the same principle is applied to disjunction, the material conditional and negation (Kleene 1952: 334). Furthermore, Simon regards both

¹⁶ In discussion, Boghossian suggested conjunction elimination as a fallback example of a non-discretionary rule if modus ponens fails. Peacocke writes of the possession-condition for the concept of conjunction, “On any theory, this possession-condition will entail that thinkers must find the transition from A and B to A compelling, and must do so without relying on any background information” (2004: 172).

truth and indefiniteness as designated (acceptable) semantic values for an assertion: what matters is to avoid falsity. In a borderline case, some speakers say “Jack is bald,” others with equal vehemence say “Jack is not bald”; they may persist even when they recognize that the dispute cannot be resolved. According to Simon, both assertions are acceptable. In answer to the question “Is Jack bald?,” even the answer “He is and he isn’t” is acceptable. Although Simon does not assign the value “T” to “Jack is bald,” that metalinguistic reservation is consistent with assenting to the sentence, that is, with believing that Jack is bald under the guise of that very sentence (similarly, supervaluationists about vagueness reject the disquotational inference from “‘Jack is bald’ is not true” to “Jack is not bald”). The joint implication of Simon’s principles is that any complex sentence formed by the application of the specified operators to simpler sentences, at least one of which is borderline, has a designated value – of course, on Simon’s view, most such sentences should not be uttered, on the pragmatic grounds that they violate the conversational maxim of relevance (Grice 1989: 27). Suppose that “A” is simply false while “B” is borderline. Consequently, for Simon, “B” is indefinite, so “A and B” is also indefinite. Thus the corresponding instance of conjunction elimination – “A and B; therefore A” – has a designated premise and an undesignated conclusion. On these grounds, Simon rejects the conclusion of that instance while accepting its premise (although he points out that asserting the premise would be pragmatically misleading in most contexts, since “B” is irrelevant to its status). In other cases, he treats the premise merely as a supposition, but still rejects the deduction from it to the conclusion. Once again, this need not reflect incompetence with the English language. Conjunction elimination is no exception to the general pattern. Arguably, violations of conjunction elimination are actual, not just possible, in the Conjunction Fallacy, a much-studied, widespread and robust psychological phenomenon in which subjects assign a higher probability to a conjunction than to one of its conjuncts.¹⁷

¹⁷ The seminal paper is Tversky and Kahneman (1983). See also Kahneman and Frederick (2002), Sides, Osherson, Bonini, and Viale (2002) and Jönsson and Hampton (2006). We can also imagine speakers who reject instances of conjunction elimination through muddling truth and conversational appropriateness. “Did she take the money and give it back? Yes. Did she take the money? No, she took-the-money-and-gave-it-back.”

No given argument or statement is immune from rejection by a linguistically competent speaker. Quine's epistemological holism in "Two Dogmas" undermines his notorious later claim about the deviant logician's predicament: "when he tries to deny the doctrine he only changes the subject" (1970: 81).

Understanding words in a natural language has much to do with the ability to use them in ways that facilitate smooth and fruitful interaction with other members of the community. That ability can be realized in indefinitely various forms. Speakers can compensate for their deviance on one point by their orthodoxy on others, their ability to predict the reactions of non-deviant speakers, their willingness in the long run to have their utterances evaluated by public standards. As we have seen, such compensation is often possible when the deviance results from localized interference in the normal practice of using a word by high-level theoretical concerns. Thus there is no litmus test for understanding. Whatever local test is proposed, someone could fail it and still do well enough elsewhere with the word to count as understanding it. Could an inferentialist reply that such objections trade on a loose everyday sense of "understanding" that must be replaced by something more precise for theoretical purposes? It is far from clear that a stricter sense would do a better job. The relevant features of the ordinary conception of understanding are not mere unreflective sloppiness. Rather, they are an appropriate response to an important constraint on a theory of linguistic meanings: that there is little point in talking about them unless they can be shared across significant differences in belief, between different individuals at the same time or the same individual at different times. They can survive factual learning and factual disagreement. Although inferentialist accounts respect the letter of that constraint, they violate its underlying spirit, by setting inflexible limits to the scope for genuine disagreement. The more holistic ordinary notion of understanding permits localized disagreement at virtually any point.

Cases of logical deviance hint at ways in which the failure of individualist accounts of meaning go deeper than the immediate lessons of the original anti-individualist arguments of Putnam (1975) and Burge (1979). Their cases are often analyzed in terms of a distinction between experts with full understanding and lay-people with partial understanding who defer to the experts, in virtue of which one may

correctly ascribe to them attitudes to the contents that experts determine.¹⁸ Such asymmetries are postulated by Putnam's Hypothesis of the Universality of the Division of Linguistic Labor:

Every linguistic community . . . possesses at least some terms whose associated "criteria" are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers depends upon a structured cooperation between them and the speakers in the relevant subsets. (Putnam 1975: 228)

But, as we have seen, experts themselves can make deviant applications of words as a result of theoretical errors and still count as fully understanding their words. Although they defer to nobody on the matters at issue, they are more than adequately integrated members of the speech community with respect to those very words. Their assignments of meaning to those words are not parasitic on the assignments that more privileged individuals make. Rather, each individual uses words as words of a public language; their meanings are constitutively determined not individually but socially, through the spectrum of linguistic activity across the community as a whole. The social determination of meaning requires nothing like an exact match in use between different individuals; it requires only enough connection in use between them to form a social practice. Full participation in that practice constitutes full understanding. That is why there is no litmus test for understanding.¹⁹

¹⁸ An example is Peacocke's discussion of deference-dependent propositional attitude ascriptions (1992: 29–33). Burge (1986) extends his earlier arguments in ways related to the arguments of this chapter, in his account of the understanding of words such as "sofa," and argues for such a deeper lesson. Goldberg (2000) replies on behalf of Burge to Bach (1988) and Elugardo (1993).

¹⁹ For a related conclusion concerning lexical competence in a shared language see Marconi (1997: 56). For the relevance of the model of full understanding as full induction into a practice to the theory of vagueness see Williamson (1994a: 211–12). It is not implied that no similar issue could arise for understanding on the part of a single isolated individual, for such an individual's meanings and concepts are constitutively determined, at least in part, by their dispositions over a range of counterfactual circumstances; those dispositions and their bearings may be hard to survey from the limited standpoint of the actual circumstances.

4

Peter and Stephen understand (1) without assenting to it; UAI fails. Someone sympathetic to the spirit of understanding-assent links might concede that much while arguing that its upshot is only a superficial loosening of those links. If the deviance results only from erroneous theorizing that overlays an ordinary understanding of the terms, may not the links still hold at the underlying level?

However, we have already seen reason to doubt that deviance can only arise from theorizing extrinsic to speakers' ordinary understanding of the words. Vann McGee's examples exert an intuitive pull on native speakers, irrespective of and even contrary to their theoretical predilections. We can also imagine untheoretical native speakers whose unreflective patterns of assent and dissent to non-metalinguistic sentences are those which Peter, Stephen, and Simon respectively recommend, although they lack the reflective capacity to rationalize those patterns by appeal to formal semantic theories. They too would be able to fit in well enough with the rest of the linguistic community, to engage smoothly in useful communication and adjust to their differences with other speakers in order not to attract too much attention. They too would use their words as words of the public language, rather than declaring unilateral linguistic independence. How do we know that there are not in fact many such native speakers of English around us? Once we concede that Peter, Stephen, and Simon are competent speakers, we can hardly refuse the same classification to other speakers merely on grounds of their unacquaintance with formal semantics.

What might be claimed, in the case of both theoretical and untheoretical deviant native speakers, is that the deviance is some kind of performance error which leaves their underlying competence intact: at some basic level they have the required dispositions, which they fail to manifest as a result of interfering factors, such as computational limitations, conflicting dispositions to take cheap and dirty intellectual short-cuts, and so on. On this view, Peter and Stephen still have a disposition to assent to (1), masked by their later theorizing; they 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.²⁰ 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 linguistic competence. For untheoretical speakers, the interfering factors are unconscious, but the effect is similar. UAI' and UAt' might therefore be watered down as follows:

- (UDAI') Necessarily, whoever understands the sentence "Every vixen is a vixen" has a disposition to assent to it.
(UDAt') Necessarily, whoever grasps the thought *every vixen is a vixen* has a disposition to assent to it.

Having a disposition to assent does not entail assenting. Thus UDAI' and UDAt' are consistent with the denials of UAI' and UAt'. Do Peter and Stephen have the disposition to assent to (1) despite happening not to assent to it? If understanding is linked to such dispositions to assent in these cases, one might even try to use that to explain how it is also linked to dispositions to know, along lines similar to those sketched in Section 2. But are UDAI' and UDAt' true?

There are two salient ways to fill out the dispositional story: at the *personal level* or the *sub-personal level*. At the personal level, the postulated dispositions require something like counterfactual conditionals to the effect that sufficient conscious reflection and exposure to further arguments would bring the person to assent. Thus Peter and Stephen would assent to (1) if only they thought about it more and talked to more experts. By contrast, at the sub-personal level, the postulated dispositions are grounded in something like an unconscious reasoning module, even if the personal-level counterfactual conditionals are false. Thus the default outcome of Peter and Stephen's underlying competence is assent to (1), even if stable dispositions from other sources irreversibly override that default.

²⁰ Eklund (2002: 262) defends such a view of logical deviance. See Martin (1994), Lewis (1997), Martin and Heil (1998), Bird (1998), and Mumford (1998) for some basic issues about masked dispositions. Harman (1999: 213) relies on defeasible inferential dispositions in his conceptual role semantics.

An analogous contrast arises for syntax. As 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 “that was” between “horse” and “raced” instead: they realize that the original string was well-formed after all; “the horse” is the object, not the subject, of “raced.” Conversely, native speakers often unreflectively accept ill-formed strings as well-formed, for example when a plural verb is separated from its singular subject by a long intervening string that includes a plural noun, but can be brought to acknowledge their mistake, as when a draft is corrected. On a personal level account, such conscious reflective judgments, actual or counterfactual, are constitutive of well-formedness. On the contrasting sub-personal level account, those 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 UDAI’ and UDAr’ from the counterexamples of Peter and Stephen. For, by hypothesis, their refusal to assent to (1) is stable under conscious reflection, exposure to further arguments and so on. Like many people, not least philosophers, they are obstinate in defense of their favorite views, willing to make whatever *ad hoc* moves are needed to retain them. One knows in advance that the task of dissuading them is hopeless, however good one’s objections: a common experience in philosophy. As Peter and Stephen became comfortable with their deviant theories they gradually ceased to feel even an initial temptation to assent to (1), we may assume, although they still remember what it was like to feel such a temptation. They assimilate the change to one in which education gradually eradicates the tendency to make a particular false assumption. Perhaps years of browbeating or social ostracism would cause them to change their minds, but that applies to almost any belief; it is poor evidence that an underlying disposition to assent was present all

along. Would Peter and Stephen assent to (1) if they lacked their conscious theoretical commitments? Perhaps not, but that counterfactual would show little. The possibility of untheoretical analogues of Peter and Stephen has already been raised. They lack the conscious theoretical commitments but still do not assent to (1). If it is objected that the untheoretical analogues, unlike Peter and Stephen, do not understand (1) with its normal English sense because they lack the required unconscious cognitive structures, that is in effect to switch to the sub-personal version of the dispositional account. On the personal level account, Peter and Stephen are *not* disposed to assent to (1). If that makes them irrationally obstinate, they are no more so than many philosophers and non-philosophers in defense of a favorite view.

The sub-personal level story has more room for maneuver in defense of UDAl' and UDAt'. It can insist that although Peter and Stephen's personal refusal to assent to (1) is stable under conscious reflection and exposure to further arguments, they retain a disposition to assent to (1) in virtue of features of their unconscious logic rules. This requires the postulated rules to be encased in some sort of psychological module, for if they consisted only in general habits of reasoning, Peter and Stephen's earlier habits could eventually be erased by their later ones, and the disposition to assent to (1) would disappear. The module must include rules for deduction, since that is the kind of reasoning relevant to (1). This module may be a component of an overall semantic module (after all, we are considering (1) as a candidate for analyticity). If the grounds for assent to (1) were merely inductive – that we have never observed a vixen that was not a vixen – people who understood (1) could reasonably refuse to assent to it 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 rules would include analogues for natural language connectives of the introduction and elimination rules in a Gentzen-style system of natural deduction. But do humans have a module that includes unconscious logic rules 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 that 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 seem 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 not surprise 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 percent (not 100 percent!) of subjects endorsed modus ponens, only 72 percent endorsed modus tollens (if A then B; not B; therefore not A), while as many as 63 percent endorsed the fallacy of affirming the consequent (if A then B; B; therefore A) and 55 percent endorsed the fallacy of denying the antecedent (if A then B; not A; therefore not B). When the antecedent is negative, affirming the consequent overtakes modus tollens in popularity.²¹ In some cases, when a further premise of the form “If C then B” is added to modus ponens only a minority endorses the inference (Byrne 1989).²² 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” (Manktelow and Over 1987, Wason and Shapiro 1971). In general, the real-life credibility or otherwise of premises and conclusion strongly influences judgments of validity and invalidity.

²¹ See Schroyens and Schaeken (2003); the percentages are as summarized by Oaksford (2005: 427).

²² Is it still modus ponens if there is an extra redundant premise? If not, then humans apply modus ponens only in the most artificial circumstances, since in practice we always have further information. Moreover, people without formal education tend to do *worst* in reasoning tasks with artificial premises from which all background information has been screened out (see Harris (2000: 94–117) for discussion). Such a restriction would make a disposition to assent to modus ponens a rather artificial test for understanding “if.”

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 among those with the intelligence of able undergraduates (Newstead *et al.* 2004; the samples in the experimental literature tend to consist of university students, since they are the most easily available subjects). 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 modular that contains formal rules to subserve conscious deduction, whether conceived as part of a language module or as part of a reasoning module.

Of course, there may be sub-personal processes whose inner workings can conveniently be represented as employing deductive rules, just as there may be sub-personal processes whose inner workings can conveniently be represented as employing differential equations, for example to process perceptual input, in even the most mathematically ignorant subjects. But that is not quite the issue. We are questioning the existence of a sub-personal basis for an unmanifested disposition to assent, that is, to perform an action at the personal level. The problem is that the data of normal performance tell against the hypothesis of a set of deductive rules (semantic or not) unconsciously employed as the primary route to conscious assent in the relevant normal cases.

A widespread, although not universal, view among 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.²³ System 1 lacks the formal rules that enable deductive rea-

²³ See Stanovich and West (2000: 659), where a list is also provided of earlier authors who have proposed similar views.

soning to succeed in the absence of helpful clues from the content of premises and conclusion. Although defeasible and only moderately reliable, 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 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 it is not sealed off in an unconscious module.

How does this picture apply to Peter and Stephen? With respect to System 1, they fall within the normal range of human variation. They are slightly unusual with respect to System 2, which 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 to understand simple sentences like (1). Any System 2 differences at issue between Peter or Stephen and average speakers of English are wholly 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 aversion to universal generalizations with empty subject terms or borderline cases seems to be within the normal range of System 1 reasoning among native speakers. On the two systems picture, there is no reason to assume that all linguistically competent speakers have an underlying disposition to assent to (1).

The two systems picture has not been conclusively established; it may turn out to need modification. Nevertheless, it throws into relief the empirical speculations on which the sub-personal understanding-disposition-to-assent links depend, 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 links are in trouble.

How can System 1 or any other system evaluate deductive arguments without using formal rules for reasoning with logical constants in natural language, even if their effect is almost swamped by associations, heuristics, and other pragmatic factors?²⁴ There are alternatives. For example, one of the main psychological theories of deductive

²⁴ For such an approach see Braine and O'Brien (1991), criticized by Evans and Over (2004: 56–9).

reasoning is currently 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 they have examined includes no counter-model, and they wrongly treat it as representative. They may erroneously classify a valid argument as invalid, because they leave the process of constructing a counter-model incomplete, under the misapprehension that there is no obstacle to completing it. Background beliefs about the specific subject matter of an argument influence its classification because they influence which mental models are constructed. 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.²⁵ But that process does not involve deductive reasoning in a natural language. Nor would natural deduction rules for the natural language connectives be very relevant; it is more like the construction of a truth-table. For example, in calculating the truth-value of a conditional in a model, one does not apply the rule of conditional proof to that very conditional if one already has the rules for constructing truth-tables.²⁶

²⁵ Mental models need not be visualized (Johnson-Laird and Byrne 1993: 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.

²⁶ Standard proofs of formalizations of (1) use conditional proof.

Evaluating a sentence in a model might involve something closer to an imaginative analogue of the processes that issue in complex perceptual judgments 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.

Therefore:

Everybody over there is wearing a hat.

But of course the final premise “Everybody over there is A, B, or C” is itself a universally quantified perceptual judgment. To suppose that it too was reached as the conclusion of a deductive argument is to start a futile regress.

Although the mental models theory does not apply to all human reasoning – for example, to the System 2 kind some humans learn to carry out in logic classes – it may apply to a high proportion of it. The theory is a salutary reminder that reasoning with logical constants need not be formal deductive reasoning, and that the empirical evidence suggests that in humans it usually is not.

One remaining concern is that logical skills must play some role in linguistic competence because logical features play a role in determining well-formedness. An example is the category of negative polarity items. Consider these sentences:

- (4) If she ate any of the cake, she was hungry.
 (5)* If she was hungry, she ate any of the cake.

“Any” is a negative polarity item. To a first approximation, the reason why “she ate any of the cake” is acceptable as the antecedent of the conditional but not as the consequent is that the antecedent is in a downward entailing (negative) context while the consequent is instead in an upward entailing (positive) context. A context *C* is upward entailing just in case whenever *A* entails *B*, *C*(*A*) entails *C*(*B*); *C* is downward entailing just in case whenever *A* entails *B*, *C*(*B*) entails *C*(*A*). Thus recognition of the logical features of contexts

seems to be needed in order to distinguish between well-formed and ill-formed sentences. But things are not so simple. Consider these sentences:

- (6) Exactly four people in the room were of any help.
- (7) Few people in the room were of any help.

Logically, “few” creates a downward entailing context; “exactly four” does not. However, (6) is acceptable provided that in the context it is taken to imply (7), but not generally otherwise. Thus the phenomenon involves a significant pragmatic element: which contexts are suitable for “any” cannot be determined on purely logico-linguistic grounds. If we disagree with the speaker of (6) about how many people were in the room or what proportion of them could have been expected to help, we may find her use of “any” inappropriate without regarding her as *linguistically* incompetent. Similarly, if a speaker has deviant views as to which contexts are downward entailing, but uses “any” in just those contexts that she treats as downward entailing, we might find her deviant use of “any” inappropriate without regarding her as linguistically incompetent, precisely because the deviation in use is explained by logical rather than linguistic unorthodoxy. Thus the role of logical knowledge in such cases does not make it part of purely linguistic competence. All our knowledge is potentially relevant to judging the appropriateness of a given use of “any.”²⁷

Suppose, nevertheless, that our classification of strings such as (4)–(7) as well- or ill-formed does depend on some prior classification of contexts as downward entailing or not. The question remains: is that classification available for unconscious reasoning that would issue in conscious assent to supposedly analytic sentences? To identify

²⁷ Ladusaw (1996: 325–37) surveys issues concerning negative polarity. Strictly speaking, the context of the antecedent of a counterfactual conditional is not downward entailing on standard logics of such conditionals, according to which strengthening of the antecedent fails; for example, although “It rained hard” entails “It rained,” “If it had rained, it would not have rained hard” does not entail “If it had rained hard, it would not have rained hard.” Nevertheless, negative polarity items are felicitous in the antecedent of counterfactual conditionals: “If you had taken any of that arsenic, you would have died” (see van Rooij (2006) for discussion).

a context as downward entailing involves a more sophisticated logical insight than identifying a particular argument as valid, since it requires the validation of an abstract pattern of argument. For example, identifying negation as a downward entailing context requires checking this schema, for arbitrary sentences “A” and “B”: If “A” entails “B” then “It is not the case that B” entails “It is not the case that A.” That is just the kind of abstract formal reasoning task on which humans perform worst. Contrast that with our high level of reliability in determining whether strings with negative polarity items are well-formed. Thus the evidence suggests that the unconscious logic in question is not at the service of the cognitive processes that normally produce conscious assent to sentences like (1). Such cases therefore fail to support a modification of the conclusions reached so far.

One special sort of case deserves separate discussion. Some metalinguistic sentences or thoughts look analytic for distinctive reasons. As observed in Chapter 2, even when a philosophical question is not itself metalinguistic, metalinguistic considerations can still help us to answer it.

Consider theoretical terms. We can understand the word “phlogiston” without believing phlogiston theory. Might we do so because we still believe that “phlogiston” is generally associated with that theory, just as one can understand a natural kind such as “gorilla” without believing the associated stereotype (“Gorillas are ferocious”) because one still believes that “gorilla” is generally associated with that stereotype (Putnam 1975)? However, such sociolinguistic beliefs are no more immune than logical beliefs from the challenge of theoretical unorthodoxy without change of meaning. If T is any version of phlogiston theory, someone can understand “phlogiston” and associate it with T without believing that it is generally associated with T, in the belief that “phlogiston” is and was generally associated not with T but with somewhat different versions of phlogiston theory. This is clear if T is a strong version of the theory. Even if T is a weak version, they may believe that the word is generally associated with a stronger version, and deny that it is *ipso facto* associated with T. On such grounds, they may even disbelieve that they themselves associate the word with T. Let such sociolinguistic beliefs be false; nevertheless, holding them is quite consistent with understanding “phlogiston.” It is futile to multiply disjuncts and restrictive clauses

in the hope of formulating a sociolinguistic claim so anodyne that anyone who understands “phlogiston” *must* accept it. The result will just be a complex theoretical claim that ordinary speakers can legitimately doubt, on the grounds that such matters are hard to determine.

A more minimalist line of argument for metalinguistic analyticities appeals to the connection between understanding and knowledge of reference. Suppose that someone understands this sentence:

(8) “Tree” applies to all and only trees.

Then they understand its constituent words, in particular “tree.” So they know what “tree” means. For common nouns, knowledge of meaning requires knowledge of application conditions. Consequently, they know that “tree” applies to all and only trees. Moreover, since knowledge entails belief, they also believe that “tree” applies to all and only trees. Thus, it seems, they should knowledgeably assent to (1). The argument generalizes to a large class of disquotational claims (the identity of the expression mentioned on the left-hand side with the one used on the right-hand side is crucial, since if they were distinct understanding of the latter would not entail knowledge about the former).

Nevertheless, those who understand (8) may refuse assent to it. Stephen is an example, since on his view a universally quantified biconditional with borderline cases for both sides is not definitely true. Indeed, some supervaluationists about vagueness even deny such disquotational principles for vague terms, such as “tree”. However erroneous such theories of vagueness, holding them is consistent with ordinary linguistic understanding of (8). If understanding really does involve tacit propositional knowledge of meaning, that knowledge may contradict conscious beliefs.

Let us grant for the sake of argument that understanding (8) entails knowing both that “tree” applies to all and only trees and that (8) means that “tree” applies to all and only trees. How then can one understand (8) without assenting to it? We lack direct conscious access to whatever tacit knowledge linguistic understanding is supposed to consist in, otherwise semantics as a branch of empirical linguistics would be much easier than it actually is. We consciously entertain the proposition that “tree” applies to all and only trees as

presented by sentence (8), or by the corresponding conscious thought “*tree*” applies to all and only trees. In tacitly knowing that “tree” applies to all and only trees (if we do), we may tacitly entertain that proposition under a quite different unconscious mode of presentation. Thus understanding-assent links fail for sentences of natural language and conscious thoughts:

- (UAI*) Necessarily, whoever understands the sentence “‘Tree’ applies to all and only trees” assents to it.
 (UAt*) Necessarily, whoever grasps the thought “*tree*” applies to all and only trees assents to it.

For if linguistic understanding involves tacit propositional knowledge of meaning, it presumably involves tacit assent to the relevant propositions under modes of presentation of some sort. Any tacit assent to the proposition that “tree” applies to all and only trees need not be to it under the modes of presentation that UAI* and UAt* require. The same difficulty arises even if we require only a disposition to assent, as in UDAI’ and UDAt’.²⁸

To determine in exactly what sense of “tacit knowledge,” if any, understanding does involve tacit propositional knowledge of meaning lies beyond the scope of this book. According to Gareth Evans (1985: 338–9):

Tacit knowledge of the syntactic and semantic rules of the language are [sic] not states of the same kinds as the states we identify in our ordinary use of the terms “belief” and “knowledge.” Possession of tacit knowledge is exclusively manifested in speaking and understanding a language; the information is not even potentially at the service of any other project of the agent, nor can it interact with any other beliefs of the agent (whether genuine beliefs or other tacit “beliefs”) to yield further beliefs. Such concepts as we use in specifying it are not concepts we need to suppose the subject to possess, for the state is inferentially insulated from the rest of the subject’s thoughts and beliefs.

Even if the contrast is less extreme than Evans argues, the lack of inferential integration is real, and crucial here. Of course, the

²⁸ See also Soames (1995) for relevant considerations.

ordinary notions of knowledge and belief may well provide appropriate templates for the construction of new notions of “tacit knowledge” and “tacit belief” of value to cognitive psychology. It can be theoretically rewarding to exploit the similarities between tacit knowledge and ordinary knowledge, but for present purposes it is the differences that matter.

Whatever the nature of tacit assent and dissent, no reflective intellectual discipline operates at the level of such assent and dissent, even if such a tacit level is necessary for its operation. Thus linguists’ tacit knowledge of their native language does not already satisfy the goal of linguistics. Similarly, philosophy as a discipline operates at the level of conscious reflection and public discussion, whatever their unconscious underpinnings. For present purposes, we may therefore restrict assent to conscious assent and maintain the generalization that there are no necessary links from understanding to assent, or even to dispositions to assent.

To summarize: The case for treating lack of a disposition to assent to (1) as lack of 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. Links from linguistic understanding to assent or to dispositions to assent fail.

5

The argument of the last two sections was at the level of language, not thought. It was directed primarily against UAI’ and UDAI’, not UAt’ and UDAt’. Could a theorist of thought maintain UAt’ or UDAt’ while acknowledging Peter and Stephen as counterexamples to UAI’ and UDAI’?

For the sake of argument, thoughts are being individuated by a cognitive criterion fine enough to suit an epistemological conception of analyticity, so we may assume that when a speaker understands a sentence, they associate it with a unique thought, in the intimate way

in which we associate the sentence “Grass is green” with the thought *grass is green*. In particular, the speaker assents to the sentence if and only if they assent to the thought. Consider Stephen (the argument is parallel for Peter). Since Stephen understands “Every vixen is a vixen,” he associates it with a unique thought *t*. Thus Stephen assents to “Every vixen is a vixen” if and only if he assents to *t*. But Stephen is an acknowledged counterexample to UAI’; he does not assent to the sentence “Every vixen is a vixen.” Therefore he does not assent to *t*. Consequently, if *t* is the thought *every vixen is a vixen*, Stephen does not assent to the thought *every vixen is a vixen*, in which case he is also a counterexample to UAt’. Thus if Stephen is not a counterexample to UAt’, the thought he associates with the sentence “Every vixen is a vixen” is not the thought *every vixen is a vixen*.

There is a parallel argument for dispositions. Stephen is an acknowledged counterexample to UDAl’; he understands “Every vixen is a vixen” while having no disposition to assent to it. We may therefore assume that he is relevantly stable; thus in all relevant situations *t* is the unique thought he associates with the sentence. Thus Stephen has a disposition to assent to “Every vixen is a vixen” if and only if he has a disposition to assent to *t*. Therefore he has no disposition to assent to *t*. Consequently, if *t* is the thought *every vixen is a vixen*, Stephen has no disposition to assent to the thought *every vixen is a vixen*, in which case he is also a counterexample to UDAt’. Thus if Stephen is not a counterexample to UDAt’, the thought he associates with the sentence “Every vixen is a vixen” is not the thought *every vixen is a vixen*.

The upshot is that theorists of thought can maintain links from understanding to assent or dispositions to assent at the level of thought while abandoning them at the level of language only if they deny that the thought Peter or Stephen associates with the sentence “Every vixen is a vixen” is the thought *every vixen is a vixen*. They may either deny that Peter and Stephen grasp the thought *every vixen is a vixen* at all or assert that they grasp the thought by some means other than that sentence and assent to it, or at least have a disposition to assent.

The thought *every vixen is a vixen* is the thought *we* associate with (1). Thus the envisaged theorist of thought is claiming that the thought we associate with (1) differs from the thoughts Peter and Stephen associate with it, even though all of us understand (1) with

its usual meaning in English.²⁹ This need not imply that (1) is indexical, expressing different propositions in the contexts of different speakers, for thoughts are not being identified with propositions. You might use the sentence “He is hungry” (pointing at me), which you associate with a demonstrative thought *he is hungry* to express the very proposition I express using the sentence “I am hungry,” which I associate with the distinct thought *I am hungry*; you associate the sentence “I am hungry” with the same thought but use it to express a different proposition, that you are hungry. For all that has been said, Peter and Stephen use (1) to express the same proposition as we do. But on what basis are the thoughts Peter and Stephen associate with (1) being distinguished from the thought we associate with (1)?

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 with (1). But what is 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 over it. To insist that the thought we associate with (1) nevertheless differs from the thoughts Peter and Stephen associate with (1) is to undermine Frege’s requirement of the publicity of senses, and in particular thoughts.

If Peter and Stephen associate (1) with different thoughts from ours, should we not understand them better by translating their idioms non-homophonically into ours? Presumably we should seek sentences other than (1) that we associate with the very thoughts they associate with (1), or at least sentences we associate with thoughts

²⁹ Neo-Fregeans such as Evans (1982: 40) sometimes claim that different speakers can achieve linguistic competence with the same proper name by associating it with different concepts (modes of presentation) of the same object. On the view envisaged in the text, phrases such as “the thought *every vixen is a vixen*” or “the concept *every*” presumably are indexical, since they refer to the thought or concept that the speaker associates with the italicized expression. Discussions of concept possession tend to use such phrases freely, without attention to such indexicality. On the envisaged view, they may require consequent revision.

more similar to the thoughts they associate with (1) than is the thought we associate with (1), and translate the 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 that 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 the thought we associate with (1), and that we should apply no 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 that 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 do not store exactly what 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 logic skills is quite consistent with reading their utterances homophonically.

Peter and Stephen are counterexamples to UAt' and UDAt'. The links from understanding to assent, or even to dispositions to assent, fail for thought as they do for language.

6

How do the considerations of preceding sections apply to traditional paradigms of analyticity? Consider:

(9) Every vixen is a female fox.

Given that “vixen” is synonymous with “female fox,” (9) results from substituting synonyms for synonyms in the logical truth (1). Hence (9) is synonymous with (1): it is Frege-analytic but not itself a logical truth. We can expect the arguments of previous sections against links from understanding to assent or dispositions to assent for examples like (1) to work at least as strongly for examples like (9). Let us check this.

We may try to reduce discussion of (9) to discussion of (1), on the grounds that the concept *vixen* just is the concept *female fox*. Thus the thought *every vixen is a female fox* just is the thought *every vixen is a vixen* (since thoughts are composed of concepts). To grasp, assent to or know a thought is just to have a relation to that thought. Consequently, to grasp, assent, or know *every vixen is a female fox* just is to grasp, assent, or know *every vixen is a vixen*. At the level of thought, the previous discussion carries over automatically. For example, in being counterexamples to the understanding-assent link for the thought *every vixen is a vixen*, Peter and Stephen are *ipso facto* counterexamples to the understanding link for the thought *every vixen is a female fox*.

At the level of language, the reduction is slightly more complicated: “vixen” and “female fox” are distinct expressions even if they are associated with the same concept. Someone can understand “female fox” without understanding “vixen.” Conversely, someone can understand “vixen” without understanding “female fox”: for instance, a native speaker of another language who is learning English understands “vixen,” because she was taught it as a synonym for a word in her native language, but has not yet encountered “female” and

“fox.” If she has mastered the construction “Every . . . is a –,” she can understand (1) without being in a position to understand (9). Someone who understands neither (1) nor (9) can assent to one of them without assenting to the other, on the testimony of someone else who tells him that the former is true without telling him that the latter is true. Nevertheless, we might try arguing that whoever understands (9) will take just the same attitudes to it as to (1).

The argument is this. Suppose that someone understands (9) (as always, with its normal English meaning). Thus she associates it with the thought *every vixen is a female fox*. Consequently, she takes an attitude A1 (such as assent or knowledge) to (9) if and only if she takes the corresponding attitude A2 to the thought *every vixen is a female fox* at the level of thought (in preceding sections, A1 and A2 were equated). Our speaker also understands (1), because it is composed entirely out of words (“vixen”) and modes of construction (“every . . . is a –”) which she understands in understanding (9). Thus she associates (1) with the thought *every vixen is a vixen*. Consequently, she takes A1 to (9) if and only if she takes A2 to the thought *every vixen is a vixen*. For the reason already given, the thought *every vixen is a vixen* is the thought *every vixen is a female fox*. Therefore she takes A2 to the thought *every vixen is a vixen* if and only if she takes A2 to the thought *every vixen is a female fox*. It follows that she takes A1 to (9) if and only if she takes A1 to (1). Thus, with respect to speakers who understand (9), discussion of (9) reduces to discussion of (1).

Whether or not the concept *vixen* is the concept *female fox*, the reduction succeeds for Peter and Stephen, since they use the concepts interchangeably and do understand (9). They are counterexamples to epistemological analyticity for (9) just as much as they are for (1), at the levels of both thought and language.

The assumption that the concept *vixen* is the concept *female fox* is controversial. 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 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 metalinguistic. Does this speaker assent to the thought *every vixen is a vixen* without assenting to the thought *every vixen is a female fox*? If so, the thoughts are distinct (which is

compatible with the identity of the proposition that every vixen is a vixen with the proposition that every vixen is a female fox), and the concept *vixen* is not the concept *female fox*.

To make a case more like those of Peter and Stephen, we can imagine that our speaker is quite familiar with the dictionary definition of “vixen” as “female fox.” He also knows that dictionaries give a second definition of “vixen” as “quarrelsome woman.” However, unlike most of us, he does not believe that these are two senses of “vixen.” Rather, he thinks that “vixen” in its primary sense applies to both female foxes and quarrelsome women. He may defend his view with sophisticated arguments from the philosophy of language, although this is not essential. He denies (9), intending “vixen” in the public sense in which it applies at least to female foxes.

Our imaginary speaker is not so different from actual native speakers of English who deny that a man who has lived with a partner for several years without getting married is a bachelor, or assert that someone who underwent a sex-change operation after giving birth is a mother without being a female parent.³⁰ Suppose that they are in fact mistaken; “bachelor” has the same intension as “unmarried man” and “mother” has the same intension as “female parent.” Thus they are mistaken about the meaning of the English words “bachelor” and “unmarried.” Nevertheless, they fall well within the range of permissible variation for linguistically competent speakers. They are only giving more weight than others to an inclination that most speakers feel in some degree to classify the cases that way. Without regarding them as having spoken parrot-fashion, we report their beliefs using the words “bachelor” and “unmarried.” We classify them as believing that some unmarried men are not bachelors and that some mothers are not female parents because we interpret them as having used the words with their normal English meanings, despite their errors. That is how they intend to be interpreted, not as using the words with idiosyncratic senses.³¹ If we believe that all unmarried

³⁰ Compare Harman (1999: 151) on problems in analyzing “bachelor” as “unmarried adult male” and Nozick (2001: 135–6) on the non-synonymy of “mother” and “female parent.”

³¹ One problem with interpreting speakers as all speaking their own idiolects is that it tends to undermine testimonial knowledge: if Y gets some knowledge from X and passes it on to Z in the same words, they do not mean in Y’s mouth what they meant in X’s.

men are bachelors and all mothers are female parents, we therefore classify their beliefs in question as untrue, for the belief that some unmarried men are not bachelors is true if and only if some unmarried men are not bachelors, and the belief that some mothers are not female parents is true if and only if some mothers are not female parents. Given that we correctly interpret them as using the words with their normal English meanings, they understand the words in the relevant sense of “understand.” Although they are ignorant of some facts about the normal English meanings of the words, such ignorance is quite compatible with linguistic competence (which is why native speakers of English take university courses in the semantics of English). Arguably, their error is not primarily semantic: they have the semantic belief that the word “bachelor” does not apply to all unmarried men because they have the non-semantic belief that some unmarried men are not bachelors and the semantic knowledge that “bachelor” applies only to bachelors; they have the semantic belief that the word “mother” does not apply only to female parents because they have the non-semantic belief that some mothers are not female parents and the semantic knowledge that the word “mother” applies to all mothers.

Such cases also help answer the objection to examples such as those in this chapter that the awkward subject who consciously denies that P also has unconscious, semantically derived knowledge (or belief) that P. When a competent native speaker denies that every unmarried man is a bachelor, the postulation of unconscious knowledge (or belief) that every unmarried man is a bachelor serves no good explanatory purpose. The speaker tends to apply “bachelor” to something once they have applied “unmarried” and “man” to it, but the tendency is defeasible. Such defeasible connections can be explained without postulation of unconscious belief in a universal generalization. In such cases, there need be no hint of the cognitive dissonance or tension that one might expect from a direct contradiction between conscious and unconscious beliefs. Given that there is no contradicted unconscious knowledge in these simple cases, it is not clear what better reason there is supposed to be in postulating it for more complex cases either.

Suppose, given the considerations above, that the concept *vixen* is not the concept *female fox*. Then the claim of epistemological analyticity is even worse off for (9) than it is for (1), at the levels of both

thought and language. Logically orthodox subjects can understand (9) and grasp the thought *every vixen is a female fox* while refusing to assent. In that case, they will also reject the corresponding inference rule with instances of the form “*a* is a vixen; therefore *a* is a female fox” (and conversely); likewise at the level of thought.³²

The underlying style of argument against links from understanding to assent or dispositions to assent is quite general. For each candidate one must still find appropriate counterexamples: since they are most convincing when unorthodoxy on the point at issue is amply compensated by orthodoxy on related points, no one counterexample will suit all cases. Nevertheless, with a little ingenuity one always succeeds.³³

³² Peter and Stephen assent to the conclusion of this inference rule whenever they assent to its premise. For some subtler problems it raises for them see Williamson (2006b: 33–4).

³³ Another application of the present style of reasoning is to claims that sorites paradoxes reveal incoherence in vague concepts. Thus Dummett (1975a) argues that observational predicates in natural language are governed by rules that infect the language with inconsistency: for example, to understand “looks red” one must be willing to apply a tolerance principle by which one can infer from “*x* is visually indiscriminable from *y*” and “*x* looks red” to “*y* looks red,” which generates sorites paradoxes because visual indiscriminability is non-transitive. More recently, Roy Sorensen (2001) has argued that linguistic competence with vague terms involves willingness to make inferences such as that from “*n* seconds after noon is noonish” to “*n* + 1 seconds after noon is noonish,” which commits us to inconsistent conclusions by sorites reasoning (given our other commitments, such as “Noon is noonish” and “Midnight is not noonish”). Matti Eklund (2002) defends a similar account of both sorites and semantic paradoxes. There are no such requirements on linguistic competence and concept possession. An ordinary speaker of English who understands “looks red” and “noonish” and has the concepts *looks red* and *noonish* in the normal way but then rejects the relevant tolerance principles in the light of the sorites paradoxes does not thereby cease to understand those expressions or to have those concepts. She might treat the premises of the tolerance principles as providing good defeasible evidence for their conclusions, without even being *disposed* to expect long chains of such reasoning to preserve truth; this attitude seems to be less than Dummett, Sorensen, and Eklund require for competence, since it is insufficient to render sorites paradoxes puzzling. In any case, even if a whole community of speakers is disposed to treat tolerance principles as obviously fallacious, it can still have terms like “looks red” and “noonish” that are just as vague as ours; speakers’ acceptance of tolerance principles is quite inessential to vagueness.

In principle, we could also explore putative links from understanding of one sentence to (dispositions to) assent to another sentence or a thought, or from grasp of one thought to (dispositions to) assent to another thought. In practice, such candidates fall to objections very similar to those already raised. Details are therefore omitted.

7

Old theories tend to survive refutation in the absence of new theories to take their place. Despite all the evidence against the existence of links from understanding to assent or dispositions to assent, 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 (1998: 132) 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.³⁴

³⁴ Jackson's application of the Ramsey-Carnap-Lewis method for defining theoretical terms to moral vocabulary (and more generally in his program 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 also 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 application of the Ramsey-Carnap-Lewis method in Boghossian (2003) see Williamson (2003a). In general, if the platitudes are weak, as we have every reason to expect, many different candidates will satisfy them. Call these the *admissible candidates*. For simplicity, think of them as properties (more accurately, they are *n*-place sequences of properties and relations, where *n* is the number

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

The notion of a shared language is vague (Jackson does not suggest otherwise). There can be sorites series of speakers in which each seems to be speaking the same language as the next but the first is clearly not speaking the same language as the last.³⁵ One reaction is that there is no such thing as a shared language, a conclusion endorsed in some form by both Noam Chomsky and Donald Davidson. Similarly, Margaret Thatcher once claimed "There is no such thing as society," and one can certainly construct sorites series in her support. But almost everything looks vulnerable to sorites series; they are a poor way to establish non-existence. Whatever exactly shared languages are, they are no mere illusion. We can follow Jackson in asking how they are possible. But there is an alternative to his answer.

of primitive predicates to which the method is being applied). The conjunction or disjunction of these admissible candidates will often not itself be an admissible candidate. Schematic example: let the platitudes be "All Fs are electrons," "Some electrons are Fs" and "Some electrons are not Fs," where the method is being applied to "F"; the conjunction of the admissible candidates is the empty property, which does not satisfy the second platitude and so is inadmissible; their disjunction is the property of being an electron, which does not satisfy the third platitude and so is inadmissible. The non-uniqueness problem for the Ramsey-Lewis-Carnap *definiens*, in effect "the property that satisfies the platitudes," is *not* that it is vague which property it denotes but that it definitely fails to denote any property at all, since many properties definitely satisfy the platitudes; neither supervaluationism nor any other theory of vagueness rescues the definition. A modified description such as "the most natural property that satisfies the platitudes" may still not solve the problem – perhaps several admissible candidates are equally natural and more natural than any others, or for every admissible candidate there is a more natural one – and in any case raises the question why the Ramsey-Carnap-Lewis method is being applied to some terms but not to the highly theoretical term "natural" itself (otherwise the problem simply recurs for "natural"). It is a mistake to assume that such problems are really problems for the linguistic practice itself rather than for the appeal to platitudes, for that is to assume that the platitudes exhaust what the practice does to secure reference for the predicate. Uses of the predicate to make controversial claims may also play a role in determining its reference, although not a naïvely descriptivist role (the account in Chapter 8 will permit this). The method of platitudes rashly throws such information away.

³⁵ Williamson (1990: 137–41) discusses sorites series for languages.

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 rough 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 idea 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.³⁶

As Kripke and Putnam argued, 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. The paradigm is their description of the role of scientific experts in fixing the reference of natural kind terms. Even if they oversimplified the relation between natural kind terms in natural language and scientific theory, a more refined account will still respect the division of linguistic labor, for distinctions between levels of expertise are observable even within the pre-scientific use of natural kind terms. Contrary to some of Putnam's less careful formulations, no canonical list of "criteria"

³⁶ A similar point is made in Schroeter and Schroeter (2006). More generally, the research program that these authors are pursuing has points in contact with the ideas of the present chapter.

for the application of the term need be available even to the most expert members of the community. Speakers may simply differ from each other in various ways in their ability to distinguish between members and non-members of the relevant kind.

The underlying insight is relevant far beyond the class of natural kind terms, as Burge observed. 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.³⁷ No asymmetries in sociolinguistic status between individual speakers are required. For instance, if I classify a shade close to orange as “red” but subsequently discover that it is classified as “not red” by most native speakers of English whose eyesight is as good as mine, I may rationally admit that I was wrong without conceding that either I misunderstood the word “red” or my visual system was abnormal or malfunctioning. One can know that “red” means *red* without being infallible as to exactly which shades count as shades of red. Even if I obstinately insist that I am right and the rest are wrong in this particular case, my assumption that “red” in my mouth is inconsistent with “not red” in theirs shows that I intend my use of “red” to be treated as the use of a word of a public language. That its reference is fixed by the pattern of use over the whole community does not entail that the majority must be right in any given case: reference can supervene on underlying facts in ways far from transparent to native speakers.

The unity of a linguistic practice, like the unity of other complex objects, has both synchronic and diachronic aspects. As usual, causal continuity is necessary but not sufficient for diachronic unity. Anaphoric pronouns constitute one paradigm of such unity: the reference of later tokens is parasitic on the reference of earlier tokens; the identity of reference results from collusion, not coincidence. Over a longer timescale, the historical chains that preserve the reference of names represent a similar form of diachronic unity. Written testi-

³⁷ 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 (1999b: 512–14).

mony and verbal testimony preserved in memory depend on such reference-preserving links. As usual, the intention to preserve reference is not guaranteed to succeed, but success is the default (Kripke 1980).

Such diachronic links can hold non-trivially even for the linguistic or conceptual practice of an isolated individual. Contrary to some readings of Wittgenstein's private language argument, what seems right to the isolated individual need not be right, given their overall use dispositions: even at the individual level, reference can supervene on underlying facts in ways far from transparent to the subject. The point of the social determination of meaning is not that meaning can never be determined individually, but that, when an individual does use a shared language as such, individual meaning is parasitic on social meaning.

A complex web of interactions and dependences can hold a linguistic or conceptual practice together even in the absence of a common creed that all participants at all times are required to endorse. This more tolerant form of unity arguably serves our purposes better than would the use of platitudes as entrance examinations for linguistic practices.

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 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.³⁸

³⁸ Davidson famously endorses a holistic principle of charity while rejecting the analytic-synthetic distinction (2001: 144–9). See Chapter 8 for more discussion of charity. Of course, he takes the notion of a shared language less seriously than here (Davidson 1986).

It is far easier and more rewarding to discuss the existence of true contradictions with a dialetheist such as Graham Priest than creationism with a Christian fundamentalist or Holocaust denial with a neo-Nazi.³⁹ The difficulty of engaging in fruitful debate with fundamentalists or neo-Nazis is not plausibly attributed to some failure of linguistic understanding on their part (or ours); it arises from their willful 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 that 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.⁴⁰ Similarly, conceptual practices do better not to restrict in advance their capacity for innovation.

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.⁴¹

³⁹ For examples of rational debate for and against a law of non-contradiction see Priest, Beall, and Armour-Garb (2004).

⁴⁰ 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).

⁴¹ Someone who understands a word without being disposed to utter it (perhaps because they find it obscene or unpronounceable) can still count as sufficiently

If we picture speaking the same language in this way, how should we picture meaning the same thing? There is no quick generalization from the former to the latter. Different uses of the same word must be causally related, at least indirectly.⁴² Creatures who are causally unrelated to us cannot use our word “not”; at best they can use a word exactly like our word in its general syntactic, semantic, and phonetic properties. But, on the usual view, their word can in principle be synonymous with ours. Synonymy does not entail causal relatedness.

Expressions are synonymous when they have exactly the same semantic properties. Fortunately, the tradition of truth-conditional semantics provides us with a rich store of such properties, if we take it seriously as a branch of linguistics and put aside Quinean reservations.

Two paradigms of a semantic property are the extension of a predicate, the set of things to which it applies, and its intension, the function that takes each circumstance of evaluation (say, an ordered pair of a world and a time) to the extension of the predicate with respect to that circumstance. For the purposes of compositional semantics, this approach can be generalized to expressions of other grammatical categories, so that they have intensions too. Thus synonymy entails at least sameness of intension. That is still a rather coarse-grained criterion, since it does not reflect internal compositional structure: “5 + 7” and “9 + 3” have the same intension. We can go more fine-grained by associating expressions with trees whose nodes correspond to their semantically significant constituents, each node being decorated with the content of the corresponding constituent; the branching structure of the tree encodes the constituency structure of the expression. Thus synonymy entails at least sameness of associated tree. This criterion is similar to Carnap’s notion of intensional isomorphism (1947: 56). In this sense not even “vixen” and “female fox” are synonymous, since they differ in semantically significant structure, unless the account can be applied at a level of deep logical form at which they turn out to have the same constituents. Something like intensional isomorphism can serve as a criterion for sameness of content expressed in a given context of utterance.

engaged in the practice of using it. The account should also be read so as to allow for understanding of dead languages.

⁴² On the metaphysics of words see Kaplan (1990).

An expression brings its linguistic meaning to a context rather than having that meaning made up in the context. Thus “I” as used by TW does not have the same linguistic meaning as “TW,” even though they have the same content (since they are unstructured rigid designators of the same object). Rather, “I” as used by TW is identical in linguistic meaning with “I” as used by any other competent speaker of English. Thus a better approximation to the linguistic meaning of an expression is its character in the sense of Kaplan (1989), the function taking each context of utterance to the content of the expression in that context.

We might go still further. For instance, so far “and” and “but” come out synonymous, since they are simple expressions that make the same contribution to truth-conditions. We might distinguish their meanings by adding as further semantic properties conventional implicatures, themselves individuated like characters.

Even without conventional implicatures, once content is individuated by intensional isomorphism, the conception of linguistic meaning as character is already exquisitely fine-grained. Nevertheless, if semantic theory discovers a need to attribute still more semantic properties, or to revise the framework already sketched, sameness with respect to the newly identified semantic properties will be required for synonymy. In any case, we need not try to circumscribe in advance exactly what properties semantic theory will need to recognize.

The point is methodological. Whether an expression in one language is synonymous with an expression in another language is not a matter of whether the two speech communities associate similar beliefs with the expressions. Rather, the practices of each community (including their beliefs) determine the semantic properties of its expressions. Synonymy is the identity of the properties so determined, irrespective of similarities in belief. It is consistent with large differences in belief (just as very different distributions can have the same mean), and non-synonymy is consistent with much smaller differences in belief (just as very similar distributions can have different means). In particular, synonymy is consistent with the total absence of shared platitudes.

The synonymy of two expressions does not entail that competent speakers treat them interchangeably, as noted in chapter 3. Someone can understand “furze” and “gorse” by learning them from ostension of different samples without appreciating their synonymy. In some

cases, even competent speakers who know two expressions to be synonymous will not treat them interchangeably. For example, the slang word “gob” means the same as “mouth,” but competent speakers are normally sensitive to whether the social context makes “gob” (but not “mouth”) inappropriate. Such differences in register are linguistic but not semantic. Consequently, knowing the meaning of an expression does not automatically qualify one for full participation in the practice of using it. Someone who acquires the word “gob” just by being reliably told that it is synonymous with “mouth” knows what “gob” means without being fully competent to use it. One does not achieve full competence with a sentence of a foreign language by learning its meaning from a phrasebook without knowing which constituent contributes what to that meaning. For a less obvious case, consider empty terms. Arguably, “phlogiston” fails to refer with respect to any circumstance of evaluation (since it designates rigidly, if at all) and any context of utterance (since it is non-indexical); it is semantically atomic and has no conventional implicatures. Those facts may completely determine its semantics, strictly speaking. Nevertheless, knowing them alone does not qualify one to participate in the linguistic practice of using “phlogiston,” since they do not distinguish it from empty terms associated with other failed theories. Although no particular piece of knowledge is necessary for participation, such abject ignorance is not sufficient. We should resist the temptation to build all qualifications for participation in the practice of using a term into its meaning, on pain of turning semantic theory into a ragbag of miscellaneous considerations (even the inclusion of conventional implicature is marginal).

What of concepts? Presumably, thinkers causally unrelated to us could have the concept *not*. Hence sameness of concept does not entail causal relatedness; it is closer to sameness of meaning than it is to sameness of word. If so, the concept *furze* may well just be the concept *gorse*. If thoughts are composed of concepts in the obvious way, then the thought *all furze is gorse* just is the thought *all furze is furze*, and whoever assents to the latter *ipso facto* assents to the former. We may sometimes be unable to determine whether we are employing two concepts or one. That makes the individuation of thoughts and concepts less accessible to the thinker than many theorists of thought have wished. For the sake of greater (but still imperfect) accessibility, they might therefore switch to individuating

concepts more like words than like meanings. In any case, the argument against epistemological analyticity at the level of thought has already been explained, in Section 5.

8

At this point, a friend of epistemological analyticity may suspect that the mistake was to go for the idea that understanding is somehow *psychologically* sufficient for assent. Instead, the suggestion is, we should go for the idea that understanding is somehow *epistemologically* sufficient for assent.⁴³ Externally, Peter and Stephen are in a position to know (or to assent with justification). They seem to be willfully and perversely turning their backs on knowledge that is available to them. It is there for the taking, but they are psychologically blocked from taking it.

We must be careful about the source of the blockage. Suppose that it is lack of logical insight. Although Peter and Stephen grasp the thought *every vixen is a vixen*, they lack the logical insight to know *every vixen is a vixen*. Other people just like Peter and Stephen except for having more logical insight do know *every vixen is a vixen*. Anyone who grasps the thought *every vixen is a vixen* and has a modicum of logical insight can know *every vixen is a vixen*. That story assigns no special role to grasp of concepts, beyond the usual role that grasping any thought plays as a precondition for knowing it: the decisive role is assigned to logical competence, not conceptual competence. For conceptual competence to play the decisive role, something like this is needed:

(KU t') Whoever knows *every vixen is a vixen* in the normal way does so simply on the basis of their grasp of the thought.

(Understand “on the basis of” more like “by an exercise of” than like “by inference from.”) Similarly, for semantic competence to play the decisive role, something like this is needed:

⁴³ Some rationalist defenders of intuition seem to have something like this in mind.

(KUI') Whoever knows "Every vixen is a vixen" in the normal way does so simply on the basis of their understanding of the sentence.

KUt' and KUI' may be plausible at first sight. They do not imply that whoever understands the sentence or grasps the thought has a disposition to assent to it, let alone to know it.

What do the definite descriptions "their grasp of the thought" in KUt' and "their understanding of the sentence" in KUI' denote? There are thick and thin candidates. The thin candidates are the mere fact that they grasp the thought and the mere fact that they understand the sentence respectively. The thick candidates are the underlying facts that constitute the respective thin candidates, the facts that realize this particular subject's understanding at this particular time. The thin candidates are exactly similar for any two people who grasp the thought or understand the sentence, since they have the same property of grasping the thought or understanding the sentence. The thick candidates may differ between any two people who grasp the thought or understand the sentence, since different underlying facts can constitute their doing so. These characterizations are schematic, but will do for present purposes.

Suppose that the definite descriptions in KUt' and KUI' denote the thick candidates. KUt' and KUI' remain somewhat plausible on this reading. Then, given the holistic picture of concept possession and linguistic understanding in previous sections, KUt' and KUI' have much less epistemological significance than might have been hoped. The facts that constitute your understanding of a given sentence include various cognitive capacities that are not in general necessary for understanding that sentence, but help to make up your particular competence with it. For example, the facts that constitute Peter's understanding of (1) include his logical capacities; the facts that constitute Stephen's understanding of (1) include his rather different logical capacities. The bases cited in KUI' and KUt' include cognitive capacities that are not in general necessary for understanding the sentence or grasping the thought. Thus the thick candidates are too thick to yield bases for analyticity; they involve cognitive capacities that are not semantic or conceptual in any relevant sense.

Suppose instead that the definite descriptions in KUt' and KUI' denote the thin candidates. But they are not the bases in any useful

sense for knowing *every vixen is a vixen* or “Every vixen is a vixen” in the normal way, although confusion with the thick candidates may suggest otherwise. The thin candidates imply no specific logical capacity at all, as Peter, Stephen and others show. It is not as though in such cases the subject’s understanding quietly tells them to assent but they override the advice; it is providing no such advice to be overridden. For the imagined overridden advice is a metaphor for the hypothesis of overridden dispositions to assent, dispositions necessary for understanding; that hypothesis was rejected in Section 4. By itself, thin understanding cannot guide our assent. Consequently, understanding in the thin sense provides no basis for assent to anyone. Of course, understanding is a precondition for knowing, and in that sense may be *part* of the basis for knowing, but that point is quite general; it is neutral between the analytic and the synthetic. Although the combination of understanding in the thin sense with the right bit of elementary but not universal logical competence is a basis for knowing (1), that point neither explains why logical knowledge is available in the armchair nor makes it distinctively conceptual or semantic. By themselves, the thin candidates are too thin to be bases for knowledge.

We could try eliminating the talk of bases, for instance in formulations like these:

- (AJt’) Whoever grasps the thought *every vixen is a vixen* and assents to it does so with justification.
 (AJl’) Whoever understands the sentence “Every vixen is a vixen” and assents to it does so with justification.⁴⁴

But such principles are false, since someone who assents because his father told him not to does so without even defeasible justification. The obvious way to avoid such counterexamples and make the connection with conceptual or semantic competence is to qualify “assents to it” by “on the basis of that grasp [understanding].” But that returns us to the difficulties of KUt’ and KUl’.

⁴⁴ The intended differences between assenting with justification in AJt’ and AJl’ and being justified in assenting in UJt and UJl are that (i) the former but not the latter entails assent and (ii) the assent in the former must be appropriately sensitive to the justification.

The problem is general. The idea that, in the cases at issue, understanding is epistemologically sufficient for assent is the idea that assent on the basis of understanding has the desired positive epistemic status. But once we disambiguate “understanding” between thick and thin candidates, we can see that the thin candidates are too thin to be bases for assent while the thick candidates are not purely semantic or conceptual. The attempt to base the epistemology of obvious truths such as (1) and (9) on preconditions for understanding them rests on a false conception of understanding.

Linguistic competence plays the same role when we know “Vixens are female foxes” as when we know “There is a vixen in the garden.” It does not gain a role just because perception loses one. The contribution of linguistic competence amounts to this: you won’t get very far if you conduct your inquiry in a language you don’t understand. Of course, that goes for *any* inquiry.

The following chapters develop a quite different account of the nature of at least some philosophical knowledge, on which linguistic and conceptual competence play only this background role, and philosophical beliefs are much less distinctive in nature than many philosophers like to think. We start with knowledge of metaphysical possibility and necessity.