Published in *Philosophical Topics* 28 (2000): pp. 211-244

Truth, Falsity and Borderline Cases*

Miroslava Andjelković

University of Belgrade

M.Andjelkovic@f.bg.ac.yu

Timothy Williamson

University of Edinburgh

Timothy.Williamson@ed.ac.uk

Introduction. According to the principle of bivalence, truth and falsity are jointly exhaustive and mutually exclusive options for a statement. It is either true or false, and not both, even in a borderline case. That highly controversial claim is central to the epistemic theory of vagueness, which holds that borderline cases are distinguished by a special kind of obstacle to knowing the truth-value of the statement. But this paper is not a defence of the epistemic theory. If bivalence holds, it presumably does so as a consequence of what truth and falsity separately are. One may therefore expect bivalence to be derivable from a combination of some principles characterizing truth and other principles characterizing falsity. Indeed, such derivations are easily found. Their form will of course depend on the initial characterizations of truth and falsity, and not all such characterizations will permit bivalence to be derived. In this paper we focus on the relation between its derivability and some principles about truth and falsity. We will use borderline cases for vague expressions as primary examples of an urgent

challenge to bivalence.

A key variable in the relation is obviously the choice of a logic. For most of the paper our background logic is classical. That choice does not automatically prejudge the issue in favour of bivalence, for the latter can fail even in a classical context on some non-standard accounts of truth and falsity; traditional supervaluationist semantics provides the best-known instance.

Our strategy is to start not with semantic theories designed to have some particular result for vague languages but with principles which seem natural from the standpoint of the theory of truth and falsity. We then explore their implications for bivalence. More specifically, simple considerations about the semantics of indexicals will lead us to replace the usual disquotational characterization of truth by one which makes explicit allowance for contextual variation in what is said by a sentence. The apparatus we use for that purpose *appears* to make room for failures of bivalence in borderline cases. For the argument for bivalence requires a principle of uniformity in what a given sentence says in a given context, and borderline cases may appear to motivate a denial of such uniformity. We will argue that the appearance is illusory.

In the final section we investigate how far our results remain available within a many-valued framework. We will show that those who reject bivalent semantics are not thereby precluded from accepting the primary principles about truth and falsity, since they might adopt a many-valued logic in which bivalence is not deducible from those principles. The philosophical adequacy or otherwise of such a logic is not our current concern.[1] Rather, we hope to separate the contribution of the primary principles about truth and falsity from the contribution of logic.

One theme of this paper is the significance of falsity. Recent philosophy has produced a vast body of research and debate on `the problem of truth' - which is of course a tissue of connected problems. There is no corresponding body of research and debate on `the problem of falsity'. By contrast, for ancient philosophers - such as Plato in the *Sophist* - falsity rather than truth was the really problematic concept. The question of bivalence concerns falsity just as much as it concerns truth. The current disparity in treatment is to some extent a matter of presentation, for most of the relevant issues about

falsity are parallel to issues about truth, and so do not require separate extended discussion. Nevertheless, it cannot be taken for granted that there are no failures of parallelism, and we will consider some candidates. In any case, we shall often need to make principles about truth interact with principles about falsity to derive our results.

A Tarskian argument. One can observe interaction between the concepts of truth and falsity in Tarski's argument for his famous biconditional about truth in his paper `The semantic conception of truth and the foundations of semantics'. We will broach some of the ideas discussed below by analysing his argument. Tarski writes (1944: 343):

> Let us start with a concrete example. Consider the sentence "snow is white". We ask the question under what conditions this sentence is true or false. It seems clear that if we base ourselves on the classical conception of truth, we shall say that the sentence is true if snow is white, and that it is false if snow is not white. Thus, if the definition of truth is to conform to our conception, it must imply the following equivalence:
>
> The sentence "snow is white" is true if, and only if, snow is white.

Appealing to a certain conception of truth, in this passage Tarski claims that a *sufficient* condition for the sentence to be true is that snow is white, and a *sufficient* condition for it to be false is that snow is not white. He then *derives* the claim that the former condition is also necessary for the sentence to be true, and requires a definition of truth to imply both its sufficiency and its necessity.[2] How does the derivation go? In obvious notation, we may formulate the premises and the conclusion of Tarski's inference thus:

(P1) $\mathbf{P} \supset \mathbf{T}(\ulcorner \mathbf{P} \urcorner)$

(P2) $\mathbf{\sim P \supset F(\ulcorner P \urcorner)}$

(C1) $\mathbf{T(\ulcorner P \urcorner) \equiv P}$

A natural reconstruction of the reasoning from (P1) and (P2) to (C1), and in particular to the left-to-right direction of (C1), is that Tarski is assuming the contrariety of truth and falsity:

(P3) $\mathbf{F(\ulcorner P \urcorner) \supset \sim T(\ulcorner P \urcorner)}$

For by transitivity we may infer $\mathbf{\sim P \supset \sim T(\ulcorner P \urcorner)}$ from (P2) and (P3), from which in classical logic contraposition, double negation introduction and elimination, conditional proof and transitivity yield $\mathbf{T(\ulcorner P \urcorner) \supset P}$; combined with (P1), that yields (C1). On an alternative reconstruction of the argument, falsity is simply defined as non-truth from the beginning, so that (P2) abbreviates $\mathbf{\sim P \supset \sim T(\ulcorner P \urcorner)}$, and (C1) follows without appeal to (P3). We have more to say below about such a definition of falsity. The former reconstruction has the merit of respecting the symmetry between truth and falsity in Tarski's explicit premises.

It is striking that Tarski bases his biconditional about truth partly on a claim about falsity. On either reconstruction, Tarski's premises also enable him to derive a corresponding disquotational biconditional about falsity:

(C2) $\mathbf{F(\ulcorner P \urcorner) \equiv \sim P}$

`Snow is white' is false if, and only if, snow is not white.

Generalizing on (C1), Tarski called a definition of truth `adequate' if all equivalences of the form `X is true if, and only if, p' follow from it, where `X' is replaced by a name of a sentence of the language for which truth is being defined and `p' is replaced by the sentence named. Similarly, generalizing on

(C2), we might call a definition of falsity ` adequate' if all equivalences of the form ` X is false if, and only if, not p' follow from it, where ` X' is replaced by the name of a sentence of the relevant language and ` not p' is replaced by a negation of the sentence named.

Tarski' s talk of the ` the classical conception of truth' refers back to a passage from Aristotle' s *Metaphysics* (Book Γ. 7, 1011^{a}: 26-28), quoted on the same page and elsewhere by Tarski:

> To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.

This passage might naturally be read as stating *sufficient* conditions for truth and falsity, exactly corresponding to (P1) and (P2). But since in the previous sentence Aristotle indicates that he is going to define truth and falsity, one would also hope to extract necessary and sufficient conditions from it, just as in (C1) and (C2) (alternatively, Aristotle might be read as defining truth and falsity directly).

Aristotle did not originate the classical conception. In Plato' s*Sophist* (240e-241a) falsity is characterized in just the same terms. Plato' s final definitions of true and false statement in that dialogue (263b) are only subtly different, and perhaps bring out the intention behind formulations such as Aristotle' sNevertheless, for convenience, we will continue to attach Aristotle' s name to the classical conception.

Given classical logic, we can easily deduce from Tarski' s premises that ` Snow is white' is either true or false. For we have **P** ∨ **~P** as an instance of the law of excluded middle, so, by the introduction and elimination rules for disjunction and *modus ponens*, (P1) and (P2) yield this:

(C3) **T(⌜P⌝) ∨ F(⌜P⌝)**

In effect, some such result was already clear to Aristotle; the passage quoted above continues ` so that he who says of anything that it is, or that it is not, will say either what is true or what is false' . Thus the

classical conception of truth and falsity seems to be straightforwardly committed to the principle of bivalence. But we must go deeper. Our formulations so far have taken no account of Plato and Aristotle's references to *saying*. They will turn out to serve an essential purpose, and to complicate the picture significantly.

Saying and disquotation. We need to develop a formal framework in which to conduct our investigation. Since we are concerned with vagueness in language, we will treat sentences as the bearers of truth and falsity. That is consistent with the idea that the truth-value of a sentence is determined by the truth-value of a proposition which it expresses. Of course, a sentence may express different propositions with respect to different contexts of utterance, if it contains indexicals or demonstratives such as `I' and `you' or `this' and `that'. Thus the truth-value of a sentence is also relative to a context of utterance.[3] Where no confusion results, we follow the common convention of using expressions of the formal language in English to refer to themselves. When using rather than mentioning such expressions as though they belonged to English, we indicate that we are doing so by not making them boldface. We use the expression **Say(s,c,P)** to mean something like: the sentence s as uttered in the context c says (or is used to say) that P (where `P' may be replaced by a declarative sentence); in terms of propositions, s in c expresses the proposition that P. We use **True(s,c)** (respectively, **False(s,c)**) to mean that s is true (respectively, false) in c.[4] One could easily extend the discussion from vagueness in language to vagueness in non-linguistic thought by letting **Say** mean the more general notion of expressing and **s** range over thought types as well as sentences.

Until the final section we will take for granted the principles of classical logic. We will help ourselves without comment to substitution instances of theorems of the classical propositional calculus and its extensions for the relevant types of quantifier, and use the rule of *modus ponens* similarly. Although these logical assumptions are of course not uncontroversial in the case of vagueness, justifying them is not our current concern. They are in any case not exclusive to a particular theory of vagueness; they are common to epistemicism, supervaluationism and some other views. Moreover, as it happens,

many - but not all - of our arguments will use only uncontroversial fragments of classical logic.

Within this framework, the natural principle about truth is that if a sentence says that something is so, then it is true if and only if that thing is so. More precisely:

(T) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \supset [\mathbf{True}(s,c) \equiv \mathbf{P}]]$

Unlike (C1), (T) is not itself a disquotational principle about truth, for it contains no occurrence of **P** in quotation marks (explicit or implicit). To recover a disquotational biconditional of the form $\mathbf{True}(\ulcorner \mathbf{P} \urcorner,c) \equiv \mathbf{P}$ from (T), we should need a corresponding auxiliary premise of the form $\mathbf{Say}(\ulcorner \mathbf{P} \urcorner,c,\mathbf{P})$. Such a premise holds if the sentence which replaces **P** contains no context-dependent elements. For example, if ` $E = mc^2$' is context-independent, then in any context it says that $E = mc^2$, and so is true if and only if $E = mc^2$. Even if **P** is replaced by a context-dependent sentence, the auxiliary premise $\mathbf{Say}(\ulcorner \mathbf{P} \urcorner,c,\mathbf{P})$ still holds if **c** refers to the context in which that premise is being asserted. For example, as uttered in the present context ` We are Europeans' says that we are Europeans, and so is true if and only if we are Europeans. But ` We are Europeans' as uttered by the Fondas does not say that we are Europeans, for *we* are Miroslava Andjelković and Timothy Williamson, and ` we' as uttered by the Fondas does not refer to us. In fact, we are Europeans but the sentence as uttered by the Fondas is not true. The disquotational biconditional holds only under restricted circumstances. But (T) does better; it handles context-dependence with ease. Since ` We are Europeans' as uttered by the Fondas does not say that we are Europeans, the corresponding instance of (T) is vacuously true. That sentence as uttered by them does say that they [the Fondas] are Europeans, and it is true as so uttered if and only if they are Europeans. Thus (T) is more basic than the disquotational biconditional; it explains both the successes and the failures of the latter.

A further advantage of (T) over the disquotational biconditional is that the latter but not the former must be revised to meet semantic paradoxes such as the Liar. For example, in some context c we can construct a self-referential sentence s to be $\mathbf{\sim True}(s,c)$. The corresponding disquotational

biconditional is **True(⌜~True(s,c)⌝,c)** ≡ **~True(s,c)**; since we have **s** = ⌜**~True(s,c)**⌝, we can deduce **True(s,c)** ≡ **~True(s,c)**, which is logically false. Thus not every instance of **True(⌜P⌝,c)** ≡ **P** holds. But this argument does not show that (T) is invalid; it merely falsifies the antecedent **Say(⌜~True(s,c)⌝,c,~True(s,c))**. Not even the sentence ` This sentence is not true' succeeds in saying of itself that it is not true.[5]

The variable **P** takes sentence position in (T). In particular, it flanks the biconditional ≡. Correspondingly, the third argument place in **Say(s,c,P)** is for a sentence, not a name (not even a name of a sentence); with respect to that argument place, **Say** works like an operator rather than a (first-level) predicate. Thus (T) involves quantification into sentence position. Such quantification might be interpreted substitutionally. The sentences substituted for **P** would be declarative sentences of the language in which the theorist is working (English, say). This would not restrict the instances of (T) which hold non-vacuously to those in which the sentence to which **s** refers is itself a sentence of English, for we can express in English what a sentence of some other language says. For example, the Serbian sentence ` Jedan sto je u sobi' as uttered in a suitable context says that one table is in the room, and so is true if and only if one table is in the room. However, if in some context a sentence s says something which cannot be expressed in English, at least in the context in which the theorist is working, then no instance of (T) holds non-vacuously for that case. (T) would still be true, but it would not be as informative about such cases as we should like it to be. We might therefore consider a non-substitutional reading of the quantifier ∀**P**, on which ∀**P Φ(P)** might be false even if **Φ(P)** is true whenever **P** is replaced by a sentence of the language in which the theorist is working. For example, given fixed assignments to the variables **s** and **c**, ∀**P ~Say(s,c,P)** might be false even though every corresponding substitution instance of the form **~Say(s,c,P)** was true. Such a non-substitutional interpretation would *not* automatically be objectual, for example over all propositions conceived as objects of a special kind. Objectual quantification is quantification whose semantics is given by non-substitutional quantification *into name position* in the metalanguage; both ` proposition' and ` object' are nouns, not sentences. One should not dismiss without argument the possibility that non-substitutional quantification into sentence

position is irreducible because its semantics can be given faithfully only by non-substitutional quantification into sentence position in the metalanguage too. It is not obviously wrong to suppose that one can understand such non-objectual non-substitutional quantification even if it cannot be expressed unequivocally in English; that might be a defect of English. We shall not attempt to decide between the substitutional and non-substitutional options here. Both are consistent with the arguments below. Even if the substitutional interpretation is not the intended one, it still provides a consistency proof of a standard theory of quantification into sentence position. The use made below of such quantification is quite undemanding, for sentence variables are not replaced by sentences that themselves contain either quantification into sentence position or one of the expressions **True**, **False** or **Say**.

Corresponding to (T), the natural principle about falsity is that if a sentence says that something is so, then it is false if and only if that thing is not so. More precisely:

(F) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \supset [\mathbf{False}(s,c) \equiv \sim\mathbf{P}]]$

For example, in a context in which the Serbian sentence ` Jedan sto je u sobi' says that one table is in the room, it is false if and only if it is not the case that one table is in the room. (F) raises the same issues of interpretation as (T); (F) stands to (C2) as (T) stands to (C1). In particular, we need the same auxiliary premise **Say(**⌜**P**⌝**,c,P)** to infer the disquotational biconditional **False(**⌜**P**⌝**,c)** ≡ **~P** from (F) as we do to infer the corresponding disquotational biconditional **True(**⌜**P**⌝**,c)** ≡ **P** from (T). We will have much more to say about both (T) and (F) below.

Bivalence. How should the principle of bivalence be formulated within this framework? The principle should not imply that non-declarative sentences are true or false, for presumably they are not intended to say that something is the case. For the same reason, the principle does not imply that a declarative sentence is true or false if it does not say that something is the case. For example, if a language teacher writes the sentence ` That belongs to her' on the board to illustrate a point of grammar, without attempting

to supply a reference for the demonstratives, the principle of bivalence does not require the sentence to be false. Thus we can reasonably build into the antecedent condition that the sentence says that something is the case, just as in (T) and (F). The usual formulation then says, on that condition, that the sentence is either true or false, where the disjunction is understood as inclusive. Since classical logic is being assumed, that is equivalent to the claim that the sentence is false if it is not true:[6]

(WB) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \supset [\sim\mathbf{True}(s,c) \supset \mathbf{False}(s,c)]]$

We call (WB) the principle of *Weak Bivalence*. It is weak because, by itself, it is consistent with the supposition that (for each fixed context) **True** and **False** stand for exactly the same property, one possessed by every sentence saying that something is so, even if there are many such sentences. In that case **True** and **False** would stand for one and the same truth-value, not two, and the term ` bivalence' would be a misnomer. We can strengthen weak bivalence by adding its converse, the principle that truth and falsity are mutually exclusive for a sentence that says that something is so (compare (P3) above):

(ME) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \supset [\mathbf{False}(s,c) \supset \sim\mathbf{True}(s,c)]]$

Suppose that a sentence is true only if it says that something is so; then we can drop the antecedent **Say(s,c,P)** and (ME) implies the simpler principle that truth and falsity are unconditionally mutually exclusive. We combine (WB) and (ME) into a single principle of *Strong Bivalence*:

(SB) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \supset [\sim\mathbf{True}(s,c) \equiv \mathbf{False}(s,c)]]$

In classical logic, we can reformulate the biconditional and negation as an exclusive disjunction: a sentence that says that something is so is either true or false and not both. (SB) stands to (WB) as exclusive stands to inclusive disjunction. For any context in which at least one sentence says that

something is so, (SB) requires **True** and **False** to stand for distinct properties, and the term ` bivalence' is appropriate. Contrary to the usual practice, we will concentrate on the strong rather than weak form of bivalence. Of course, (SB) is equivalent to (WB) given the almost universally accepted principle (ME). Nevertheless, (SB) does better than (WB) in making some logical and philosophical connections salient.[7] Dialetheists think that paradoxical sentences such as the Liar are both true and false; on some readings they accept (WB) but reject (SB) and (ME). We discuss the status of (SB), (WB) and (ME) below.

The logical relations between (T), (F) and (SB) are rather simple: any two of them entail the third. To check that, note that their main biconditionals are (classically equivalent to) **~True(s,c)** ≡ **~P**, **False(s,c)** ≡ **~P** and **~True(s,c)** ≡ **False(s,c)** respectively and that the biconditional behaves symmetrically and transitively. Thus, given (T), (F) is equivalent to (SB); given (F), (T) is equivalent to (SB); given (SB), (T) is equivalent to (F).

For future reference, it is convenient to label the two directions of (T) and (F) separately:

(T→) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say(s,c,P)} \supset [\mathbf{True(s,c)} \supset \mathbf{P}]]$

(T←) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say(s,c,P)} \supset [\mathbf{P} \supset \mathbf{True(s,c)}]]$

(F→) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say(s,c,P)} \supset [\mathbf{False(s,c)} \supset \mathbf{\sim P}]]$

(F←) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say(s,c,P)} \supset [\mathbf{\sim P} \supset \mathbf{False(s,c)}]]$

The Tarskian argument discussed above corresponds to the derivation of (T→) from (F←), (T←) and (ME). In effect, we also saw that Aristotle derived (WB) from (T←) and (F←). More generally, one can easily check that these logical relations hold amongst the unidirectional principles:

(T←) and (F←) entail (WB);

(T→) and (F→) entail (ME);

(F←) and (ME) entail (T→);

(F→) and (WB) entail (T←);

(T←) and (ME) entail (F→);

(T→) and (WB) entail (F←).

By considering unintended interpretations of **True** and **False** and applying the Peano-Padoa method, one can also check that no further entailments obtain amongst the six halves of (T), (F) and (SB) beyond those implicit in the foregoing list, if **True** and **False** are treated as primitive predicates, not as logical constants. For example, if one interprets **False** normally but **True** as applying to nothing, then (ME), (T→), (F→) and (F←) all hold whilst (WB) and (T←) do not hold, so neither (WB) nor (T←) follows from (ME), (T→), (F→) and (F←). Thus a subset of the six principles non-trivially entails (WB) (that is, without containing (WB) itself) only if it contains (T←). Similarly, if one interprets **True** normally but **False** as applying to nothing, then (ME), (T→), (T←) and (F→) all hold whilst (WB) and (F←) do not hold, so neither (WB) nor (F←) follows from (ME), (T→), (T←) and (F→). Thus a subset of the six principles non-trivially entails (WB) only if it contains (F←). Therefore, a subset of the six principles non-trivially entails (WB) only if it contains both (T←) and (F←), in which case the entailment is reducible to the first item on the list above (one noted by Aristotle).

Epistemicists about vagueness typically argue from (T) and (F) to (SB), or more specifically, following Aristotle, from (T←) and (F←) to (WB) (Williamson 1994: 188-189 and 1995). We will not discuss here whether they are entitled to assume (F) in arguing for (SB). For present purposes, what we should note is that someone could consistently accept both (T) and classical logic while still rejecting (F), and therefore (SB), and therefore epistemicism.[8] Thus the notion of falsity should play a significant role in the discussion of vagueness; principles such as (F) need explicit attention. Unfortunately, the literature has neglected (F) and cognate principles about falsity. We will do something about that deficiency here.

Falsity and negation. Can the notion of falsity be defined in terms of the notion of truth? If so, we might be able to use the definition to reduce (F) to (T), and therefore to derive (SB) from (T). In that case, the neglect of the notion of falsity would be more apparent than real, since the notion would have been treated implicitly in the treatment of the notion of truth. Presumably such a definition should not use auxiliary conceptual resources so strong that they would suffice by themselves for a direct definition of falsity without the detour through truth.

The simplest attempted definition of falsity is as non-truth: $\mathbf{False(s,c)} =_{def} \mathbf{\sim True(s,c)}$. It was used in the second reconstruction above of Tarski's argument from (P1) and (P2) to (C1). Under this definition, (SB) becomes a trivial logical truth, and (F) reduces to (T). However, the definition has very implausible consequences. A sentence which does not say that something is so counts as false simply because it is not true. Indeed, mountains and lakes count as false for the same reason. We could revise the definition by defining falsity as the conjunction of non-truth with saying that something is so: $\mathbf{False(s,c)} =_{def} \mathbf{\sim True(s,c) \,\&\, \exists P\, Say(s,c,P)}$. According to epistemicism, such a definition would give at least extensionally correct results. But it would not be acceptable to those who reject (SB); a less controversial definition would be preferable. Moreover, it employs quantification into sentence position; given such quantification, one might as well define falsity directly, rather than making an unnecessary detour through the notion of truth (see below).

Falsity is often defined as the truth of the negation. For example, Michael Dummett writes (1981: 108):

> It is commonly observed that our intuitive application of the term `false' is largely governed by the principle that a statement is false if and only if its negation is true, supplemented by a general disposition on our part to construe as the negation of a statement the simplest plausible candidate for that role.[9]

We discuss below what counts as the negation of a sentence. Such a definition of falsity may be

understood as implying that an item is false only if it has a true negation, and therefore only if it has a negation, so that it is not false if it has no negation at all. Since lakes and mountains have no negations, they are not false. If a sentence s fails to say that something is so, then either s lacks a negation or its negation also fails to say that something is so, for what the negation of s says would be the contradictory of what s says; either way, s does not count as false.

Let us use the singular term **Ns** to refer to a negation of the sentence to which the term **s** refers. If the latter has several verbally different negations, we do not worry which of them the former refers to. Thus the definition of falsity as the truth of the negation yields this principle:

(FN) $\forall s \forall c[\mathbf{False}(s,c) \equiv \mathbf{True}(Ns,c)]$

We do not make the biconditional in (FN) conditional on the antecedent **Say(s,c,P)**, for if s fails to say that something is so then, as noted above, Ns also fails to say that something is so; consequently, neither **False(s,c)** nor **True(Ns,c)** holds, and therefore the biconditional holds vacuously.

Obviously, (FN) does not help unless we can independently characterize negation (**N**). The easiest way to do so is by using negation in the metalanguage to say what Ns says:

(N) $\forall s \forall c \forall P[\mathbf{Say}(s,c,P) \equiv \mathbf{Say}(Ns,c,\sim P)]$

Note that, given (N), we cannot automatically assume that a sentence is the negation of its negation. For two applications of (N) yield **Say(s,c,P)** ≡ **Say(NNs,c,~~P)**; on a fine-grained notion of saying, **Say(s,c,~~P)** is not equivalent to **Say(s,c,P)** (we might still allow that a sentence is a *contradictory* of its negation). On a very fine-grained view, to which we do not commit ourselves here, the negation of a sentence consists of a negation operator and that sentence, the former applied to the latter.

Given (FN) and (N), we can reduce (F) to (T) and thereby derive (SB). For as a special case of (T) we have **Say(Ns,c,~P)** ⊃ **[True(Ns,c)** ≡ **~P]**, which by (N) yields **Say(s,c,P)** ⊃ **[True(Ns,c)** ≡ **~P]**,

which by (FN) yields **Say(s,c,P)** ⊃ **[False(s,c)** ≡ **~P]** and therefore (F). Once we have (T) and (F), we can derive (SB).

Even without (FN), we can use (T) and (N) to derive a principle related to (SB) (although without the implication that the sentences takes one of two values) with **True(Ns,c)** in place of **False(s,c)**: in a context in which a sentence says that something is so, either it or its negation is true. Thus the rejection of (F) by itself would not avoid all the controversial consequences of an epistemic theory of vagueness, since one would still be committed to the claim that when a vague sentence in a borderline case says that something is so, either it or its negation is true, and not both, even though we have no idea how to find out which of the two contradictory sentences is true. However, not all opponents of epistemicism need have the same ultimate grounds; different opponents may reject different consequences. Let us return to the original principle (SB). To assess it, we must consider the notion of falsity itself, and therefore (FN).

What happens to (FN) if s has no negation? If s is not a candidate for truth or falsity - for example, if s is a lake or mountain - then (FN) holds vacuously. But might not s be a candidate for truth or falsity and still have no negation, because it is in an expressively impoverished language? Creatures might in principle communicate information about their environment to each other in a limited system of signals without negation. Perhaps some animals actually do so. Such signals could be true or false, depending on the relation between the signalled state of the environment and its real state. Suppose, for example, that the animals can signal that there is food over there. Thus, when there is food over there, they can communicate truly that there is food over there. When there is no food over there, they can communicate falsely that there is food over there. That does not require them to have the capacity to negate a signal; they may be unable to signal that there is no food over there. Of course, questions can be raised about the legitimacy of attributing propositional content to signals in so simple a system. Nevertheless, we see no decisive obstacle to such attributions. A creature or other system can hardly engage in behaviour that has the function of communicating true information unless it also has the power to get things wrong, the power to represent something as so when it is not so. The capacity to represent

(in a full-blooded sense of the term) involves the capacity to misrepresent.[10] If the capacity to negate signals is not strictly necessary for communicating true information, it is also not strictly necessary for communicating false information.

A friend of (FN) might reply that a negation operator N can always in principle be added to the system, so that one can define the falsity of s as the truth of its hypothetical negation Ns. This proposal must be formulated with some care, for the falsity of s in a given situation is equivalent to the truth of its negation in that very situation, not in a counterfactual system in which negation has been introduced into the system. For the presence of negation in the signal system might indirectly affect the aspect of the environment about which information is being communicated, by affecting the behaviour of the animals.[11]

One could construct an artificial code with a self-referential signal that says falsely of itself that it has a negation. In the counterfactual situation in which that signal has a negation, it is true. Presumably such a code could not be anyone' s first language, but it is not clear why that should be relevant.

In practice, the introduction of a negation operator can alter the use of the original signals. One reason is that the absence of a signal can itself pragmatically communicate the information that the situation is not of a kind to evoke a signal. The significance of that information depends on whether a negative signal is available as well as a positive one. Suppose, for example, that signals convey information about how full a waterhole is. If only a positive signal is available, it might be most efficient to emit the signal when the waterhole is at least half full, for then in appropriate circumstances the absence of a signal will convey the information that the waterhole is less than half full. Thus the system of two options {signal, no signal} partitions degrees of fullness from 0 to 1 into two intervals [0,½) and [½,1], each of width ½ (where the width of a set is defined as the least upper bound of the distances between its members). That is optimal in the sense that the closed interval [0,1] cannot be partitioned into two sets each of width less than ½. The width of the set in which a signal is emitted is some measure of the signal' s informativeness. But the whole interval can be partitioned into three intervals each of width ⅓ and cannot be partitioned into three sets each of width less than ⅓. Thus if there is a system of

three available options {positive signal, negative signal, no signal}, it might be most efficient to emit the positive signal when the waterhole is at least two-thirds full, the negative signal when it is at most a third full, and no signal in between. This simplified example shows that the pragmatic significance of the positive signal is sensitive to the availability of the negative symbol. In the long run this pragmatic sensitivity might affect the semantic properties of the positive signal, in particular its truth-conditions and falsity-conditions.

Intuitively, the complications of working out what might happen in counterfactual circumstances in which a signal had a negation are quite irrelevant to the truth or falsity of its actual tokens. Nor is the definition of the falsity of s as the *actual* truth of a merely *possible* or abstract negation of s particularly attractive. There is a simpler way.

We propose to treat falsity as a primary notion on a par with truth. If negation has a role to play in the characterization of falsity, it is negation as used in the metalanguage by the theorist, not negation as used in the object-language by the speakers under study. For example, ~ occurs in (F) as the metalanguage negation, whereas **N** is mentioned in (FN) as the object-language negation. Given (T) and (N), (F) is equivalent to (FN) whenever **Ns** is well-defined, so (F) and (FN) do not compete for correctness, although they may compete for primacy.

Even in a language without negation, vagueness can arise in virtue of borderline cases. In some contexts, a negationless signal may be neither clearly true nor clearly false. A sorites series may be possible for it: for instance, a series of contexts, each indiscriminable from the next, in the first of which it is true and in the last of which it is false. Although one needs logical constants to formulate the sorites reasoning explicitly, one does not need them merely to have difficulty in using the signal correctly in such a series of contexts. For example, the case may be classified differently depending on which end of a sorites series it is approached from. Without the logical constants, one cannot properly reflect on the significance of the phenomena of vagueness, but at least some of the phenomena themselves can occur.

The possibility of vague but negationless signals also suggests that borderline cases should not

primarily be conceived as those in which speakers are poised symmetrically between a positive and a negative assertion, since the latter is not always an option. Certainly the pragmatic significance of the absence of a positive assertion does not give it the semantic significance of a negative assertion.[12]

The concepts of truth and falsity are parallel in most respects, but not in all, at least as they are explained by (T) and (F). For (F) uses negation in the metalanguage at a point at which (T) does not use any operator at all. We have not defended the use of negation rather than some other operator to define falsity, although it seems very natural; if (F) is coherent, what does it characterize other than falsity? By substituting other operators in the metalanguage for ~, one can construct a variety of other concepts. For example, if ~ is replaced by □, read ` necessarily' , the result characterizes the concept of necessary truth. If ~ is replaced by Δ~, where Δ is read ` clearly' , a concept of clear falsity is characterized, and one should not expect to derive the analogues of (SB) or (WB). Within classical logic, the principle of bivalence stands or falls with the connection between falsity and standard negation.

Definitions of truth and falsity. One might be tempted to regard (T) and (F) as implicit definitions of **True** and **False** respectively. That would be a mistake. Even granted the correctness of (T) and (F), in one respect they are too weak to constitute definitions; in another respect they are too strong to do so. They guarantee neither uniqueness nor existence.

In what follows we will treat **True** and **False** as so far uninterpreted symbols, but the other primitive terms of the formal language as already interpreted. We can therefore ask how various principles constrain the interpretation of **True** and **False**. (T) and (F) are too weak to define **True** and **False** because they are quite neutral about the application of those terms to a sentence when it says nothing. Presumably a sentence is neither true nor false in a context in which it says nothing, but (T) and (F) impose no such constraint. They are equally consistent with the unappealing supposition that a sentence is both true and false in a context in which it says nothing. Consequently, they do not characterize **True** and **False** uniquely. To be more precise, let (T_1) and (T_2) be the results of subscripting **True** in (T) by ` 1' and ` 2' respectively. Then from (T1) and (T_2) on the assumption **Say(s,c,P)** one can

deduce $\mathbf{True_1(s,c)} \equiv \mathbf{P}$ and $\mathbf{True_2(s,c)} \equiv \mathbf{P}$ and therefore $\mathbf{True_1(s,c)} \equiv \mathbf{True_2(s,c)}$, giving:

(T!-) $\forall \mathbf{s} \forall \mathbf{c} \forall \mathbf{P}[\mathbf{Say(s,c,P)} \supset [\mathbf{True_1(s,c)} \equiv \mathbf{True_2(s,c)}]]$.

But one cannot drop the assumption and assert unconditionally:

(T!) $\forall \mathbf{s} \forall \mathbf{c} \forall [\mathbf{True_1(s,c)} \equiv \mathbf{True_2(s,c)}]$.

Thus truth is not *the* value of **True** which satisfies (T), because there is no such unique value. Similarly, if (F_1) and (F_2) are the results of subscripting **False** in (F) accordingly, they entail:

(F!-) $\forall \mathbf{s} \forall \mathbf{c} \forall \mathbf{P}[\mathbf{Say(s,c,P)} \supset [\mathbf{False_1(s,c)} \equiv \mathbf{False_2(s,c)}]]$.

But one cannot drop the assumption and assert unconditionally:

(F!) $\forall \mathbf{s} \forall \mathbf{c} \forall [\mathbf{False_1(s,c)} \equiv \mathbf{False_2(s,c)}]$.

Thus falsity is not *the* value of **False** which satisfies (F), because there is no such unique value. We can fill these gaps by explicitly stipulating that a sentence is true or false only if it says something:

(T+) $\forall \mathbf{s} \forall \mathbf{c}[\mathbf{True(s,c)} \supset \exists \mathbf{P\ Say(s,c,P)}]$

(F+) $\forall \mathbf{s} \forall \mathbf{c}[\mathbf{False(s,c)} \supset \exists \mathbf{P\ Say(s,c,P)}]$

With these additions, truth and falsity are uniquely characterized, in the sense that (T!) follows from (T_1), (T_2), (T_1+) and (T_2+) (the last two being the results of subscripting **True** in (T+) accordingly), and (F!)

follows from (F_1), (F_2), (F_1+) and (F_2+). It is worth noting that (T+) and (F+) rely more heavily on quantification into sentence position than do (T) and (F). For one could drop the universal quantifiers from (T) and (F) and treat the remainders simply as schemata, whereas no such treatment is possible for the existential quantifiers in (T+) and (F+).

A deeper problem is that (T) and (F) are too strong to be mere definitions of **True** and **False**, because they are *creative*. That is, (T) has a consequence not involving **True** and (F) has a consequence not involving **False**. More specifically, each of them entails a principle of *uniformity* to the effect (at least in the classical context) that everything said by a given sentence in a given context has the same truth-value:

(U) $\forall s \forall c \forall \mathbf{P} \forall \mathbf{Q}[[\mathbf{Say}(s,c,\mathbf{P}) \,\&\, \mathbf{Say}(s,c,\mathbf{Q})] \supset [\mathbf{P} \equiv \mathbf{Q}]]$

For the antecedent of (U) yields $\mathbf{True}(s,c) \equiv \mathbf{P}$ and $\mathbf{True}(s,c) \equiv \mathbf{Q}$ by instances of (T), from which $\mathbf{P} \equiv \mathbf{Q}$ follows. Similarly, the antecedent of (U) yields $\mathbf{False}(s,c) \equiv \mathbf{\sim P}$ and $\mathbf{False}(s,c) \equiv \mathbf{\sim Q}$ by instances of (F), from which $\mathbf{\sim P} \equiv \mathbf{\sim Q}$ follows, and therefore $\mathbf{P} \equiv \mathbf{Q}$ too.

Let L be a language without the predicates **True** and **False** but in which (U) can be formulated, and consider a theory Θ in L of which (U) is not a theorem. If one adds **True** to L and (T) to Θ, the result is not a conservative extension of Θ, because (U) is a theorem of the new theory in the old language without being a theorem of the old theory. For the same reason, if one adds **False** to L and (F) to Θ, the result is again not a conservative extension of Θ. If (T) and (F) were genuine definitions, their addition would yield conservative extensions of Θ. They behave more like theories than definitions.

A question now arises for the epistemicist derivation of the bivalence principles (SB) and (WB). It relies on (T) and (F), which are logically stronger than definitions of **True** and **False**. Thus the opponent of epistemicism might accept classical logic but block the derivation of (SB) and (WB) by rejecting (T) or (F) or both, while insisting that this does not amount to redefining **True** and **False**. In particular, the anti-epistemicist might reject (U), and therefore both (T) and (F), on the grounds that (U)

fails in borderline cases for a vague sentence s. Perhaps they will suggest that a vague sentence says many things, corresponding to its different possible sharpenings; in a borderline case, some of these things differ from others in truth-value, contrary to (U).

We will postpone asking whether (U) really does fail in borderline cases, and suppose for the sake of argument that it does. If so, what happens to the bivalence principles (SB) and (WB)? In order to answer that question, we need characterizations of truth and falsity. (T) and (F) will not do for present purposes because they entail (U), which is being supposed to fail. We therefore need to characterize truth and falsity in some alternative way. Within the present framework, the natural idea is to define truth and falsity explicitly by quantifying into sentence position.

A standard proposal is that a sentence is true if something that it says to be so is so. The biconditional corresponding to such a definition is this:[13]

(TDEF1) $\forall s \forall c[\mathbf{True}(s,c) \equiv \exists P[\mathbf{Say}(s,c,P) \ \& \ P]]$

The obvious analogue for falsity of such a definition is that a sentence is false if something that it says to be so is not so. The corresponding biconditional stands to (F) as (TDEF1) stands to (T):

(FDEF1) $\forall s \forall c[\mathbf{False}(s,c) \equiv \exists P[\mathbf{Say}(s,c,P) \ \& \ \sim P]]$

Like (T+) and (F+), both (TDEF1) and (FDEF1) make much more essential use of quantification into sentence position than do (T) and (F). Schemata without such quantification are no substitute for (TDEF1) and (FDEF1).

Before continuing with the main line of argument, we pause to consider an objection. By explicitly defining truth, for instance by (TDEF1), are we not liable to incur inconsistency, by Tarski's theorem on the undefinability of truth? Fortunately, we can avoid this danger by not permitting the unrestricted application of the expression **Say** to sentences of the metalanguage as values of the term in

its first argument-place (**s**). This still permits **Say** to be applied to sentences of many diverse languages, provided that their content is not of a special metalinguistic kind. Such a restriction requires extensive discussion, but our present concern is not with the semantic paradoxes. Quantification into sentence position does not itself introduce any inconsistency. As noted above, one can give it a consistency proof by using a substitutional interpretation. Indeed, one could give an alternative consistency proof by using a different unintended interpetation in which the semantic value of the sentential variable **P** is simply its truth-value, although that would not permit **Say** to behave in the intended way, as an intensional operator with respect to **P**. One could accommodate some of that intensionality by treating the semantic value of **P** as the set of possible worlds in which it is true, although even that would not capture its full intended meaning, since we take **Say** to behave *hyper*intensionally: a sentence may say that P without saying that Q even though in all possible worlds P if and only if Q. For example, ` 2+2=4' says that 2+2=4 without saying that 7+5=12, even though all mathematical truths are true in all possible worlds and therefore in the same possible worlds. Given the limited nature of the present use of quantification into sentence position, as noted above, we can consistently treat the semantic values of variables in sentence position as structured or otherwise finely grained propositions, thereby allowing for hyperintensionality. We will not pursue these issues here, but simply repeat that none of the logical framework within which we are operating falls to a semantic paradox.

Since (TDEF1) and (FDEF1) are in effect definitions, they do not entail (U). Therefore, they do not entail (T) or (F). More specifically, one can easily check that (TDEF1) entails (T←), and therefore does not by itself entail (T→) (for if it did, it would entail (T)), and that (FDEF1) entails (F←), and therefore does not by itself entail (F→) (for if it did, it would entail (F)).

We can check that (TDEF1) and (U) together do entail (T→), and therefore (T), by an obvious argument. Similarly, (FDEF1) and (U) together entail (F→), and therefore (F). This shows that (U) is the *strongest* consequence of the conjunction of (T) and (F) to contain neither **True** nor **False**, in the sense that if **A** contains neither **True** nor **False** and (T) and (F) together entail **A** then (U) entails **A**. For (U), (TDEF1) and (FDEF1) together entail (T) and (F); thus if (T) and (F) entail **A**, then by transitivity (U),

(TDEF1) and (FDEF1) entail **A**; but if we replace **True** and **False** throughout the deduction by the right-hand sides of (TDEF1) and (FDEF1) respectively, the result trivializes (TDEF1) and (FDEF1) and is still a valid deduction of **A** from (U), since the premises and conclusion do not contain **True** or **False**. Thus (U) exhausts the non-conservative aspect of (T) and (F).

What are the implications of (TDEF1) and (FDEF1) for bivalence? As already noted, they entail (T←) and (F←) respectively, which together entail (WB), so (TDEF1) and (FDEF1) jointly yield weak bivalence. But they do not yield the strong bivalence principle (SB), because they do not yield (ME). In fact, one can easily show that if any sentence is a counter-example to (U), then (TDEF1) counts it as true and (FDEF1) counts it as false. Thus if (U) fails for vague sentences in borderline cases, (TDEF1) and (FDEF1) make those sentences both true and false. Such a consequence would generally be regarded as unacceptable.

One or two theorists of vagueness do suggest that a vague sentence is true if and only if at least one of its admissible sharpenings is true, and is false if and only if at least one of its admissible sharpenings is false (Hyde 1997). Such a view is know as *subvaluationism*, by analogy with supervaluationism (discussed below). Subvaluationism resembles the envisaged combination of (TDEF1) and (FDEF1) with the negation of (U). The various things said by a sentence correspond to its various sharpenings. However, subvaluationism has highly counter-intuitive consequences. Consider, for example, `small' as a vague predicate of natural numbers in a context in which one admissible sharpening puts its cut-off point between *n*-1 and *n* while another equally admissible sharpening puts the cut-off point between *n* and *n*+1. Then the conjunctions `*n*-1 is small and *n* is not small' and *n* is small and *n*+1 is not small' both count as true according to subvaluationism, because each is true on at least one sharpening. Moreover, since we might know that both sharpenings were admissible, we might know that both conjunctions were true. Presumably, therefore, we should assert each of them if the question arises. But we are not entitled to assert (rather than stipulate) of any place in particular that the cut-off point for a vague predicate comes there. Still less are we entitled to assert both conjunctions, for they are mutually inconsistent; not even subvaluationism makes their conjunction `*n*-1 is small and *n* is not small and *n* is

small and n+1 is not small' true. Consequently, we reject subvaluationism. It faces all the main problems of supervaluationism and more besides. In what follows we assume the principle (ME) that truth and falsity are mutually exclusive. Thus, in the absence of (U), we reject the combination of (TDEF1) and (FDEF1).

One could preserve both (ME) and either one of (TDEF1) and (FDEF1) while rejecting (U) by treating truth and falsity as contradictories. For example, suppose that one accepts (TDEF1). Then to treat falsity as non-truth is to accept the dual of (FDEF1):

(FDEF2) $\forall$**s**$\forall$**c[False(s,c)** $\equiv$ $\forall$**P[Say(s,c,P)** $\supset$ **~P]]**

A sentence is false if nothing that it says to be so is so. (TDEF1) and (FDEF2) jointly entail the full bivalence principle (SB) itself, and therefore both (ME) and (WB). But since they merely define **True** and **False**, they do not entail (U). Thus they do not jointly entail (T) or (F). In particular, (FDEF2) does not entail (F): although it entails (F$\rightarrow$), it does not entail (F$\leftarrow$). In any context in which (U) fails, the relevant sentence counts as true by (TDEF1) and as not false by (FDEF2). Therefore, if (U) fails for all vague sentences in borderline cases, all such sentences count as true and not false - and one can know that. Thus the combination of (TDEF1), (FDEF2) and the negation of (U) represents a route to strong bivalence incompatible with that taken by the epistemicist.

A similar route to strong bivalence equally inconsistent with epistemicism combines (FDEF1) with the negation of (U) and a definition of truth as non-falsity. It yields the dual of (TDEF1):

(TDEF2) $\forall$**s**$\forall$**c[True(s,c)** $\equiv$ $\forall$**P[Say(s,c,P)** $\supset$ **P]]**

A sentence is true if everything that it says to be so is so. (TDEF2) and (FDEF1) jointly entail (SB), and therefore both (ME) and (WB). Since they merely define **True** and **False**, they do not entail (U) or (T) or (F). In particular, (TDEF2) does not entail (T): although it entails (T$\rightarrow$), it does not entail (T$\leftarrow$). In any

context in which (U) fails, the relevant sentence counts as not true by (TDEF2) and as false by (FDEF2). Therefore, if (U) fails for vague sentences in borderline cases, such sentences count as false and not true - and one can know that.

Although the two combinations just considered secure strong bivalence, they do so at a high cost to classical intuitions. They undermine the natural conception of truth and falsity as somehow parallel notions: one of the defining conditions is quantified universally, the other existentially. Moreover, the presence of just one of (TDEF1) and (FDEF1) suffices to yield some of the counter-intuitive consequences of subvaluationism. For example, given the persuasive principle (N) linking what a sentence says to what its negation says, one can show that in any context in which a sentence is a counter-example to (U), so is its negation: from **Say(s,c,P) & Say(s,c,Q)** one can derive **Say(Ns,c,~P) & Say(Ns,c,~Q)**. Now suppose that (U) fails in a borderline case *n* for the vague predicate ` small' . Then (TDEF1) counts both `*n* is small' and*ì* is not small' as true. That (FDEF2) does not count `*n* is small' as false too is not much consolation, for that is achieved only by not counting a sentence with a true negation as false; (FN) is violated from right to left. Similarly, (FDEF1) counts both `*n* is small' and*ì* is not small' as false. That (TDEF2) does not count*ì* is not small' as true too is equally little consolation, for that is achieved only by not counting the negation of a false sentence as true; (FN) is violated from left to right. Both (TDEF1) and (FDEF1) are to be rejected. What remains is the combination of (TDEF2) and (FDEF2).

Given (TDEF2) and (FDEF2), failures of (U) produce truth-value gaps rather than truth-value gluts. If vague sentences in borderline cases are counter-examples to (U), then such sentences are neither true nor false. This view has an obvious analogy with supervaluationism, on which a sentence is true if and only if all its admissible sharpenings are true, and is false if and only if all its admissible sharpenings are false. As with subvaluationism, the various things said by a sentence would correspond to its various sharpenings.

An immediate problem is that (TDEF2) and (FDEF2) together do not make truth and falsity mutually exclusive, even in the presence of (U). If a sentence says nothing, they count it as vacuously

both true and false. Indeed, they count mountains and lakes as both true and false. To overcome this problem, one can add clauses requiring that a sentence is true or false only if it says something:

(TDEF2*) $\forall s \forall c[\mathbf{True(s,c)} \equiv [\exists \mathbf{P\ Say(s,c,P)} \ \mathbf{\&}\ \forall \mathbf{P[Say(s,c,P)} \supset \mathbf{P]]]}$

(FDEF2*) $\forall s \forall c[\mathbf{False(s,c)} \equiv [\exists \mathbf{P\ Say(s,c,P)} \ \mathbf{\&}\ \forall \mathbf{P[Say(s,c,P)} \supset \mathbf{\sim P]]]}$

This modification preserves the analogy with supervaluationism; the latter does not permit sentences to be vacuously both true and false, for admissible sharpenings are defined at the level of a language as a whole and supervaluationists assume that there is always at least one such sharpening. (TDEF2*) and (FDEF2*) jointly entail (ME), (T→) and (F→); they do not entail (U), (T), (T←), (F), (F←), (WB) or (SB). Since counter-examples to (U) must satisfy the extra clause $\exists \mathbf{P\ Say(s,c,P)}$, they still count as neither true nor false by (TDEF2*) and (FDEF2*). If borderline cases falsify (U), they are classified as involving truth-value gaps.

The case for the uniformity principle. In the preceding discussion we assumed, for the sake of argument, that (U) can fail: a sentence may in the same context say something that is so and something that is not so. If so, the supposed failure appears to give succour to supervaluationism. It is time to evaluate the assumption that (U) can fail.

There is a non-technical notion of saying on which to say something can also be to say some of its immediate logical consequences. In that sense, saying that it is cold and wet might involve both saying that it is cold and saying that it is wet; saying that Nenad is absent might involve saying that someone is absent. That notion of saying is clearly irrelevant to the present problem, for it would yield counter-examples to (U) even in unproblematic non-borderline cases. For example, perhaps it is cold but not wet; perhaps Nenad is present but someone else is absent. Since the relevant instances of (T) and (F) should hold in such unproblematic cases, so should their consequences, the relevant instances of (U).

Thus we should interpret **Say(s,c,P)** in (T), (F) and (U) to mean something like: s says in c *just* that P. But then how can (U) fail? If, in a given context, a sentence says *just* that P and says *just* that Q, how could the proposition that P be anything other than the proposition that Q? Similarly, if we read **Say(s,c,P)** as something like ` The propositional content of s in c is that P' , then the uniqueness implied by the definite article leaves no obvious room for (U) to fail.

Let us consider an example. Mary is thinking about a very important decision that she is going to make. John is playing with dice. In the first situation, they are in the same room but completely unaware of each other. If asked ` Is the die cast?' , John will treat the interrogative sentence as raising the question whether the die has been thrown, and Mary as raising the question whether the decision has been made. John answers ` No' since the die is still in his hand, and Mary, who does not see or hear John, answers ` Yes' . Obviously one sentence can say two different things at the same time, but can we make it say them in the same context? Let us give the die to Mary, and ask her verbally the same question immediately afterwards. We know that Mary has made up her mind, since her answer was ` Yes' the first time. But now the die is in her hand and the question is repeated. She is confused. Why would she be confused if this were not a context in which the interrogative sentence asks two different things? What can she answer? It would be sensible for her to say ` No, but the decision is made' . By putting it that way she gives a privileged status to the literal meaning of the question, but she answers it in both its senses. Now, being completely sure that Mary is aware of the two meanings of the question ` Is the die cast?' , we may be confused when Mary without being asked says ` The die is cast' . We cannot so easily say that the sentence must mean that the die has been thrown. It might mean that the decision has been made, it might even mean both. But when can it mean both? If she wants to convey more than one piece of information, she will often believe that all the pieces of information are correct, and if she is right the uniformity principle will hold. But of course she might be lying or mistaken on one matter without being so on the other. Whatever she wants to convey, she can hope for success only if the pieces of information do not contradict each other, but propositions that are not contradictories or contraries can still have opposite truth-values; the proposition that the decision has been made is not contradictory or contrary

to the proposition that the die has been thrown, but one may be true while the other is false. Thus the uniformity principle is still threatened.

Suppose that Mary says ` The die is cast' , intending thereby both to say that the die has been thrown and to say that the decision has been made. Does her sentence say two things in the same context? If she is mistaken about the die but not about the decision, does that falsify (U)? Perhaps Mary intends to convey the conjunction of two pieces of information. She might have achieved the same end less economically by saying ` The die is cast and the die is cast' , meaning by the first conjunct that the die has been thrown and by the second that the decision has been made. The conjunction is simply false because its first conjunct is false. She has said more than one thing in the non-technical sense noted above in which, in saying that it is cold and wet, one both says that it is cold and says that it is wet. The difference is that in Mary' s case the two conjuncts lack separate surface representations. Presumably they have separate representations at the level of logical form. Mary' s intention makes it clear that in *this* context the sentence expresses a single proposition, a conjunction. If the question ` Is the die cast?' is asked in a similar sense, the correct answer is simply ` No' . Now consider the alternative case: in saying ` The die is cast' , Mary intends to perform two separate speech acts. In uttering the interrogative sentence ` Is the die cast?' in a similar way, she would intend to ask two separate questions. In this case, it is not unreasonable to postulate two simultaneous contexts with the same speaker, one corresponding to her intention to say that the die has been thrown, the other to her intention to say that the decision has been made. Since the difference in intention is built into the description of the case, an appeal to it is well-motivated, not *ad hoc*. Contexts can differ in subtle as well as obvious ways. Alternatively, if one individuates sentences semantically, one might even argue that the phonetic string ` The die is cast' represents two semantically distinguished sentences. Either way, (U) is preserved without being trivialized. We will *not* brush aside challenges to (U) from vagueness by multiplying contexts or sentences.

Suppose that Harry is a borderline case for baldness. It is natural to hold that the sentence ` Harry is bald' says in the present context that Harry is bald. But, if so, why think that it says something else

too? *What* else does it say? If the sentence does not say that Harry is bald, the question arises again: what does it say? `Harry is bald' seems to say just that Harry is bald. Even the semantic paradoxes do not seem to be counter-examples to the qualified disquotational principle that *if* the sentence ⌜P⌝ says anything at all in this context, it says just that P. Thus it is hard to understand how (U) could fail in the present context. But if it cannot fail in the present context, then it cannot fail in any ordinary context, for there is nothing special about the present context. Since borderline cases such as Harry occur in the present context, vagueness would not falsify (U).

Someone might use the supervaluationist or subvaluationist notion of an admissible sharpening in an attempt to explain a sense of **Say** on which (U) could fail: a sentence would say its admissible sharpenings. Such an explanation is only as clear as the notion of an admissible sharpening. Supervaluationists have great difficulty in giving an adequate account of that notion in such a way that it does not reduce to something epistemic (Williamson 1994: 164, 1995 and 1997: 216-217, the last of which responds to McGee and McLaughlin 1995).

Fortunately, the uniformity principle (U) can be supported by arguments more rigorous than the foregoing remarks. Our strategy is as follows. We start with (TDEF2*), a definition of truth of the sort that someone who rejects (U) might accept, and show that under plausible assumptions about compositionality it still leads to (U). Since (TDEF2*) makes **True** behave in a distinctively supervaluationist fashion only if (U) fails, the appearance of support lent by (TDEF2*) to supervaluationism is illusory. We will also show that under plausible assumptions about compositionality (TDEF1) too leads to (U). Thus the appearance of support lent by (TDEF1) to subvaluationism is equally illusory.

Our object assumes that sentences s and t can be connected by a material biconditional into a sentence Est (≡ is the material biconditional in the metalanguage). Although the argument does not apply directly to sentences in a language without such a connective, that is not a serious limitation. For a language with a material biconditional may contain arbitrarily vague sentences, or sentences with any other features that might be thought to undermine (U). Thus if (U) holds for languages with a material

biconditional, the relevant theories that predict the failure of (U) are false, and there is no longer any reason to reject (U). For notational simplicity, we will not make explicit provision in the formal argument for sentences in languages without a biconditional.

We will now argue from (TDEF2*) to (U). We first note a plausible principle of compositional semantics. Roughly, since E expresses ≡, Est says the biconditional of what s says and what t says:[14]

(E1) $\forall s \forall t \forall c \forall P \forall Q[[Say(s,c,P) \,\&\, Say(t,c,Q)] \supset Say(Est,c,P \equiv Q)]]$

Next, we note that if a sentence says something then its biconditional with itself is true:

(E2) $\forall s \forall c \forall P[Say(s,c,P) \supset True(Ess,c)]$

Supervaluationists would certainly accept (E2), since every sharpening of Ess is true; it is a classical tautology. Now, as a special case of (E1), we have:

(1) $\forall s \forall c \forall P \forall Q[[Say(s,c,P) \,\&\, Say(s,c,Q)] \supset Say(Ess,c,P \equiv Q)]]$

Next, recall that (TDEF2*) yields (T→), the half of (T) acceptable to supervaluationists. A special case of (T→) is:

(2) $\forall s \forall c \forall P \forall Q[Say(Ess,c,P \equiv Q) \supset [True(Ess,c) \supset [P \equiv Q]]]$

From (1) and (2) we have:

(3) $\forall s \forall c \forall P \forall Q[[Say(s,c,P) \,\&\, Say(s,c,Q)] \supset [True(Ess,c) \supset [P \equiv Q]]]$

But (E2) allows us to discharge the condition **True(Ess,c)** from (3) and thereby derive (U). Once we have (U), we can easily recover (T) from (TDEF2*) and (F) from (FDEF2*). The bivalence principles (SB) and (WB) then follow as before from (T) and (F).

Conversely, we can explain (E2) on general grounds by deriving it from (T) and (E1). For an instance of (T) is **Say(Ess,c,P≡P) ⊃ [True(Ess,c) ≡ [P≡P]]**, which yields **Say(Ess,c,P≡P) ⊃ True(Ess,c)**, while an instance of (E1) is **Say(s,c,P) ⊃ Say(Ess,c,P≡P)**; together, these give (E2).

Although (TDEF2*) and (FDEF2*) appear to invite a supervaluationist treatment of vagueness, they do not really do so. Such a treatment would involve the denial of (U). The foregoing argument shows that that in turn would require the supervaluationist to deny (E1). But that is too high a price to pay, for it destroys our conception of what biconditionals say.

There is a similar argument for (U) from (TDEF1), the truth definition with an apparently subvaluationist flavour. The argument also assumes that each relevant sentence s has a negation Ns; this assumption is harmless in the present dialectical situation for a reason just like that already given in relation to the assumption that the relevant sentences are closed under the biconditional function E. In place of (E2), this argument needs:

(E2*) **∀s∀c ~True(NEss,c)**

(E2*) does not need the condition that NEss says something, for if it says nothing it is certainly not true. Subvaluationists should accept (E2*), since no admissible sharpening of NEss is true; it is a classical contradiction. As before, we derive (1) from (E1). We then apply (N) to (1) to reach:

(1*) **∀s∀c∀P∀Q[[Say(s,c,P) & Say(s,c,Q)] ⊃ Say(NEss,c,~[P≡Q])]**

Next, recall that (TDEF1) yields (T←), the half of (T) acceptable to subvaluationists. A special case of (T←) is:

(2*) $\forall s \forall c \forall \mathbf{P} \forall \mathbf{Q}[\mathbf{Say}(\mathbf{NEss},c,\sim[\mathbf{P} \equiv \mathbf{Q}]) \supset [\sim[\mathbf{P} \equiv \mathbf{Q}] \supset \mathbf{True}(\mathbf{NEss},c)]]$

From (1*) and (2*) we have:

(3*) $\forall s \forall c \forall \mathbf{P} \forall \mathbf{Q}[[\mathbf{Say}(s,c,\mathbf{P}) \,\&\, \mathbf{Say}(s,c,\mathbf{Q})] \supset [\sim[\mathbf{P} \equiv \mathbf{Q}] \supset \mathbf{True}(\mathbf{NEss},c)]]$

But (E2*) allows us to apply *modus tollens* to the final conditional in (3*) and thereby derive (U). Once we have (U), we can easily recover (T) from (TDEF1) and (F) from (FDEF1).[15] The bivalence principles (SB) and (WB) then follow as usual from (T) and (F). Although (TDEF1) and (FDEF1) appear to invite a subvaluationist treatment of vagueness, they do not really do so. Such a treatment would involve the denial of (U). The argument shows that that would require the subvaluationist to deny (E1). As before, that is too high a price to pay.

The crucial premise in both arguments for (U) is (E1). It articulates the natural way in which a biconditional sentence says something when its constituent sentences say something. One might come to doubt (E1) by equating saying something with not clearly not saying it. This can be formalized in terms of the ` clearly' operator Δ. Given that the antecedent of (E1) is clearly sufficient for its consequent, we have:

(E1Δ) $\forall s \forall t \forall c \forall \mathbf{P} \forall \mathbf{Q}\ \Delta[[\mathbf{Say}(s,c,\mathbf{P}) \,\&\, \mathbf{Say}(t,c,\mathbf{Q})] \supset \mathbf{Say}(\mathbf{Est},c,\mathbf{P} \equiv \mathbf{Q})]]$

But from (E1Δ) we cannot infer:

(E1$\Delta\Delta$) $\forall s \forall t \forall c \forall \mathbf{P} \forall \mathbf{Q}[[\sim\Delta\sim\mathbf{Say}(s,c,\mathbf{P}) \,\&\, \sim\Delta\sim\mathbf{Say}(t,c,\mathbf{Q})] \supset \sim\Delta\sim\mathbf{Say}(\mathbf{Est},c,\mathbf{P} \equiv \mathbf{Q})]]$

Similarly, when $\Box$ and $\Diamond$ are read as ` it is necessary that' and ` it is possible that' respectively, $\Box[[\mathbf{P} \,\&\, \mathbf{Q}] \supset \mathbf{R}]$ does not generally entail $[\Diamond\mathbf{P} \,\&\, \Diamond\mathbf{Q}] \supset \Diamond\mathbf{R}$. For example, if **P** expresses a contingency

then $\Box$**[[P & ~P]** $\supset$ **[P & ~P]]** is trivially true while **[**$\Diamond$**P &** $\Diamond$**~P]** $\supset$ $\Diamond$**[P & ~P]** is false. For analogous reasons, if it is not clear that s does not say that P and not clear that s does not say that Q, it does not follow that it is not clear that Ess does not say that P if and only if Q. It may be clear that Ess does not say that P if and only if Q just because it is clear that P and not Q. But if it is clear that Ess does not say that P if and only if Q, then that is so because it is clear that s does not *both* say that P *and* say that Q. Thus (E1) and (E1Δ) themselves are unthreatened. Saying something should not be confused with the more complex matter of not clearly not saying it. There is a natural and theoretically central notion of saying which satisfies (E1), (E1Δ) and (U). That is the notion needed in an account of truth and falsity. If we so wish, we can then introduce a secondary notion of quasi-saying by the formula **~**Δ**~Say(s,c,P)**, but quasi-saying something does not amount to saying it. One might for a moment be tempted to suppose that quasi-saying is a more precise notion than saying, because it takes vagueness explicitly into account with its use of the clarity operator. However, the phenomenon of higher-order vagueness implies that quasi-saying is a vague notion too, just like saying. Unlike saying, quasi-saying has no special theoretical significance.

Extension to many-valued logic. How far do our results depend on the underlying framework of classical logic that we have assumed so far? We will argue that principles about truth and falsity such as (T) and (F) are not specific to two-valued logic. There is a plausible case for them in a many-valued context too. Our aim is not to show that they follow from *all* non-classical views (they do not), but to argue that they are not already tantamount to the two-valued conception. To do that, we will prove that (T) and (F) follow from (U) and a closely related principle in a wide class of many-valued logics defined by some natural simplifying assumptions. We suggest that (U) and related principles are themselves non-committal between two-valued and many-valued logic.

According to those who apply many-valued logic to the problem of vagueness, borderline cases have an intermediate status between truth and falsity. On this view, for example, the choice between ` I like him' and ` I do not like him' is sometimes a matter of politeness, not of truthfulness, since the two

statements are equally far from truth and falsity. **~P** will have such a neutral status if and only if **P** also has a neutral status. Many-valued semantics generalizes the two-valued truth-functional account of the logical constants: when a complex sentence is formed by applying a logical constant to simpler sentences, the value of the complex sentence is calculated as a function of the values of the simpler ones. For convenience we will refer to all the values in question as truth-values. In three-valued logic they are truth, falsity and neutrality; in fuzzy logic, they form a continuum of degrees from perfect truth to perfect falsity.

We assume a background notion of a model, commenting only on some distinctive features. Expressions are assigned semantic values in a model relative to assignments of semantic values to variables. Variables in name position are assigned objects. Variables in sentence position are assigned something like propositions, not just truth-values, for intuitively **Say(s,c,P)** and **Say(s,c,Q)** can differ in truth-value relative to the same assignments to **s** and **c** even when **P** and **Q** themselves take the same truth-value. To say one truth is not to say them all. Thus the semantic value of a formula (a proposition) must be distinguished from its truth-value. In a given model, its truth-value is a function of its semantic value but not usually *vice versa*. We assume that there are only finitely many truth-values and that they are linearly ordered by a relation $\leq$. By contrast, fuzzy logic uses a continuum of truth-values, to which our apparatus only approximates. On other views, the relevant ordering is merely partial. Given suitable assumptions, our arguments could be generalized to infinite lattices of truth-values, but for simplicity we will discuss only the finite case here. Since our main concern is with truth-values, by the ` value' of a formula we will mean its truth-value. The maximum and minimum values are t (` truth') and f (` falsity') respectively. At least one but not every one of these values is *designated*. An argument is *valid* if and only if in every model in which every premise takes a designated value, the conclusion also takes a designated value. We assume that the designated values are closed upwards: for any values v_1 and v_2, if $v_1 \leq v_2$ and v_1 is designated then v_2 is also designated. Therefore, since at least one value is designated and t is the maximum value, t is designated, and since at least one value is not designated and f is the minimum value, f is not designated. We say that a formula *holds* if it takes some designated value and

that it *fails* if it takes some undesignated value.

As usual in many-valued semantics, the value of a conjunction is the minimum of the values of its conjuncts, and the value of a disjunction is the maximum of the values of its disjuncts (all relative to an assignment of semantic values to variables). Since the designated values are closed upwards, it follows that a conjunction holds if and only if every conjunct holds, and that a disjunction holds if and only if some disjunct holds. Quantified sentences are treated analogously. The value of **∀x A[x]** (**∃x A[x]**) on an assignment *a* of semantic values to variables is the minimum (maximum) value of **A[x]** on an assignment differing from *a* at most over its assignment to **x** (a variable in either sentence or name position). Consequently, **∀x A[x]** (**∃x A[x]**) holds relative to *a* if and only if **A[x]** holds relative to every (some) assignment differing from *a* at most over **x**. In the particular case of three-valued logic, it follows that the value of **∀x A[x]** is t if the value of **A[x]** is t for every assignment, it is n if the value of **A[x]** is n for some assignment and f for none, and it is f if the value of **A[x]** is f for some assignment. Similarly, the value of **∃x A[x]** is t if the value of **A[x]** for some assignment is t, it is n if the value of **A[x]** is n for some assignment and t for none, and it is f if the value of **A[x]** is f for every assignment.

As is standard, we assume that **A ⊃ B** takes the value t whenever **B** takes a value no lower than the value taken by **A**. In particular, **A ⊃ B** takes the value t in all the following cases: when **A** takes the value f; when **B** takes the value t; when **A** takes the same value as **B**. We also assume that if **A** takes the value t then **A ⊃ B** takes the same value as **B**. The latter equation might be extended to the remaining cases too, but we do not need a ruling for them here. The equivalence ≡ is defined in the natural way as a conjunction of implications; the value of **A ≡ B** is the same as the value of **[A ⊃ B] & [B ⊃ A]**, that is, the minimum of the values of **A ⊃ B** and **B ⊃ A**. Thus **A ≡ B** takes the value t whenever **A** takes the same value as **B**. The most natural assumption about the truth-table for **~** is that it is monotonic decreasing in the sense that the value of **~A** is lower than the value of **~B** if and only if the value of **B** is lower than the value of **A**; in the finite case that assumption determines the table for **~** uniquely.

So far, it may remain tempting to suggest that holding (taking a designated value) is really just being true, and that failing (taking an undesignated value) is really just being false. But suppose that for

some value n, sentences take the value n when and only when their negations do. There is such a ` neutral' value n when and only when the number of values is odd, in particular when it is three (of course, n will not be t or f). Then, whether n is designated or undesignated, it is not the case that **~P** takes a designated value if and only if **P** takes an undesignated value. Thus the proposed equation would falsify the claim that **~P** is true if and only if **P** is false (compare (FN) and Dummett 1981: 421-6 and 1991: 49). If **~** is genuine negation, the system is genuinely non-classical.

The case when **P** and **~P** both take the neutral value n brings out the contrast with supervaluationism most sharply. For then the biconditional **P** ≡ **~P** takes the designated value t, so **~[P** ≡ **~P]** takes the undesignated value f, and is therefore not a theorem. Since **~[P** ≡ **~P]** is a classical tautology, it is valid for both supervaluationists and subvaluationists. They have no analogue of the verification of a conditional by the assignment of the same intermediate value to its antecedent and consequent.

We will make the simplifying assumption that sentences of the form **Say(s,c,P)** take only the ` classical' values t and f. In effect, we idealize away vagueness in the notion of saying. Our reason is not that there is no such vagueness, but that it is not our present concern. The assumption implies nothing about the relation between **P** and **Q** when both **Say(s,c,P)** and **Say(s,c,Q)** take the value t. In particular, it does not follow that **P** and **Q** themselves must take the value t or f. That something vague is said can itself be precisely true or false. Our aim is to show that, even in a non-classical framework within which truth and falsity are assumed not to be jointly exhaustive alternatives for sentences of the object-language, concepts (**True** and **False**) can still be characterized by the principles (T) and (F). As is usual, we will use a classical metalogic in arguing about non-classical many-valued logic. That is in effect to ignore higher-order vagueness. We have no alternative; advocates of many-valued logic have not solved the problem of using it as its own metalogic.

We first show that if **True** is defined by (TDEF1), then the argument from the uniformity principle (U) to (T) is valid. The result of replacing **True(s,c)** in (T) by its *definiens* from (TDEF1) (with a change of bound variables for clarity) is this:

(T*) $\forall s \forall c \forall \mathbf{Q}[\mathbf{Say(s,c,Q)} \supset [\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}] \equiv \mathbf{Q}]]$

Let v_a be the function from formulas to their truth-values in some fixed model relative to an assignment *a*. A **P**-variant of *a* is any assignment differing from *a* at most over **P**. We first establish a lemma:

LEMMA. Either $v_b(\mathbf{Say(s,c,P)}) = f$ for every **P**-variant *b* of *a* or for every formula **A** there is a **P**-variant *d* of *a* such that $v_d(\mathbf{Say(s,c,P)}) = t$ and $v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ A}]) = v_d(\mathbf{A})$.

PROOF: Suppose that for some **P**-variant *b* of *a*, $v_b(\mathbf{Say(s,c,P)}) \neq f$. By the assumption about **Say**, $v_b(\mathbf{Say(s,c,P)}) = t$. By the semantic rule for $\exists$, $v_b(\mathbf{Say(s,c,P)\ \&\ A}) \leq v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ A}])$ and for some **P**-variant *c* of *a*, $v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ A}]) = v_c(\mathbf{Say(s,c,P)\ \&\ A}) = \min\{v_c(\mathbf{Say(s,c,P)}), v_c(\mathbf{A})\}$. There are two cases to consider. If $v_c(\mathbf{Say(s,c,P)}) = t$ then $v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ A}]) = \min\{t, v_c(\mathbf{A})\} = v_c(\mathbf{A})$ and we can put $d = c$. If $v_c(\mathbf{Say(s,c,P)}) = f$ then $v_b(\mathbf{A}) = \min\{t, v_b(\mathbf{A})\} = \min\{v_b(\mathbf{Say(s,c,P)}), v_b(\mathbf{A})\} = v_b(\mathbf{Say(s,c,P)\ \&\ A}) \leq v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ A}]) = \min\{f, v_c(\mathbf{A})\} = f$. Thus $v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ A}]) = f = v_b(\mathbf{A})$ and we can put $d = b$. ■

We can now show that the argument from (U) to (T*) is valid. Suppose that (U) holds relative to *a*. We must show that (T*) also holds relative to *a*. Since (U) is closed and *a* is arbitrary, it suffices to show that $v_a(\mathbf{Say(s,c,Q)} \supset [\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}] \equiv \mathbf{Q}])$ is designated. If $v_a(\mathbf{Say(s,c,Q)}) = f$, then $v_a(\mathbf{Say(s,c,Q)} \supset [\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}] \equiv \mathbf{Q}]) = t$, which is designated. Otherwise, $v_a(\mathbf{Say(s,c,Q)}) = t$ and $v_a(\mathbf{Say(s,c,Q)} \supset [\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}] \equiv \mathbf{Q}]) = v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}] \equiv \mathbf{Q})$. Thus we need only show that $v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}] \equiv \mathbf{Q})$ is designated. Let *b* be the **P**-variant of *a* that assigns to **P** the semantic value that *a* assigns to **Q**. Then $v_b(\mathbf{Say(s,c,P)}) = v_a(\mathbf{Say(s,c,Q)}) = t$. Thus, by the lemma (with $\mathbf{A} = \mathbf{P}$), there is a **P**-variant *d* of *a* such that $v_d(\mathbf{Say(s,c,P)}) = t$ and $v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}]) = v_d(\mathbf{P})$. Consequently, $v_a(\exists \mathbf{P}[\mathbf{Say(s,c,P)\ \&\ P}] \equiv \mathbf{Q}) = v_d(\mathbf{P} \equiv \mathbf{Q})$, so we need only show that $v_d(\mathbf{P} \equiv \mathbf{Q})$ is designated. But since $v_a((U))$ is designated, $v_d(\mathbf{[Say(s,c,P)\ \&\ Say(s,c,Q)]} \supset \mathbf{[P \equiv Q]})$ is also designated. Moreover, *a* and *d* assign the same semantic value to **Q**, so $v_d(\mathbf{Say(s,c,Q)}) = v_a(\mathbf{Say(s,c,Q)}) = t$. Thus

v_d([**Say(s,c,P) & Say(s,c,Q)**]) = min{t,t} = t, so v_d([**Say(s,c,P) & Say(s,c,Q)**] ⊃ [**P** ≡ **Q**]) = v_d(**P** ≡ **Q**). Therefore v_d(**P** ≡ **Q**) is designated. That completes the proof.

Two remarks are worth making on the proof. They will generalize to the arguments about (F) and (SB) mentioned below.

(i) Inspection of the proof reveals something stronger than the validity of the argument from (U) to (T*): relative to *a*, the matrix of (T*) takes either the value t or the value taken by the matrix of (U) relative to some **P**-variant of *a*. It follows that the value of (T*) is at least as high as that of (U). Thus even if we had taken the alternative course of defining an argument to be ` valid' if and only if in every model the value of the conclusion is not lower than the values of all the premises, the argument would still have counted as valid in that sense.

(ii) The proof uses no assumptions at all about the truth-tables for **~** and ≡. The only assumptions that it uses about the truth-table for ⊃ concern the cases when the value of the antecedent is either t or f.

For the case of falsity, we need a slight variation on (U):

(U~) **∀s∀c∀P∀Q[[Say(s,c,P) & Say(s,c,Q)] ⊃ [~P ≡ ~Q]]**

Of course (U~) is equivalent to (U) in classical logic, because **~P** ≡ **~Q** is equivalent to **P** ≡ **Q**. They are also equivalent in some many-valued logics, but not in all. For suppose that **A** ⊃ **B** takes the same value as **B** when that is less than the value of **A**. Let **P** take the value t and **Q** an intermediate value n_1. Thus **~P** takes the value f and **~Q** takes some intermediate value n_2. Consequently, although **Q** ⊃ **P** and **~P** ⊃ **~Q** both take the value t, **P** ⊃ **Q** takes the value n_1 while **~Q** ⊃ **~P** takes the value f. Thus the values of **P** ≡ **Q** and **~P** ≡ **~Q** are n_1 and f respectively. We therefore cannot assume that (U~) is always derivable from (U) in the many-valued context, for n_1 might be designated. However, once we add the assumption (N) (which retains its plausibility in the many-valued context), **Say(s,c,P)** and **Say(s,c,Q)** must have the same values (either t or f, by the hypothesis about saying) as **Say(Ns,c,~P)** and

Say(Ns,c,~Q) respectively, so (U~) will in effect be guaranteed as a special case of (U).

Just as the definition of **True** by (TDEF1) validates the argument from (U) to the principle (T) about truth, so the definition of **False** by (FDEF1) validates the argument from (U~) to the principle (F) about falsity. The result of replacing **False(s,c)** in (F) by its *definiens* from (FDEF1) (with a change of bound variables) is this:

(F*) $\forall s \forall c \forall \mathbf{Q}[\mathbf{Say}(s,c,\mathbf{Q}) \supset [\exists \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \,\&\, \sim\mathbf{P}] \equiv \sim\mathbf{Q}]]$

The argument from (U~) to (F*) can be shown to be valid by an argument just like the one above, with **~P** and **~Q** substituted for **P** and **Q** respectively (except as arguments for **Say**).

Perhaps more surprisingly, a slight variation of the argument shows that if **True** and **False** are defined by (TDEF1) and (FDEF1) respectively, then the argument from (U~) to the strong bivalence principle (SB) is also valid. But does not the many-valued semantics presuppose the failure of bivalence? The point is that many classically equivalent formulations of bivalence are not equivalent to each other in the many-valued context. In the latter, there is no valid argument from (U) or (U~) or both to this variant of weak bivalence:

$\forall s \forall c \forall \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \supset [\mathbf{True}(s,c) \vee \mathbf{False}(s,c)]]$

On expansion by (TDEF1) and (FDEF1), this becomes:

(WB$\vee$) $\forall s \forall c \forall \mathbf{P}[\mathbf{Say}(s,c,\mathbf{P}) \supset [\exists \mathbf{Q}[\mathbf{Say}(s,c,\mathbf{Q}) \,\&\, \mathbf{Q}] \vee \exists \mathbf{Q}[\mathbf{Say}(s,c,\mathbf{Q}) \,\&\, \sim\mathbf{Q}]]$

For suppose that each sentence in a given context expresses at most one proposition. Then (U) and (U~) will both take the value t, but (WB$\vee$) need not take a designated value. For, on the supposition, if **Say(s,c,P)** takes the value t (relative to an assignment), then the disjunctive consequent of (WB$\vee$) will

take the same value as **P** ∨ **~P**, and that value will be at least as high as the value of (WB∨) itself. If **P** takes an intermediate value, then **~P** also takes an intermediate value, and consequently so does **P** ∨ **~P**; in that case, (WB∨) does not take the value t. Thus if t is the only designated value, (WB∨) does not hold. In the corresponding expansion of (SB), the consequent is **~∃Q[Say(s,c,Q) & Q]** ≡ **∃Q[Say(s,c,Q) & ~Q]**, which on the same supposition as before takes the value of **~P** ≡ **~P** (that is, t) when **Say(s,c,P)** takes the value t.

We still need to assess the plausibility of (U) and (U~) in the many-valued context. Most of the same considerations still apply. What is worth remarking is that if t is not the only designated value, then in some systems (U) and (U~) hold even if a given sentence in a given context expresses propositions of different values. Consider, for example, five-valued logic with the values t, n+, n, n- and f in descending value. Suppose that the designated values are t and n+, and that **A** ⊃ **B** takes the value n+ when **B** takes a value one step lower than that of **A**. In that case **~A** will take a value one step lower than that of **~B**, so both **A** ≡ **B** and **~A** ≡ **~B** will take the value n+. Thus in a model in which no sentence expresses two propositions differing in value by more than one step (for each context), both (U) and (U~) take designated values. In consequence, they do not automatically require sentences to express propositions of the same value. The disjunctive weak bivalence principle (WB∨) can still fail, because **P** ∨ **~P** can take the undesignated value n. Of course, we cannot expect ⊃ and ≡ to behave transitively in that logic; the arguments from **A** ⊃ **B** and **B** ⊃ **C** to **A** ⊃ **C** and from **A** ≡ **B** and **B** ≡ **C** to **A** ≡ **C** count as invalid. Some fuzzy logicians have welcomed such invalidities as a way of disarming the sorites paradox. Alternatively, a many-valued logician might maintain transitivity and affirm (U) and (U~) on a stricter reading on which they do require sameness of value. In any case, many-valued logic provides no reason to reject (U) or (U~).

We conclude that in many-valued logic the correctness of (T), (F) and (SB) as principles about truth or falsity does not depend on the disjunctive weak bivalence principle (WB∨). A similar argument could be constructed in terms of other definitions of truth and falsity, such as (TDEF2) and (FDEF2) or (TDEF2*) and (FDEF2*).

There remains a problem about the connection between the metalinguistic concepts of the truth-values t and f and the object-language concepts expressed by **True** and **False** respectively. For suppose that no sentence can express more than one proposition in a single context, and consider an assignment relative to which **Say(s,c,P)** takes the value t but **P** takes the neutral value n. Whether one defines **True** by (TDEF1), (TDEF2) or (TDEF2*), **True(s,c)** will also take the value n relative to that assignment. Thus one adopts a neutral attitude towards **True(s,c)**. But intuitively one is not supposed to adopt a neutral attitude towards the assignment of the value t to **P**, for that assignment is incompatible with the assignment of n to **P**. Thus **True** in the object-language seems not to correspond exactly to the assignment of t in the metalanguage. Yet t was informally explained as truth, and **True** is intended as a formal rendering of that notion. Does this mismatch destabilize the many-valued semantics? We will not attempt to answer that question here. What we have argued is that, on plausible assumptions, the many-valued framework permits one to define notions that do satisfy the principles (T), (F) and (SB), so that the use of principles of that form does not by itself commit one to the classical framework. One of us is independently committed to classical logic; the point here is that that is a *further* commitment. Our definitions of truth and falsity do not necessarily depend on whether we accept classical or non-classical logic.

Notes

* We thank Rastko Jovanović and Peter Milne for very useful comments on a draft of this paper.

1 The second author has criticized accounts of vagueness based on many-valued logic elsewhere (Williamson 1994: 96-141).

2 We use ` necessary' and ` sufficient' without modal force to emphasize the direction of the implications.

3 The present formulation follows Williamson 1995, whereas Williamson 1994: 187-188 treats utterances as truth-bearers. For present purposes nothing turns on the contrast between utterances and sentences in contexts. For the comparison between propositions and linguistic items as truth-bearers in relation to vagueness see the opening parts of Schiffer 1999 and Williamson 1999a.

4 If the sentence s contained modal or tense operators, we might handle their semantics by a further relativization of **Say**, **True**, **False** to the possible world and time with respect to which the sentence is being evaluated (more generally, to a circumstance of evaluation in the sense of Kaplan 1989), which may differ from the possible world and time at which it is being treated as uttered. Since our present concern is not with such operators, we ignore this further kind of relativization.

5 For further discussion see Williamson 1994: 197 and 1998.

6 Dummett (1991: 75-81 and 1995: 203-204) has argued that bivalence should be formulated as the principle that every statement is *determinately* either true or false, claiming that without the

qualification the principle holds in some intuitively non-bivalent quasi-supervaluationist semantic frameworks. But the proper response to that difficulty is to challenge the interpretations assigned by such semantics to **True** and **False** (or to one or more of the logical constants). However many explicit qualifications one builds into a principle, one cannot render it incapable of being misunderstood. The attempt to avoid misunderstanding does not justify complicating the principle itself.

7 Williamson (1994, 1995 and elsewhere) uses ` bivalence' for weak bivalence. The idea of working with strong bivalence was proposed in an earlier version of Andjelković (1999).

8 Andjelković (1999) points out that such a combination is possible; Williamson (1999b) responds. The present paper arose out of reflection on that exchange.

9 See also Dummett 1995: 211.

10 See Dretske 1986 and 1997: 4 on what it takes for a creature to have the capacity to misrepresent its environment.

11 This is an instance of a very general problem for conditional analyses; see Shope 1978.

12 Russell 1923 gives a vivid sense of the primitive level at which vagueness arises, even though he mischaracterizes the phenomenon (for more discussion see Williamson 1994: 52-69). Dummett argues that ` the very concept of the truth of a statement, as distinct from the cruder concept of justifiability, is required only in virtue of the occurrence, as a constituent of more complex sentences, of the sentence by means of which the statement is made' (1993: 193). However, his case for taking the epistemic concept of justifiability as the starting-point for interpretation is inadequately made out.

13 In a similar definition Wright (1992: 31) writes **(∃!P){"P" says that P & P}**, presumably intending by **!** to impose a uniqueness requirement. He stipulates that the quantifier is ` of course' substitutional. Note that, as stated, the requirement seems to be consistent with **(∃P){"P" says that P & ~P}** (contrast **(∃!P){"P" says that P} & (∃P){"P" says that P & P}**). A problem for such a formulation is that the expansion of **∃!P A(P)** as something like **∃P∀Q[A(Q) ≡ P=Q]** is ill-formed, since **P** and **Q** are supposed to be variables in sentence position whereas the arguments of the identity sign **=** take name position. To avoid this problem one might replace **P=Q** by something like **∀Ψ[Ψ(P) ≡ Ψ(Q)]** or **∀s∀c[Say(s,c,P) ≡ Say(s,c,Q)]**. In the present context, a further problem for such a definition is that a vague sentence is supposed to have distinct sharpenings even when the case is *not* borderline, although they would then all have the same truth-value; on that view, the requirement would count vague sentences as not true even in non-borderline cases.

14 (E1) could plausibly be stated with ≡ in place of ⊃, just as (N) is. However, some philosophers might object on the grounds that the order of the flanking sentences in a biconditional is irrelevant. They would accept:

∀s∀t∀c∀P∀Q[Say(Est,c,P≡Q) ⊃ Say(Ets,c,P≡Q)]

But from that principle and the biconditional strengthening of (E1) one can derive the obviously false conclusion:

∀s∀t∀c∀P∀Q[[Say(s,c,P) & Say(t,c,Q)] ⊃ [Say(t,c,P) & Say(s,c,Q]]

No such problem affects the original version of (E1).

15 Conversely, we can derive (E2*) from (T), (T+), (N), (E) and other principles and thereby explain it on general grounds. An instance of (T) is **Say(NEss,c,~[P≡P]) ⊃ [True(NEss,c) ≡ ~[P≡P]]**, which yields **Say(NEss,c,~[P≡P]) ⊃ ~True(NEss,c)**. To discharge the antecedent we need two compositional principles of a different form, to the effect that complex sentences say something only if their constituent sentences do:

∀s∀c∀S[Say(Ns,c,S) ⊃ ∃R Say(s,c,R)]

∀s∀t∀c∀R[Say(Est,c,R) ⊃ ∃P∃Q[Say(s,c,P) & Say(t,c,Q)]]

Instances of those principles yield **Say(NEss,c,S) ⊃ ∃P Say(s,c,P)**. We can combine that with (T+) to derive **True(NEss,c) ⊃ ∃P Say(s,c,P)**. But instances of (N) and (E) yield **Say(s,c,P) ⊃ Say(NEss,c,~[P≡P])**. Putting the pieces together, we have **True(NEss,c) ⊃ ~True(NEss,c)** and therefore **~True(NEss,c)**.

Bibliography

Andjelković, Miroslava. (1999). ` Williamson on bivalence' *Acta Analytica* 14 (issue 23): 27-33.

Aristotle. (1924). *Metaphysics*, trans. W. D. Ross. Oxford: Clarendon Press.

Dretske, Fred. (1986). ` Misrepresentation' , in R. Bogdan, ed., *Belief: Form, Content and Function*. Oxford: Clarendon Press.

Dretske, Fred. (1997). *Naturalizing the Mind*. Cambridge, Mass. and London: MIT Press.

Dummett, Michael. (1981). *Frege: Philosophy of Language*, 2nd ed. London: Duckworth.

Dummett, Michael. (1991). *The Logical Basis of Metaphysics*. London: Duckworth.

Dummett, Michael. (1993). *The Seas of Language*. Oxford: Clarendon Press.

Dummett, Michael. (1995). ` Bivalence and vagueness' *Theoria* 61: 201-216.

Fine, Kit. (1975). ` Vagueness, truth and logic' *Synthese* 30: 265-300. Reprinted in Keefe and Smith (1996).

Hyde, Dominic. (1997). ` From heaps and gaps to heaps of gluts' *Mind* 106: 440-460.

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

Keefe, Rosanna, and Smith, Peter, eds. (1996). *Vagueness: A Reader*. Cambridge, Mass. and London: MIT Press.

McGee, Vann, and McLaughlin, Brian. (1995). ` Distinctions without a difference' *Southern Journal of Philosophy* 33 (supplement): 203-251.

Russell, Bertrand. (1923). ` Vagueness' *Australasian Journal of Philosophy and Psychology* 1: 84-92. Reprinted in Keefe and Smith (1996).

Shope, Robert. (1978). ` The conditional fallacy in contemporary philosophy' *Journal of Philosophy* 75: 397-413.

Schiffer, Stephen. (1999). ` The epistemic theory of vagueness' . In J. Tomberlin, ed., *Philosophical*

Perspectives 13: Epistemology. Oxford and Boston, Mass.: Blackwell.

Tarski, Alfred. (1944). ` The semantic conception of truth and the foundations of semantics' . *Philosophy and Phenomenological Research* 4: 341-375.

Williamson, Timothy. (1994). *Vagueness*. London and New York: Routledge.

Williamson, Timothy. (1995). ` Definiteness and knowability' *Southern Journal of Philosophy* 33 (supplement): 171-191.

Williamson, Timothy. (1997). ` Imagination, stipulation and vagueness' . In E. Villanueva, ed., *Philosophical Issues 8: Truth*. Atascadero CA: Ridgeview.

Williamson, Timothy. (1998). ` Indefinite extensibility' *Grazer Philosophische Studien* 55: 1-24.

Williamson, Timothy. (1999a). ` Schiffer on the epistemic theory of vagueness' . In J. Tomberlin, ed., *Philosophical Perspectives 13: Epistemology*. Oxford and Boston, Mass.: Blackwell.

Williamson, Timothy. (1999b). ` Andjelkovi on bivalence: a reply' *Acta Analytica* 14 (issue 23): 35-8.

Wright, Crispin. (1992). *Truth and Objectivity*. Cambridge, Mass. and London: Harvard University Press.