Truth and What Is Said[*]

Elia Zardini

Arché, Philosophical Research Centre for Logic, Language, Metaphysics and Epistemology

Department of Logic and Metaphysics

University of St Andrews

July 12, 2008

Abstract

A notion of *truth* as applicable to events of assertoric use (*utterances*) of a sentence token is arguably presupposed and required by our evaluative practices of the use of language. The truth of an utterance seems clearly to depend on what the utterance *says*. This fundamental dependence seems in turn to be captured by the schema that, if an utterance u says that P, then u is true iff P. Such a schema may thus be thought to constitute a suitable basis for an adequate theory of utterance truth, so much so that it seems straightforwardly to avoid the problems arising from *context dependence* and the *semantic paradoxes* which notoriously beset theories of utterance truth based on a

simple *disquotational* schema. The paper argues that appearances are deceptive in both cases. On the one hand, the schema cannot allow for plausible if not uncontroversial *non-indexical* forms of context dependence, arising from the possibility that what an utterance says can be the case or not relative to different situations and that the truth of an utterance u of a sentence φ arguably depends on the truth of φ at the situation "associated" with u. On the other hand, a *quantified utterance-truth* variation on the liar paradox shows that the schema entails some consequence φ and at the same time the untruth of any utterance of φ; moreover, a resilient *quantified propositional* variation on the contingent liar paradox is offered, which only relies on resources usually employed by theories of utterance truth based on the schema.

1 Introduction and Overview

Sometimes truth discriminates between different *tokens* of the same sentence *type*,[1] as, for example, between a token of the sentence 'I'm a philosopher' produced by me and a token of the very same sentence produced by Victoria Beckham. In such cases, we need a more fine-grained notion than that of sentential truth, a notion which can differently apply to different tokens of the same sentence. But sometimes truth discriminates even between different *events of assertoric use*[2] of the same sentence token, as for example, between an event of use of a token inscription of the sentence 'This site is dangerous' in a construction area and an event of use of the very same token inscription on a playground (cf Percival [1994], pp. 203–205).[3] In such cases, we need a more fine-grained notion than that of sentence-token truth, a notion which can differently apply to different utterances of the same sentence token.

At a first glance, the notion of utterance truth (as well as that of sentence-token truth), uncritically employed in the preceding examples, may well seem, at best, rather awkward and stretched, if not a straightforward category mistake. For it may well seem that we don't usually attribute the property of being true to spatiotemporally located objects in general, let alone human actions (of course, in the relevant sense—not in the sense operative in phrases like 'a true repentance'!). But simple grammatical or lexical modifications already suffice to cast a totally different light on the situation: granted, we hardly ever apply the adjective 'true' to speech acts, but our 'truth'-talk abounds with other adjectival, adverbial and nominal locutions such as 'I want a truthful answer', 'He spoke truly', 'I will tell the truth' etc. and with adjectives such as 'correct', 'right', 'accurate' etc., all of which (on at least some of their readings) do seem to track a particular property utterances have

just in case they are, very roughly, *objectively correct* (only "objectively" correct, for it may well be that other, non-truth-related features—e.g. epistemic ones—are relevant for the assessment of the overall correctness of an utterance). Arguably, such a property bears enough similarity to features of the notion(s) of truth as applied, say, to sentences or propositions as to deserve to be designated by the same word. In any case, it apparently plays such an important role in our evaluative practices of the use of language as to deserve specific philosophical consideration—if you don't want to call it 'utterance truth', call it 'utterance (objective) correctness'. This is what will occupy us in the following. We will remain neutral with regard to the questions as to whether there are other *truth bearers* (sentences, propositions, intentional mental states) and what conceptual priorities there are (if any) among the different notions of truth applying to these different kinds of truth bearers.

The rest of the paper is organized as follows. Section 2 presents two prominent theories of utterance truth and their respective motivations. Section 3 discusses the import of various context-dependence phenomena on the viability of these theories, arguing that neither of them is likely to hold in full generality. Section 4 focuses on a range of strengthened semantic paradoxes, showing how, in a classical framework and under some additional highly plausible assumptions, both theories are inconsistent. Section 5 draws some conclusions and indicates further directions of research suggested by them.

2   Two Theories of Utterance Truth

2.1   Truth and What Is Said

It is no more than a platitude that the truth of an utterance depends on what the utterance *says*.[4] For example, it is no more than a platitude that an utterance of the English sentence 'Snow is white' is true partly in virtue of its saying that snow is white. (Of course, it is also true partly in virtue of snow's being white. In general, and famously, an utterance's truth depends both on what the utterance says and on how the world is. I leave open here whether there is any interesting sense in which, sometimes, an utterance is true *wholly* in virtue of what it says (and so, in a good sense, *analytically* true).) Nor does it seem to be a mystery why truth and what is said are so connected. Intuitively, an utterance's relation to what it says does not float in the void: by saying that P,[5] an utterance u relates to a certain situation s as well and *represents s as* being such that P—in other words, u says that P *of s*. By representing s as being such that P, u comes to possess a

particular *standard of correctness* such that, by this standard, u is correct iff s is how u represents it to be (namely, such that P); to say that u is true is just to say that u meets this particular standard. To continue with the previous example, the relation of your utterance in the actual world of the English sentence 'Snow is white' to what it says does not float in the void: by saying that snow is white, your utterance typically relates to the actual world as well and *represents it as* being such that snow is white—in other words, your utterance says that snow is white *of the actual world*. By representing the actual world as being such that snow is white, your utterance comes to possess a particular *standard of correctness* such that, by this standard, your utterance is correct iff the the actual world is how your utterance represents it to be (namely, such that snow is white); to say that your utterance is true is just to say that your utterance meets this particular standard.

Of course, much more would need to be said in order to substantiate the foregoing considerations. Firstly, an account is needed of what it takes for an utterance to be about one situation rather than another. Such an account would fall into the domain of *pragmatics* and focus on how speakers' mental states and salient features of their environment conspire to determine a specific situation as the target of a certain occasion of speech. I will have something to say in this respect in section 3.4, mainly to stave off a common misconception that unduly limits the range of situations an utterance can be about. Secondly, an account is needed of which situations there are (at least as far as the needs of the semantic interpretation of natural languages are concerned). Such an account would fall into the domain of *semantics* and clarify whether the semantic interpretation of natural languages needs to posit entities relative to which what an utterance says is the case, and, if so, how fine-grainedly individuated they must be: can they be identified with possible worlds or must they be allowed to differ from one another in other respects as well, such as times, standards, places etc.? I will have something to say in this respect in section 3.3, arguing that the possibility of a quite fine-grained individuation of situations must be taken very seriously, and that, in any case, it is very likely that some dimension or other will have to be countenanced relative to which what an utterance says is the case. Thirdly, an account is also needed of what it takes for an utterance to say what it says. Such an account would fall into the domain of the *theory of linguistic content* and illuminate its psychological, social and environmental grounds. I will have nothing to say here on this major issue in the philosophy of language.

Incomplete as they may be, the foregoing considerations do enshrine a certain conception and may be thought to lead to the following schematic

4

what-is-said theory of utterance truth:

(WIS) For every utterance u, if[6] u says that P, then [u is true iff P].[7]

It must be stressed that 'u says that P' in (WIS) is understood to mean, somewhat strictly, what 'u exactly says that P' does—that is, to be true only if, for no X, u says that X and its being the case that X entails its being the case that P but not *vice versa* (cf Andjelković and Williamson [2000], pp. 230–232). A broader understanding of 'u says that P' would for instance be the one according to which, if an utterance says that $X_0, X_1, X_2 \ldots$ and its being the case that $X_0, X_1, X_2 \ldots$ entails its being the case that X_i, then the utterance also says that X_i—that is, an understanding according to which saying is *closed under logical consequence*.[8] Under such a broad understanding, (WIS) would be blatantly inconsistent. For example, your utterance of the English sentence 'Snow is black' would also say that [either snow is black or snow is white], since it says that snow is black, and its being the case that snow is black entails its being the case that [either snow is black or snow is white]. Your utterance would thus be true (using the instance of (WIS) obtained by substituting the English sentence 'Either snow is black or snow is white' for 'P') as well as untrue (using the instance of (WIS) obtained by substituting the English sentence 'Snow is black' for 'P').

A similar problem arises if an utterance is produced to say more than one thing. For example, you might be discussing face-to-face with your partner whether to go to the river or to the mountains this afternoon while also discussing at the phone with your financial consultant whether to go to the bank or to the stock exchange this morning. You might utter the English sentence 'I love going to the bank' with the intention of saying something both to your partner and to your financial consultant. In such a case, your utterance of 'I love going to the bank' would be produced to say both (correctly, let us suppose) that you love going to the river bank and (incorrectly, let us suppose) that you love going to the money bank. It would thus be true (using the instance of (WIS) obtained by substituting 'You love going to the river bank' for 'P') as well as untrue (using the instance of (WIS) obtained by substituting 'You love going to the money bank' for 'P').

For simplicity's sake, in the following we will screen off both broader understandings of 'u says that P' and the existence of such "double-talk" cases. One way of accommodating the latter—which would allow one to preserve (WIS) as it stands—would be to exploit the fine-grainedness of individuation afforded by our talk of "events of use" and to maintain quite generally

that, in any *i*ple-talk case, actually i many (possibly spatiotemporally colocated) events of assertoric use (i.e. utterances) take place. Alternatively, if one wished to cope with both problems in one fell swoop, a good start would be constituted by a natural quantificational variation on (WIS):

(WIS$^\forall$) For every utterance u, u is true iff, [[for some X, u says that X] and, [for every X, if u says that X, then X]].

The arguments to follow, suitably modified, will go through for (WIS$^\forall$) just as well as for (WIS).

2.2 Truth and Disquotation

The considerations supporting (WIS) differ remarkably from those concerning the *sui generis equivalence* which sometimes seems to obtain between 'P' and 'An utterance of 'P' is true' (on which we can afford to be briefer given the extensive attention they have received in the recent truth debate). Sometimes, for example, to say that an utterance u of the English sentence 'Snow is white' is true seems to amount to no more (and, to be sure, no less) than simply saying that snow is white.[9] On such occasions, in our talking about u's being true, our glance is no more directed towards a complex linguistic object and its (and its parts') multifarious relations to snow and being white than it is in our talking about snow's being white; *modulo* the existence of u, what it takes for u to be true just is what it takes for snow to be white. In this sense, u is true quite independently of the fact that we use the words 'snow', 'is' and 'white' for saying what we say when we utter the English sentence 'Snow is white', just as snow's being white is so independent. This second kind of consideration may in turn be thought to lead to the following schematic *disquotational theory of utterance truth*:

(D) For every utterance u, if u is an utterance of 'P', then [u is true iff P].

Note that both (WIS) and (D) are examples of what we may call '*correlation principles*': principles connecting the truth of an utterance of 'P' with its being the case that P. The validity of a correlation principle of some form or other would seem central to many different conceptions of truth, for instance both to those belonging to the correspondence tradition and to those belonging to the deflationist tradition. It is thus worthwhile investigating the extent to which the validity of these two prominent correlation principles can be plausibly upheld.

In effect, it is by no means implausible to think that the considerations supporting the correlation principle (WIS) are *more fundamental* than those supporting the correlation principle (D), and hence that (WIS)—as against (D)—should constitute the basis of an adequate theory of utterance truth.[10] Indeed, not only does (WIS) seem to be supported by more fundamental considerations than (D) is, but some theorists (see e.g. Williamson [1998], pp. 11–15) also think that, whilst retaining all the advantages offered by (D) to a theory of utterance truth, (WIS) avoids some of its major drawbacks, which concern *context dependence* and the *semantic paradoxes*. In the rest of this paper, I shall argue that this assessment of the dialectical situation is substantially incorrect: a (WIS)-based theory of utterance truth fails to handle some *non-indexical* context-dependence phenomena we might wish to recognize in natural languages, and variously fails to handle some *strengthened* semantic paradoxes.

3   Context Dependence

3.1   Correlation Principles and Indexicality

A prominent and much studied source of context dependence in natural languages is *indexicality*. Fortunately, we don't need to settle on an exact definition of this semantic property; for our purposes, the rough characterization of indexicality will do as the property which a sentence exemplifies just in case different utterances of it can *refer to different things* (objects, properties, relations etc.), even if the linguistic meaning of the sentence is kept fixed across such uses (more on indexicality and its relations to other sources of context dependence in section 3.2).[11]

The inadequacy of (D) in handling even the most simple cases of indexicality-related context dependence is glaring. Consider an utterance u_0 of the English sentence 'I'm British' produced by Tony Blair. u_0 is true, since 'I' as it appears[12] in it refers to Blair and Blair is indeed British. Consider next my utterance u_1 of the instance of (D) 'If u_0 is an utterance of 'I'm British', then [u_0 is true iff I'm British]'. As a matter of fact, u_1 has a true antecedent and a false consequent (the consequent being false since its left-hand side is true and its right-hand side false). The source of the problem here is clearly that the referent of 'I' varies from Blair's utterance of 'I'm British' to my subutterance[13] of the same sentence (as it constitutes the right-hand side of the consequent u_1) in such a way as to make the former true and the latter false (thereby making u_1 false).

At this point, the advocate of (WIS) will hasten to point out that the previous counterexample is no threat to (WIS), since it is not the case that u_0 says that I'm British, and so I cannot detach the (false) consequent of my utterance of the instance of (WIS) 'If u_0 says that I'm British, then $[u_0$ is true iff I'm British]'. True, u_0 does say that Blair is British, and so I can detach the consequent of my utterance of the instance of (WIS) 'If u_0 says that Blair is British, then $[u_0$ is true iff Blair is British]', but both sides of the detached biconditional are true. Hence, no immediate problem from indexicality would seem to arise for (WIS).

Of course, it is one thing to show that the most natural recipe for a counterexample from indexicality does not work against (WIS), another thing to show that no such counterexample exists. To this effect, the advocate of (WIS) may wish to expand on the previous consideration and point out that the *provisional* formulation of (WIS) offers a presumably complete control over indexicality-related context dependence, requiring, as it does, that its being the case that P be what an utterance says in order for it to be the necessary and sufficient condition for the utterance's truth. For a purported counterexample to (WIS) from indexicality would presumably exhibit an indexicality-induced variation from the utterance u_0 mentioned on the left-hand side to the subutterance s_0 constituting the right-hand side of the embedded biconditional of the target utterance u_1 of (an instance of) (WIS); but a necessary condition on any indexicality-induced variation is that any such variation imply a variation in the things u_0 and s_0 *refer to*, which presumably implies a variation in what u_0 and s_0 *say*. This in turn falsifies the antecedent of u_1. To see this, reflect that the subutterance s_1 which is involved in u_1's antecedent and which specifies what u_0 is supposed to say is a subutterance of the same sentence as s_0 in the same context as s_0, and hence presumably says what s_0 says; since presumably this is not what u_0 says, u_1's antecedent is false. u_1's antecedent being false, u_1 itself is true—or so (a sketch of) an argument for a safety result concerning (WIS) would go.[14] Were this or a similar argument to go through, indexicality-related context dependence would quite generally be shown to pose no threat to (WIS).[15] I will henceforth assume that some such argument is successful, and turn to investigate how (WIS) fares with respect to another possible source of context dependence.

3.2 Varieties of Context Dependence

Not every form of context dependence need be reducible to indexicality. As I will be using the phrase, 'context dependence' is meant to capture the

very general and abstract property which a sentence exemplifies just in case different utterances of it *can have different truth values.* Consider for example an utterance u_0 of the English sentence 'Riga is the capital of Latvia in 2005 AD' produced in the actual world and an utterance u_1 of the same English sentence produced in a world where Daugapils instead of Riga is the capital of Latvia in 2005 AD. Intuitively, there is no indexicality whatsoever in the English sentence 'Riga is the capital of Latvia in 2005 AD': both u_0 and u_1 refer to Riga, Latvia, 2005 AD and (we may assume) the function of being-the-capital-of-x-in-y. Indeed, intuitively, both u_0 and u_1 simply say that Riga is the capital of Latvia in 2005 AD. However, intuitively, u_0 is true and u_1 is false (in the worlds in which we are considering them), and so the English sentence 'Riga is the capital of Latvia in 2005 AD' exhibits a form of context dependence which is not reducible to indexicality.

Indeed, there seems to exist a whole hierarchy of distinct specific phenomena, all of which variously exemplify the abstract property of context dependence:

Conventionality The same sentence can be uttered in two different *languages*: (we can assume that) a (vocal) utterance of 'Empedocles leaped' in English is true, while a (vocal) utterance of the same sentence in German is false.

Ambiguity Even keeping the language fixed, two utterances of the same sentence can have different *meanings*: an utterance of 'Barclays is a bank' meaning that Barclays is a money bank is true, while an utterance of the same sentence meaning that Barclays is a river bank is false.

Indexicality Even keeping the linguistic meaning fixed, two utterances of the same sentence can refer to different *things*: an utterance of 'I'm British' referring to Blair is true, while an utterance of the same sentence referring to me is false.

Locality Even keeping fixed what is referred to, two utterances of the same sentence can say the same thing about two different *situations*: an utterance of the sentence 'Riga is the capital of Latvia in 2005 AD' talking about the actual world is true, while an utterance of the same sentence talking about a world where Daugapils is the capital of Latvia in 2005 AD is false.

A possible gap in the taxonomy should be apparent. Is it not possible that two utterances of the same sentence refer to the same things (so that

indexicality effects are ruled out) and say what they say about the same situation (so that locality effects are ruled out), but are nevertheless such that one is true and the other is false? Consider for example the case of Pia (adapted from Travis [1997], p. 89), who has just painted green the originally russet leaves of her Japanese maple. Pia's neighbor, a photographer, asks her for a green subject and Pia, pointing at the maple, utters the English sentence 'The leaves are green'. A friend of Pia's, a botanist, asks her for green leaves on which to conduct a biological experiment and Pia, pointing again at the maple, utters once more 'The leaves are green'.

It might seem natural to describe the example as one in which both of Pia's utterances refer to the leaves and to the property of being green (setting aside irrelevant issues of tense) and say what they say about the same situation (a situation in which Pia's painted maple prominently features), and yet as one in which the former utterance is true and the latter one is false, since what it takes for the former to be true is, roughly, that the maple's leaves look green under normal conditions, whereas what it takes for the latter to be true is, roughly, that the maple's leaves be green as their natural color. (It might even seem natural to describe the example as one in which, owing to their different truth conditions, the two utterances say different things, even though this further claim is not essential for the example to preserve much of its thrust. If this further claim is correct, the source of context dependence in question would have to be located in the hierarchy just sketched below indexicality—since what is referred to is kept fixed—but above locality—since what is said varies.) If this description of the situation is on the right track, then there can be context dependence even in the absence of indexicality and locality (a view vigorously defended e.g. in Recanati [2004]). Be that as it may, having clarified the space of available options here, it will do no harm for our main purposes to ignore this possibility as well as to assume that the identity of what an utterance says is determined as soon as the indexical expressions that appear in it have (contextually) been assigned a referent, so that difference in what is said requires difference in some such assignment.[16]

3.3 The Variation of a Proposition's Truth Value

We can henceforth set aside conventionality and ambiguity and focus on the contrast between indexicality and locality, recasting the main point of the previous considerations as follows. *One* way for different utterances of the same sentence to differ in truth value is for them to happen to say different things, in virtue of their being produced in different contexts—indexicality,

as in the previous example of different utterances of the English sentence 'I'm British', one produced by Blair and one produced by me. But *another* way for them so to differ *even while saying the same thing* would be for them to happen to talk about different situations, again in virtue of their being produced in different contexts—locality, as in the previous example of different utterances of the English sentence 'Riga is the capital of Latvia in 2005 AD', one produced in the actual world and one in a world where Daugapils is the capital of Latvia in 2005 AD. Thus, a link seems to exist between the *context dependence* of a sentence (whether indexicality-induced or locality-induced) and its *different contexts of utterance*. But how should this link be more precisely understood?

In formal semantics, for expressively rich enough languages, we are used to define truth of a sentence relative to a sequence of what we may call '*indexical parameters*' (agent, audience, time etc.). The values of these parameters are what is relevant for the interpretation of indexical expressions ('I', 'you', 'now' etc.) occurring in the sentence considered as uttered in a particular context. Under our current assumptions (section 3.2), what is so determined at this level of semantic interpretation are the contents of our utterances—what is said.

However, many definitions of sentential truth are not only relative to a sequence of indexical parameters, but also relative to a further sequence of what we may call '*evaluational parameters*' (world, time, standards etc.; cf the fundamental distinction between *contexts* and *circumstances of evaluation* in Kaplan [1989]). This is so because what is determined at the level of the interpretation of indexical expressions featuring in an utterance of a sentence—what is said—is also assumed to be something capable of being the case or not only relative to a particular situation; the further relativization of sentential truth to a sequence of evaluational parameters is exactly intended to represent this fact. For ease of discussion, we can henceforth use the first-order locution ⌜The proposition expressed by τ is true⌝[17] and its like as a handy short for the second-order locution ⌜What is said by τ is the case⌝[18] and its like. We can then observe that a variety of dimensions have been seriously considered as possible dimensions of variation of a proposition's truth value (and thus, at the level of formal semantic representation, as evaluational parameters to which sentential truth is relative). Prominent in the literature are:

Worlds Rather uncontroversially, a proposition can vary its truth value across different worlds, as the proposition that Riga is the capital of Latvia in 2005 AD does.

Times More controversially, a proposition might vary its truth value across different times, as the proposition that Socrates is young (if there is such a thing) might do.

Standards Even more controversially, a proposition might vary its truth value across different standards, as the proposition that Steve is rich (if there is such a thing) might do.

Places Equally controversially, a proposition might vary its truth value across different places, as the proposition that it is raining on 14/04/2006 (if there is such a thing) might do.

I think it's fair to say that the admission of any of these (or other) dimensions as dimensions of variation of a proposition's truth value is a matter of controversy in contemporary philosophy of language on which the jury is still out (contrast this with indexicality, which is uncontroversially admitted to be exhibited by e.g. 'I', 'you', 'now' etc.), even though variation with respect to worlds is still the majority view among philosophers and variation with respect to times is not uncommonly accepted. Since the problem for (WIS) to which I'm building up in this part of the paper relies on the variation of a proposition's truth value with respect to some dimension or other, I think it's important briefly to lay out what seem to me to be the most important arguments in favor of such variation. This is of course not the place properly to evaluate any of them: my main aim in this part of the paper is to establish the conditional conclusion that, if such variation exists, then (WIS) must be restricted, and, by presenting some of the arguments in favor of the variation, give a hint of the serious philosophical challenges which must be faced if the denial of the antecedent of this conditional is to be upheld on behalf of (WIS) (in the following, my talk of "counterexamples" and the like should be understood in a corresponding conditionalized fashion).

What reasons can we have to admit a particular dimension as a dimension of variation of a proposition's truth value? Four main quite distinct lines of argument suggest themselves.[19] I will list them in order of what I think is their increasing strength, oftentimes exemplifying them with respect to a specific dimension of variation (as the reader will see, at least some of them possess different degrees of plausibility in relation to different dimensions of variation).

Firstly, a particular dimension might be acknowledged as a dimension of variation of a proposition's truth value on the grounds that this would most naturally explain a wide pattern of *speech behavior*, involving *attitude reports* and *truth attributions*. For example, suppose that, in Italy, I utter the English

sentence 'It's noon' and that, in England, Nancy utters the same sentence at the same time. In such a situation, all the following attitude reports would seem felicitous: 'Elia says that it's noon, and Nancy believes it', 'It's noon, and that's what's said by both Elia and Nancy', 'Elia and Nancy make the same statement'. The felicity of such reports is most naturally explained by postulating that what Elia and Nancy say or state is a place-unspecific proposition, which, keeping the time fixed, is true at some places (say, Italy) and false at others (say, England). The same postulation would also most naturally explain why, in the same situation, one can felicitously make the truth attribution: 'What Elia says is true in Italy, but not in England' (see Kaplan [1989], p. 504 for an early mention of this kind of argument and Cappelen and Hawthorne [2009] for a critical discussion of some of these data).

Secondly, a particular dimension might be acknowledged as a dimension of variation of a proposition's truth value on the grounds that this would seem to be required by the presence in the language of the corresponding *intensional operators* (usual candidates are, for example, 'it is necessary that' for worlds, 'always' for times, 'loosely speaking' for standards etc.), which, very roughly, form a sentence φ_1 from a sentence φ_0 in such a way that the truth value of φ_1 is a function of the distribution of the truth values of φ_0 *over a certain range of models*. These models may then be taken to represent the relevant range of worlds, or times, or standards etc. and the varying distribution of truth values along them may be taken to represent the variation of a proposition's truth value along these dimensions (see Lewis [1980], pp. 36–40; Kaplan [1989], pp. 502–504 for influential statements of the argument and Cappelen and Hawthorne [2009] for a critical discussion).

Thirdly, a particular dimension might be acknowledged as a dimension of variation of a proposition's truth value on the grounds that propositions whose truth values are sensitive to the dimension in question would seem to be the best candidates for being the content of certain *egocentric attitudes*. For example, John, a hungry and unexperienced mushroom hunter being faced with the choice of either eating an amanita or throwing it away, might know full well, for every world w, whether amanitas are poisonous in w, and still keep holding the amanita in his hands puzzled as to what to do. When, on being informed by his fellow hunters about the relevant facts, he suddenly throws the amanita away, it is very intuitive to describe these circumstances as ones in which John has acquired a new belief which contributes to explaining his actions—namely, the belief that amanitas are poisonous. However, were the content of this belief something which is either necessarily true or necessarily false (i.e. true at every or no world), it would be hard to see how

the belief could have been so acquired and have such effects—after all, a necessarily true or false content would be something to the effect that amanitas are poisonous in a certain world, and John already believed or disbelieved all the contents of this kind. Together with the plausible assumption that the contents of our believings are also contents of our sayings, the desired result of variation of a proposition's truth value across worlds would then follow (see Lewis [1979] for a specific development of the argument and Recanati [2007], pp. 239–266 for a critical discussion).

Fourthly, a particular dimension might be acknowledged as a dimension of variation of a proposition's truth value on the grounds that doing so would seem to be required by a certain for many appealing metaphysics of the discourse in question, or, at least, by the almost equally appealing desideratum that natural languages usually latch onto the central features of such metaphysics. For example, if the proposition expressed by an utterance at time t of the English sentence 'Socrates is sitting' were either eternally true or eternally false, that utterance would presumably express something that for all intents and purposes can be regarded as equivalent with the proposition that Socrates is sitting at t. That property of Socrates—to be sitting at t—is either eternally possessed by him or eternally not possessed by him; it is not a property that he can acquire or lose. Indeed, given the arbitrariness of the example, arguably no utterance could express a proposition attributing a property that can be acquired or lost. (I'm availing myself to 'property'- and 'fact'-talk. As in many other cases, this is not essential to the point being made, but it greatly facilitates its verbal presentation.) But facts consisting in the possession or lack of properties that can be acquired or lost would seem to be a central feature of what is for many an appealing metaphysics of time and change (see Dummett [1997], pp. 53–54 for similar ideas).

In what follows, I wish to remain neutral as to which, if any, specific dimensions of variation should be accepted (on the basis of the previous or other lines of argument). All I want to retain from the previous discussion is the modest conclusion that at least some such dimension is at least likely to be recognized by our best semantic theories of natural languages—as will be seen, the force of the counterexamples to (WIS) is likely not to be ultimately affected by the specific identities of such dimensions.

3.4   Utterance Truth and Locality

The foregoing by itself still leaves open how truth *of an utterance* connects with the relativized definition of *sentential* truth given in the formal seman-

tics. Possibly barring some benign indeterminacy, it is possible to associate each utterance with a unique sequence of indexical parameters, representing the utterance's operative context. Assuming, as usual, that a sequence of indexical parameters (properly or improperly) contains a sequence of evaluational parameters, the associated sequence of indexical parameters will determine a corresponding sequence of evaluational parameters, representing in turn the situation the utterance talks about. Then, exploiting the definition of sentential truth relative to sequences of indexical and evaluational parameters supplied by the formal semantics and setting aside issues (which need not concern us here) about the possible gap between model-relative and absolute semantic notions, it is very natural to assume the following principle *connecting utterance truth with relativized sentential truth*:

(UTRST) For every utterance u, u is true iff the sentence u is an utterance of is true (in the "intended model" of the language) relative to the sequence of indexical parameters associated with u and to the sequence of evaluational parameters thereby determined.[20]

Note that (UTRST) provides in effect a necessary and sufficient condition for utterance truth. It is not however a principle in the same ballpark as (WIS) and (D), since it is not a correlation principle.

Unfortunately, given (UTRST) and given the assumption, which I have discussed in section 3.3, of the existence of dimensions of variation of a proposition's truth value (which is reflected, at the level of formal semantic representation, in the fact that the relativization of sentential truth to evaluational parameters is non-vacuous), counterexamples to (WIS) are forthcoming. For instance, assuming standards to be one of the evaluational parameters, if you and I happen to have different standards for richness, so that Steve meets yours but fails to meet mine, your utterance u_0 of the English sentence 'Steve is rich' still says that Steve is rich (just as, even in a world where Daugapils is the capital of Latvia in 2005 AD, an utterance of the sentence 'Daugapils is the capital of Latvia in 2005 AD' produced by an English speaker still says that Daugapils is the capital of Latvia in 2005 AD). However, my subutterance of the English sentence 'u_0 is true' is true (since the English sentence 'Steve is rich' is true relative to your standards, and (UTRST) holds no matter what the standards are) and my subutterance of the English sentence 'Steve is rich' is false (since that sentence is false relative to my standards), thereby making false my utterance u_1 of the instance of (WIS) 'If u_0 says that Steve is rich, then [u_0 is true iff Steve is rich]' (since u_1 has a true antecedent and a false consequent, consisting in a true left-hand side and in a false right-hand side).[21]

I want to emphasize that, in the proposed counterexample, standards have been chosen as the relevant dimension of variation only for ease of exposition. With only a little more ingenuity, analogous counterexamples could have been given using e.g. times instead. Suppose that, at the last second of one of Italy's matches, Stephen Hawking started to utter the English sentence 'Italy is playing well' and that, as it happens, the complex device he uses to communicate his thoughts takes a couple of seconds to finish to perform the utterance by completing the release of the relevant soundwaves. Suppose also that I'm listening to Hawking's utterance, and so receive the relevant soundwaves no sooner than at the first second after the end of the match, when Italy is not playing at all and *a fortiori* is not playing well. Suppose finally that it is common knowledge between me and Hawking that his utterances suffer from this inconvenient stretching along the temporal dimension beyond the specific time they talk about. Then Hawking's utterance u_0 of the English sentence 'Italy is playing well' still says that Italy is playing well. However, my subutterance of the English sentence 'u_0 is true' is true (since the English sentence 'Italy is playing well' is true relative to the specific time Hawking's speech is about, and (UTRST) holds no matter what the time is) and my subutterance of the English sentence 'Italy is playing well' is false (since that sentence is false relative to the time of my speech), thereby making false my utterance u_1 of the instance of (WIS) 'If u_0 says that Italy is playing well, then [u_0 is true iff Italy is playing well]' (since u_1 has a true antecedent and a false consequent, consisting in a true left-hand side and in a false right-hand side).

It is however worth stressing that the structure of the counterexamples might not seem to be straightforwardly extendible to worlds, which are after all the only relatively uncontroversial case of dimension of variation of a proposition's truth value. For, in order to have such counterexamples, we might seem to need to be able to attribute truth or falsity *simpliciter* to an utterance produced in a world relevantly different from the actual world ("truth or falsity *simpliciter*": not just truth or falsity *at that world*), just as, in order to obtain the counterexample from standards, we have attributed truth *simpliciter* to an utterance produced by a speaker with relevantly different standards for richness. For consider an utterance u of the English sentence 'Daugapils is the capital of Latvia in 2005 AD' produced in a world w where Daugapils is the capital of Latvia in 2005 AD. Even if we assume a necessitated version of (WIS), this only gives us (in a normal modal logic) that, [if, [at w, u says that Daugapils is the capital of Latvia in 2005 AD], then, [at w, u is true iff Daugapils is the capital of Latvia in 2005 AD]], from which we can get, by discharging the antecedent, that, [at w, u is true iff

Daugapils is the capital of Latvia in 2005 AD]. This in turn entails (again, in a normal modal logic) that, [[at w, u is true] iff [at w, Daugapils is the capital of Latvia in 2005 AD]], from which we can get, by discharging the left-hand side, that, [at w, Daugapils is the capital of Latvia in 2005 AD]. This is of course not a problematic consequence of (the necessitated version of) (WIS), since, even though it is not the case that Daugapils is the capital of Latvia in 2005 AD, it is indeed the case that, at w, Daugapils is the capital of Latvia in 2005 AD.

Attribution of truth or falsity *simpliciter* to an utterance produced in a world relevantly different from the actual world can be made sense of granted certain not uncontroversial metaphysical assumptions, such as the existence of very fragile, world-bound *possibilia*, for which even *trans-world* attribution of properties *simpliciter* would presumably make sense. Such an assumption is not unreasonable especially for the ontological category of events (to which utterances belong)—in fact, it is automatically correct if utterances are identified in the relevant respects with the occurrences mentioned in note 1. Indeed, it is not only philosophical, but also ordinary patterns of speech behavior that evidence the naturalness of such an assumption even for the ontological category of objects: 'In one scenario, Elia has a lavish dinner; in the other scenario, Elia misses dinner. *Elia as he is in the former scenario* is happy; *Elia as he is in the latter scenario* is sad' (note also the apparent attribution of happiness *simpliciter* and sadness *simpliciter* to Elia's relevant modal slices).

Under this assumption, a counterexample to (WIS) analogous to the previous ones can be developed with respect to worlds. Suppose that an utterance of the English sentence 'Daugapils is the capital of Latvia in 2005 AD' is produced in a world w where Daugapils is the capital of Latvia in 2005 AD. That utterance as produced in w (let's call this world-bound *possibile* 'u_0') still says that Daugapils is the capital of Latvia in 2005 AD. However, my (actual) subutterance of the English sentence 'u_0 is true' is true (since the English sentence 'Daugapils is the capital of Latvia in 2005 AD' is true relative to w, and (UTRST) holds no matter what world is realized) and my (actual) subutterance of the English sentence 'Daugapils is the capital of Latvia in 2005 AD' is false (since that sentence is false relative to the actual world), thereby making false my (actual) utterance u_1 of the instance of (WIS) 'If u_0 says that Daugapils is the capital of Latvia in 2005 AD, then [u_0 is true iff Daugapils is the capital of Latvia in 2005 AD]' (since u_1 has a true antecedent and a false consequent, consisting in a true left-hand side and in a false right-hand side). Again, this is of course not the place properly to evaluate the arguments in favour and against the existence of

world-bound *possibilia* and the meaningfulness of trans-world attribution of properties *simpliciter* to them. Yet, I think it's fair to say that it would seem to be highly unsatisfactory (and surprising) if the ultimate tenability of a theory of utterance truth were to depend crucially on the falsity of some controversial but not unreasonable metaphysical doctrine.

Moreover, the metaphysical complication induced by worlds seems to be only apparent. The appearance seems to arise from the usual presupposition that the world *in which the utterance is produced* must be the world *which features in the relevant indexical (and hence evaluational) parameter* (just as it is usually presupposed that the time, standards and place at which the utterance is produced must be the time, standards and place which feature in the relevant indexical (and hence evaluational) parameters). But it needn't be so. Some uses of language seem to be best accounted for by rejecting that presupposition. A note on a piece of paper, reading 'I will be at the usual place in two hours from now', written at 4 pm and left under one's lover's office door in both parties' full awareness that it will only be read at 6 pm, is best interpreted as referring with 'now' to 6 pm rather than to the time of utterance, and as being true iff, at 6 pm, one will be at the usual place in two hours from then. Analogously, during an actual conversation on Dostoevsky's greatest characters, one's remark 'Stavrogin might easily have committed suicide earlier than he actually did' is best interpreted as referring with 'actually' to the world (or range of worlds) where the fiction of the *Demons* is true rather than to the world of utterance, and as being true iff, at that (those) world(s), Stavrogin might easily have committed suicide earlier than he did there (see Predelli [2005], pp. 40–75).

This liberal view of contexts allows for a particularly straightforward counterexample to (WIS) which only uses worlds. Suppose that an utterance u_0 of the English sentence 'Stavrogin committed suicide' is produced during an actual conversation on Dostoevsky's greatest characters and that I'm commenting on it during a non-literary, factual conversation where everyone knows full well that in fact Stavrogin never existed as a human person and *a fortiori* never committed suicide. u_0 still says that Stavrogin committed suicide. However, my subutterance of the English sentence 'u_0 is true' is true (since the English sentence 'Stavrogin committed suicide' is true relative to the world (or range of worlds) where the fiction of the *Demons* is true, and (UTRST) holds no matter what world is realized) and my subutterance of the English sentence 'Stavrogin committed suicide' is false (since that sentence is false relative to the actual world), thereby making false my utterance u_1 of the instance of (WIS) 'If u_0 says that Stavrogin committed suicide, then [u_0 is true iff Stavrogin committed suicide]' (since u_1 has a true

antecedent and a false consequent, consisting in a true left-hand side and in a false right-hand side).

In addition to *fictional* discourse, another kind of discourse that lends itself very well to the above treatment is *suppositional* discourse. In the course of a conversation, it might be supposed for the sake of argument that, in 2005 AD, the official and operative site of the Latvian Parliament, Prime Minister, President etc. is Daugapils rather than Riga. An utterance u_0 of the English sentence 'Daugapils is the capital of Latvia in 2005 AD' produced during such a conversation seems to be best evaluated with respect to the world (or range of worlds) where the supposition made for the sake of argument is true, and hence as being true iff, at that (those) world(s), Daugapils is the capital of Latvia in 2005 AD.

A counterexample to (WIS) very similar to the one from fictional discourse can then be developed. Suppose that I'm commenting on u_0 during a non-suppositional, factual conversation where everyone knows full well that in fact it is not the case that Daugapils is the capital of Latvia in 2005 AD. u_0 still says that Daugapils is the capital of Latvia in 2005 AD. However, my subutterance of the English sentence 'u_0 is true' is true (since the English sentence 'Daugapils is the capital of Latvia in 2005 AD' is true relative to the world (or range of worlds) where the supposition made for the sake of argument is true, and (UTRST) holds no matter what world is realized) and my subutterance of the English sentence 'Daugapils is the capital of Latvia in 2005 AD' is false (since that sentence is false relative to the actual world), thereby making false my utterance u_1 of the instance of (WIS) 'If u_0 says that Daugapils is the capital of Latvia in 2005 AD, then [u_0 is true iff Daugapils is the capital of Latvia in 2005 AD]' (since u_1 has a true antecedent and a false consequent, consisting in a true left-hand side and in a false right-hand side). Once again, this is of course not the place properly to evaluate the arguments in favor and against this liberal view of contexts and the related theories of fictional and suppositional discourse employed in the construction of the last counterexamples. Yet, I think that an analogous closing remark applies in this case as well: it would seem to be highly unsatisfactory (and surprising) if the ultimate tenability of a theory of utterance truth were to depend crucially on the falsity of some controversial but not unreasonable semantic doctrine.

Of course, in response to such counterexamples, the advocate of (WIS) could (and, in my view, should) restrict the quantifier over utterances in (WIS) to utterances sharing the evaluational parameters of her utterances of (instances of) (WIS).[22] The counterexamples would then be blocked. But, of course, so could one do in the case of (D) for the indexical parameters of an

utterance. What the counterexamples thus show is that (WIS) is bound to be no less restricted and contextual than (D). Moreover, given that something along the lines of (WIS) is probably our best shot at stating a plausible and natural correlation principle, we can tentatively conclude that, insofar as our understanding of truth does consist in grasping some such principle, it is bound to be irreducibly contextually restricted.

4 Semantic Paradoxes

4.1 The Standard (WIS) Response to the Semantic Paradoxes

Famously, if the object language is expressively rich enough, a sentence λ_0 (the "*liar sentence*") can be constructed which is (at least) equivalent to the English sentence 'λ_0 is not true'.[23] To fit our utterance-oriented setting, we might assume to have a language so expressive as to allow e.g. to name with 'μ' a particular utterance of the English sentence 'μ is not true'. Since μ is an utterance of 'μ is not true', (D) entails that μ is true iff μ is not true, which is a classical contradiction. Therefore, (D) is classically inconsistent when its quantifier ranges over utterances of expressively rich enough languages.

Again, at this point, the advocate of (WIS) will hasten to point out that the *provisional* formulation of (WIS) offers a way out of paradox: all an analogous argument proves in the case of (WIS) is that (WIS) *and* the assumption that μ *says* that μ is not true are jointly classically inconsistent, which she blithely regards as a *reductio* of the assumption that μ says that μ is not true. Since it seems rather uncontroversial that, if it manages to say anything at all, μ says that μ is not true, she should conclude that μ does not say anything at all. Surprisingly enough, on this view (augmented with the additional, extremely plausible assumption that an utterance is true only if it says something), it is indeed the case that μ is not true, but, somewhat mysteriously, this truth fails to be expressed by the grammatically well-formed utterance μ. Yet, the semantic paradoxes are just that—*paradoxes*—and one might well wish to pay this cost.

Another relatively minor cost, already alluded to in note 20, is that the view under the consideration would undermine the unrestricted validity of (UTRST). For presumably the theorist's utterance θ of the English sentence 'μ is not true' should count as true. By (UTRST) (left-to-right), 'μ is not true' is true (in the "intended model" of English) relative to the sequence of

indexical parameters associated with θ and to the sequence of evaluational parameters thereby determined. However, on the face of it, these two sequences are exactly the same sequences relevant for the evaluation of μ's truth. Since 'μ is not true' is true (in the "intended model" of English) relative to them, by (UTRST) (right-to-left) μ itself should count as true, contrary to what θ says. Something has to give here, and it may well be that the best way to go for the view under consideration is to restrict (UTRST) in some principled way.

4.2   Revenge: Quantified Utterance-Truth Liar

However, the costs are even higher than that. For a natural variation on the liar sentence in utterance-theoretic terms suffices to reduce the advocate of (WIS) to pragmatic inconsistency. We can assume that our language is expressively rich enough as to allow us to construct the sentence:[24]

λ_1 Every utterance of λ_1 is not true

Now, take an utterance u of λ_1. (D) entails that u is true only if every utterance of λ_1 is not true, which, together with u's being an utterance of λ_1, entails, by *reductio*, that u is not true. Therefore, by universal generalization, (D) entails that every utterance of λ_1 is not true. But, certainly, we want to allow that there might be an utterance u of λ_1. However, entailing as we have just seen (left-to-right) that every utterance of λ_1 is not true, (D), together with u's being an utterance of λ_1, entails both that u is not true (straightforward) and that u is true (right-to-left). Therefore, if there are utterances of λ_1 at all (which we can safely assume in the following, having just made it true!), (D) is false (assuming classical logic, which I will do up until the end of section 4.3).

Again, at this point, the advocate of (WIS) might suggest the following train of thought:

> All an analogous argument proves in the case of (WIS) is that (WIS) *and* the assumption that any utterance u of λ_1 says that every utterance of λ_1 is not true are jointly classically inconsistent, which I can blithely regard as a *reductio* of the assumption that u says that every utterance of λ_1 is not true. Since I agree that, if it manages to say anything at all, u says that every utterance of λ_1 is not true, I conclude that u does not say anything

at all. And since I also agree that an utterance is true only if it says something, I conclude that it is indeed the case that every utterance of λ_1 is not true.

Unfortunately, by so concluding, she is accepting something which, jointly with some trivial syntactic facts, entails the untruth of her last utterance. In the presence of λ_1 (or of some of its like) and of the extremely plausible assumptions (a) that any utterance of λ_1 says that every utterance of λ_1 is not true if it says anything at all and (b) that saying something is necessary for utterance truth, a (WIS)-based theory of utterance truth embarrassingly entails something which cannot be recognized as such in thought and talk. It cannot be so recognized because doing so would require uttering something which is, by the very same lights of an advocate of (WIS), untrue, and accepting something which is known to be untrue does seem to do violence to our notion of truth (*pace* the recent heroic attempt of Maudlin [2004]). But, presumably, it is a necessary condition on the *rational* acceptability of a theory that it at least be possible to follow through its logical consequences in thought and talk. Thus, even if it were *consistent*, a (WIS)-based theory would not be *rationally acceptable*.[25]

4.3 Revenge: Contingent Quantified Propositional Liar

I would like to close by casting doubt even on the sheer consistency of the package (classical logic, 'say that'-constructions and second-order quantification) on which a (WIS)-based theory of utterance truth typically relies.[26] It is well-known that sometimes paradox is *conditional on empirical contingencies*. The sentence:

λ_2 Every utterance of the sentence mentioned in such-and-such paper to illustrate a standard contingent semantic paradox is not true

(where 'such-and-such paper' is short for a non-indexical definite description picking out in actuality this paper) might have not been paradoxical (and, indeed, plain false) had I been so logically untalented as to choose 'Snow is white' in its stead as an illustration of a standard contingent semantic paradox. According to the advocate of (WIS), strange as it may seem, any utterance of λ_2 does not manage to say anything in actuality, but the pill is sweetened by conceding that, of course, it might easily have said just what it does appear to say—had I only been less logically talented.

Now, let λ_3 be the following displayed sentence:

$$\lambda_3 \quad \exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), \neg X),$$

where '$\exists!^2$' is a generalized second-order uniqueness quantifier (see note 18), 's' means what 'the sentence mentioned in such-and-such paper to illustrate a final problem for (WIS)' does (same convention as before for 'such-and-such paper'), $\ulcorner \mathit{Of}(\tau_0, \tau_1) \urcorner$ means what $\ulcorner \tau_0$ is an utterance of $\tau_1 \urcorner$ does, $\ulcorner N(\tau) \urcorner$ means what $\ulcorner \tau$ is uttered in non-paradoxical circumstances $\urcorner$ does, '$>$' expresses counterfactual implication and $\ulcorner \mathit{Say}(\tau, \varphi) \urcorner$ means what $\ulcorner \tau$ says that $\varphi \urcorner$ does.

Let me first offer a brief informal version of the argument. On the one hand, since, as a matter of fact, λ_3 is the sentence mentioned in such-and-such paper to illustrate a final problem for (WIS), what would be said in non-paradoxical circumstances by an utterance of the sentence mentioned in such-and-such paper to illustrate a final problem for (WIS) is the case iff what would be said in non-paradoxical circumstances by an utterance of λ_3 is the case. On the other hand, clearly, what would be said in non-paradoxical circumstances by an utterance of λ_3 is that what would be said in non-paradoxical circumstances by an utterance of the sentence mentioned in such-and-such paper to illustrate a final problem for (WIS) is not the case. It is therefore the case iff what would be said in non-paradoxical circumstances by an utterance of the sentence mentioned in such-and-such paper to illustrate a final problem for (WIS) is not the case. Thus, putting the two pieces together, we have that what would be said in non-paradoxical circumstances by an utterance of the sentence mentioned in such-and-such paper to illustrate a final problem for (WIS) is the case iff what would be said in non-paradoxical circumstances by an utterance of the sentence mentioned in such-and-such paper to illustrate a final problem for (WIS) is not the case, which is a classical contradiction.

Now for a more formal and detailed version of the argument.[27] As a matter of fact:

(1) $\lambda_3 = s,$

and so, by Leibniz' Law,

(2) $\exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), X) \equiv \exists!^2 X(\forall u(\mathit{Of}(u, \lambda_3) \supset (N(u) > \mathit{Say}(u, X))), X).$

However, given her concessions in the case of standard contingent semantic paradoxes, the advocate of (WIS) should have no objection to granting the following *disquotational* principle about what is said:

(WISD) $\forall u(Of(u, `\varphi\text{'}) \supset (N(u) > Say(u, \varphi)))$

if φ does not contain indexical expressions, and to the following *comprehension* principle about what is said:

(WISC) $\exists!^2 X(\forall u(Of(u, `\varphi\text{'}) \supset (N(u) > Say(u, X))), \top)$

(where '$\top$' is a logical constant which takes as value, in every model $\mathfrak{M}$, $\mathfrak{M}$'s domain of entities (whatever they may be) which are the values of 0ary second-order variables) if φ is not ambiguous. But λ_3 meets both these conditions,[28] and so the following instances of, respectively, (WISD) and (WISC) should be acceptable to her:

(3) $\forall u(Of(u, `\exists!^2 X(\forall u(Of(u, s) \supset (N(u) > Say(u, X))), \neg X)\text{'}) \supset (N(u) > Say(u, \exists!^2 X(\forall u(Of(u, s) \supset (N(u) > Say(u, X))), \neg X))));$

(4) $\exists!^2 X(\forall u(Of(u, `\exists!^2 X(\forall u(Of(u, s) \supset (N(u) > Say(u, X))), \neg X)\text{'}) \supset (N(u) > Say(u, X))), \top)$.

Since the definition of λ_3 ensures:

(5) $\lambda_3 = `\exists!^2 X(\forall u(Of(u, s) \supset (N(u) > Say(u, X))), \neg X)\text{'}$,

by Leibniz' Law we have:

(6) $\forall u(Of(u, \lambda_3) \supset (N(u) > Say(u, \exists!^2 X(\forall u(Of(u, s) \supset (N(u) > Say(u, X))), \neg X))))$

from (3) and:

(7) $\exists!^2 X(\forall u(Of(u, \lambda_3) \supset (N(u) > Say(u, X))), \top)$

from (4). Now, the following *instantiation* principle about what is said is apt to strike one as a (second-order) logical truth:[29]

24

(WISI) $(\forall u(\mathit{Of}(u, \tau) \supset (N(u) > \mathit{Say}(u, \varphi))) \land \exists!^2 X(\forall u(\mathit{Of}(u, \tau) \supset (N(u) > \mathit{Say}(u, X))), \top)) \supset (\exists!^2 X(\forall u(\mathit{Of}(u, \tau) \supset (N(u) > \mathit{Say}(u, X))), X) \equiv \varphi),$

and the relevant instance for 'λ_3' and '$\exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), \neg X)$' is:

(8) $(\forall u(\mathit{Of}(u, \lambda_3) \supset (N(u) > \mathit{Say}(u, \exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), \neg X))))) \land \exists!^2 X(\forall u(\mathit{Of}(u, \lambda_3) \supset (N(u) > \mathit{Say}(u, X))), \top)) \supset (\exists!^2 X(\forall u(\mathit{Of}(u, \lambda_3) \supset (N(u) > \mathit{Say}(u, X))), X) \equiv \exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), \neg X)).$

But (6), (7) and (8) jointly entail:

(9) $\exists!^2 X(\forall u(\mathit{Of}(u, \lambda_3) \supset (N(u) > \mathit{Say}(u, X))), X) \equiv \exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), \neg X).$

This, together with (2), yields:

(10) $\exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), X) \equiv \exists!^2 X(\forall u(\mathit{Of}(u, s) \supset (N(u) > \mathit{Say}(u, X))), \neg X),$

which, together with (7) and the classically valid rule:

(R$_2$) $\exists!^2 X(\Phi(X), \Psi(X)) \equiv \exists!^2 X(\Phi(X), \neg \Psi(X)), \exists!^2 X(\Phi(X), \top) \vdash \exists!^2 X(\Phi(X), \Psi(X) \land \neg \Psi(X)),$

finally yields contradiction.

Although it does not exploit (WIS) directly (albeit a modified, clumsier version of it could), the construction of such contradiction crucially relies on the notion—so dear to the advocate of (WIS)—of what is said and on the same self-referential mechanism exploited by more familiar contingent semantic paradoxes. Clearly, it cannot be avoided simply by claiming that an utterance of λ_3 fails to say what it seems to say.[30]

On any plausible regimentation of 'what an utterance says is the case'-locutions (as the one adopted here), the argument does require *impredicative* second-order quantification. Discussion of the (dubious) philosophical viability of a ramified hierarchy of ranges for the second-order quantifier(s) lies

however outside the scope of this paper. Here, in a spirit similar to that of my closing comment in section 3, I would just like to remark that, once such seemingly *ad hoc* restrictions on quantification are allowed, it is hard to see how an analogous restriction on the (first-order) quantification over utterances on behalf of (D), blocking self-reference at least for liar-like sentences, would be any worse. In my view, it is more plausible to point the finger at the apparently unwarranted strength of the underlying (classical) logic, which allows us to go (both in the case of the quantified utterance-truth liar and in the case of the contingent quantified propositional liar) from apparently acceptable *conditional* statements of the form 'P only if it is not the case that P', reflecting the connection between a sentence and its negation typical of some paradoxical sentences, to apparently unacceptable *categorical* statements of the form 'It is not the case that P' (which in turn quickly lead to contradiction).

5   Conclusions and Glimpses Beyond

(WIS) is arguably the most plausible and natural correlation principle on the market. Its conditionalization of the connection between the truth of an utterance of 'P' and its being the case that P on the utterance's saying that P seemed to hold the key to a unified solution of the problems arising for other correlation principles (such as (D)) from context dependence and the semantic paradoxes.

I have argued that this is an illusion generated by the failure of taking into account the full range of possible varieties of context dependence (especially, locality) and of semantic paradoxes (especially, quantificational paradoxes). I have also briefly suggested that, as far as the semantic paradoxes are concerned, a correlation principle may be saved by an appropriate weakening of the underlying logic, even though, given context dependence, we cannot expect such a principle to hold in full generality. In closing, I should also like to suggest that such a failure indicates that our understanding of truth is precisely *not* exhausted by any correlation principle, so that a more fundamental principle should be sought on the basis of which successes and failures of correlation principles may be explained. Both these suggestions will however have to be taken up on another occasion.

Notes

[1]Henceforth, for short, '*sentence*'. As our interest will exclusively be with the vagaries of words-world relations rather than with those of words-words relations, sentences (as well as, more generally, expressions) will be individuated until further notice (note 24) at a purely syntactic level, yet fine-grainedly enough as to resolve any syntactic ambiguity. For example, there are (at least) two sentences corresponding to the single string of words 'necessarily'ˆ'everything'ˆ'exists'. Accordingly, quotation (and semi-quotation) will be understood as picking out such fine-grained syntactic entities. Also, I will use the phrase 'the English sentence s' as synonymous with 'the sentence s as used in English'.

[2]Henceforth, for short, '*utterances*'. The same fine-grainedness of individuation can be achieved by focusing on pairs of (semantically unambiguous representations of) sentences and contexts with suitable parameters (the "*occurrences*" of many formal semanticists) rather than on concrete events of use. Indeed, pairs of (semantically unambiguous representations of) sentences and contexts would lend themselves to a smoother semantic theorizing, especially if the theory of utterance truth is to be connected in a systematic way with the theory of validity, which will arguably assign (model-relative) semantic values to pairs of (semantically unambiguous representations of) expressions and contexts rather than to concrete events of use of an expression (see Kaplan [1989], pp. 522–523, 546). However, as this distinction is rather tangential to the issues we'll be mostly concerned with here, for ease of discussion I'll mainly talk about concrete events of use. I also wish to remain neutral with regard to the question whether contexts themselves should be taken as concrete (centered) situations of use or as abstract sequences of suitable parameters.

[3]Not all events of use of a sentence token are assertoric: one could e.g. use a token of the sentence 'I'm a philosopher' to show a fine example of calligraphy, and no relevant notion of truth would seem to apply to such an event of use (even though that event of use might still count in some sense as correct or incorrect depending on whether the token shown is indeed a piece of beautiful handwriting). Throughout, I will presuppose an intuitive understanding of what it is for an event of use to be assertoric, without trying to address the difficult philosophical task of providing an account of the nature of assertion. Also, I don't wish to presuppose that no interesting notion of truth applies to any event of non-assertoric use of a sentence token: in the following, I myself will extend 'truth'-talk in a natural way in this direction in section 3.4, and there might well be an interesting sense in which e.g. conjectures (in the "act sense", not just in the "object sense") can count as true or false. I'm confident that the discussion to follow will be highly relevant to such a notion as well, but, to fix ideas, I propose to focus attention on utterance truth. Thanks to Laurence Goldstein for making me think hard about these issues.

[4]It is common to think that there is a sense of 'say' in which a typical utterance of the English sentence 'Most students passed the exam' says (among other things) that not every student passed the exam and also to think at the same time that the fact that every student passed the exam does not tell against the truth of the utterance. Under such a combination of views (which I myself am least inclined to reject), some version or other of the much debated distinction between what an utterance *semantically expresses* and what an utterance *pragmatically conveys* will be needed. For all those of us who accept some version or other of this distinction, 'say' and its like should of course be read throughout as latching onto what an utterance semantically expresses.

[5]Throughout, I will use 'P' as schematic for declarative English sentences (whereas

'' P '' sometimes refers schematically to a sentence, some other times refers to a sentence schema—as has happened e.g. at the beginning of this footnote). Also, I will use 'φ' (possibly with subscripts) as a metalinguistic variable ranging over such sentences. Finally, I will use 'X' (possibly with subscripts) as a (bindable) second-order 0ary variable— if you like, as a (bindable) propositional variable, even though—some informal glosses notwithstanding—I will officially remain neutral as to the vexed question of the proper interpretation of second-order quantification.

[6]I will only consider theories whose background logic is classical and which use 'if' to express classical material implication.

[7]Throughout, I use square brackets to disambiguate scope in English.

[8]Such an understanding figures prominently in the theory of utterance truth of Thomas Bradwardine, a venerable philosopher and logician belonging to the school of the Oxford *calculatores* in the early 13th century (see Read [2002]). For our purposes, the main principle of Bradwardine's theory can be treated as equivalent with (WIS$^\forall$) below in the text. Thanks to Stephen Read for extremely fruitful discussions of Bradwardine's theory over the years.

[9]Save, very plausibly, for an additional presupposition of existence (of u) which is arguably present in the former but, of course, not in the latter case (see Field [1994], pp. 250–251).

[10]Needless to say, connections between truth and what is said have formed one of the major foci of much twentieth-century philosophy of language, especially in the tradition of truth-conditional semantics, seeking to illuminate the notion of what is said *via* the notion of truth. However, a (WIS)-based theory of (utterance) truth pursues rather the *converse* project, seeking to illuminate the notion of (utterance) truth *via* the notion of what is said.

[11]Thus, I'm going to use 'indexicality' in a broader sense than the one in use in some contemporary debates, according to which e.g. an apparently $i$ary predicate is "indexical" iff it refers in different contexts to different $i$ary properties (rather than referring in different contexts to the same $i+1$ary property with different objects being supplied in different contexts as arguments for the $i+1$th place of the property). I stress that, in the broad intended sense, even though good enough for our purposes the characterization is rough (one aspect in which it might be rough will be hinted at in note 28).

[12]For lack of a better term, I use '*appear*' and its like to denote the relation between an expression and an utterance that involves some tokens of the expression, while I reserve '*occur*' and its like to denote the relation between a simpler expression and a more complex expression that involves it. Much more would need to be said on the metaphysics of linguistic entities than I'll be able to do here.

[13]By '*subutterance*' I mean the event of (typically non-assertoric) use of an expression token as involved in the possibly wider event of assertoric use of a sentence token (i.e. utterance). Take for example an utterance of the English sentence 'If snow is white, then something is white'. The particular event of (non-assertoric) use of the sentence token of the English sentence 'Snow is white' involved in such utterance is a subutterance. In the following, I extend in a natural way the notions of referring, saying, truth and the like so that they apply also to subutterances of sentences. I also count utterances as subutterances.

[14]This very same feature of (WIS) makes also dispensable in it any reference to *languages*. (We can assume that) it is not the case that a (vocal) utterance by a monolingual German speaker of the sentence 'Empedocles leaped' is true iff Empedocles leaped (rather, it is true iff Empedocles loves), but that is no problem for (WIS), as such an utterance does not say that Empedocles leaped. (D) itself can be made to accommodate conventionality either by individuating sentence types semantically, or by relativizing truth to the language of the metatheory (in our case, English), or by expanding the antecedent of (D) to something equivalent to 'u is an utterance of 'P' in the language of the metatheory' (which, in our case, comes down to expanding it to 'u is an utterance of 'P' in English'). Analogous points apply for ambiguity (see section 3.2 for more on these issues).

[15]Notice that (WIS) may nevertheless be trivially subject to counterexamples from indexicality. Consider an utterance u_0 of the English sentence 'I'm lying now' produced by me at a time t at which I'm lying on my bed. u_0 is true, since 'now' as it appears in it refers to a time (t) at which I am indeed lying. Consider next a very slow utterance u_1 of the instance of (WIS) 'If u_0 says that I'm lying now, then [u_0 is true iff I'm lying now]', produced in such a way that its antecedent is still produced at t, whereas its consequent is produced at a subsequent time at which I'm standing. As a matter of fact, u_1 has a true antecedent and a false consequent (the consequent being false since its left-hand side is true and its right-hand side false). This kind of putative counterexample crucially relies on the semantic hypothesis that it is possible to assign different referents to different appearances of the same expression in the same utterance. It would bring out a different way in which indexicality can falsify instances of a correlation principle, not by affecting in a different way the interpretation of two subutterances of an expression one of which is mentioned and the other used in an utterance of (an instance of) the principle (as happens in our counterexamples to (D)), but by affecting in a different way the interpretation of two subutterances of an expression which are both used in an utterance of (an instance of) the principle. I mention this kind of putative counterexample only to set it aside as irrelevant to our concerns, since, once the semantic hypothesis on which it relies is granted, *every* principle (even logical truths!) would be subject to similar "counterexamples". Thanks to Crispin Wright for raising this issue.

[16]An even further level of variation would be given by the recognition of an implicit *relativity* in utterance truth (MacFarlane [2005]). Even keeping fixed the situation talked about, two (or even one and the same!) utterances of the same sentence could then be differently assessed by different speakers: an utterance of the sentence the English sentence 'Codfish tastes good' as assessed by me would be true, while an utterance of the same sentence (even that very same utterance!) as assessed by my sister would be false. On some versions of the view ("*truth relativism*" properly called), this is so because, even though the utterance can be assessed by me and by my sister as saying the same thing (namely, that codfish tastes good), what is said can be assessed by me as being the case and by my sister as not being the case. On some other versions of the view ("*content relativism*"), this is so because the utterance can be assessed by me as saying e.g. that codfish tastes good to me and by my sister as saying that codfish tastes good to her. Utterance-truth relativity (in any form) is however a much more controversial "phenomenon" than those mentioned in the text and raises issues that cannot be properly dealt with within the boundaries of this paper. I will therefore have to ignore relativist views here (save for a brief remark in note 20).

[17]Whose logical form, we may assume, is '$\exists!^1 x(F(\tau, x), Gx)$', where '$\exists!^1$' is a generalized first-order uniqueness quantifier (namely, such as to take as value, in every model $\mathfrak{M}$,

the set $\{\langle A, B \rangle : |A| = 1 \text{ and } A \subseteq B\}$, where '$A$' and '$B$' are variables ranging over the power-set of $\mathfrak{M}$'s domain).

[18]Whose logical form, we may assume, is '$\exists!^2 X(F(\tau, X), X)$', where '$\exists!^2$' is a generalized second-order uniqueness quantifier (namely, such as to take as value, in every model $\mathfrak{M}$, the set $\{\langle A, B \rangle : |A| = 1 \text{ and } A \subseteq B\}$, where '$A$' and '$B$' are variables ranging over the power-set of $\mathfrak{M}$'s domain of entities (whatever they may be) which are the values of 0ary second-order variables).

[19]Thanks to François Recanati for very helpful discussions on this topic.

[20]Substantially the same connection between utterance truth and relativized sentential truth is endorsed e.g. in Kaplan [1989], pp. 522–523, 547. To stress, (UTRST) is extremely plausible, so much as to be nowadays very widely accepted. Actually, we'll see in section 4.1 that the advocate of (WIS) is surprisingly committed to denying the unrestricted validity of (UTRST), if the theory is supposed to do some substantial work in avoiding the semantic paradoxes. I regard this as a very counterintuitive consequence of the theory, and, in any event, I don't see how the alleged failures of (UTRST) in the case of the semantic paradoxes do anything to motivate similar failures in the case of the counterexamples to (WIS) from context dependence which follow in the text. Another threat to (UTRST) is constituted by some of the relativist views mentioned in note 16. Even though this is not the place properly to evaluate such views, in relation to our dialectic I should remark that it is at best unclear that the usual arguments supporting relativism can be used to justify a violation of (UTRST) in every case in which a counterexample to (WIS) from context dependence threatens. Thanks to Timothy Williamson for discussion of (UTRST).

[21]I note in passing that such counterexamples from standards would call for a substantial qualification of the celebrated argument in Williamson [1992], pp. 145–149 against denials of bivalence for utterances concerning borderline cases. The argument reduces to contradiction denials of bivalence by using (WIS) (and assuming 'iff' to be contraposable). For all the argument shows, denials of bivalence of utterances made at relevantly different contexts of utterance are consistent.

[22]As both Julie Hunter and Dan López de Sa have emphasized to me, an alternative reaction would be to introduce in the language a shifting operator ⌜relative to τ's evaluational parameters⌝ such that ⌜Relative to τ's evaluational parameters, φ⌝ is true relative to a sequence of indexical parameters s_0 and to a sequence of evaluational parameters s_1 iff φ is true relative to s_0 and to the sequence of evaluational parameters determined by the sequence of indexical parameters associated with the referent of τ relative to s_0 and to s_1. One could then endorse the following variation on (WIS):

(WIS^R) For every utterance u, if u says that P, then [u is true iff, relative to u's evaluational parameters, P].

(WIS^R) is of course proof against the kind of counterexample I've been developing in the text. But (WIS^R) is no correlation principle and, under this respect, one might feel that it is closer to being a fancy object-linguistic rewriting of (UTRST) than to being a revision of (WIS). Moreover, an analogous maneuver appealing to shifting operators is available in the case of (D) (where the operators in question would have to be "monsters", in the sense of Kaplan [1989], pp. 510–512).

[23]There are of course many other extremely interesting semantic paradoxes beyond the liar (such as e.g. Curry's). Since, however, the full complexity of the phenomenon is

somehow tangential to the points I want to discuss, in the following I will focus attention on liar-like paradoxes.

[24]To avoid what would now be needless complications, assume henceforth a richer, semantic conception of a sentence (as well as, more generally, of expressions), such that a sentence is individuated also by its linguistic meaning (so that e.g. λ_1 essentially means that every utterance of λ_1 is not true), and let quotation (and semi-quotation) work accordingly. Ignore also possible ambiguities (if there are any).

[25]After completing a draft of this section, I was informed by Graham Priest that a similar argument can be found in Hazen [1987]. Hazen however uses similar considerations only to attack Jean Buridan's specific theory of utterance truth, failing to modify them in such a way that they become relevant to a wider range of theories (like the one based on (WIS)) which hold that paradoxical utterances are not true. Smiley [1993], pp. 25–26 constitutes an early attempt at a generalization—any such attempt, however, will have to face the complexities that might be induced by an utterance-based approach to the semantic paradoxes, as pointed out by Hinckfuss [1991]. Fortunately, such complexities can be ignored in the specific case of (WIS) (and (WIS$^\forall$)), for which the argument I gave in the text goes through.

[26]True, we officially stated (WIS) as a *first-order schema* rather than as a *second-order sentence*, and that has been just as good for the purposes of our discussion so far. However, the schematic formulation of (WIS) suffers from the usual logical weakness of schematic principles in general: in particular, it does not constrain in the intended way the truth of utterances saying things that are not said by any utterance in English (Andjelković and Williamson [2000], pp. 216–217 make an analogous point in discussing the shortcomings of a substitutional interpretation of second-order quantification). Moreover, second-order quantification is apparently involved in many crucial 'say that'-constructions, like 'Everything he said is the case'. In any event, the revenge paradox to follow in the text will be of interest to those of us who wish to leave open the possibility of making use of such resources in their theories. Note that the use which the paradox makes of these resources is quite limited and consistent with a substitutional interpretation. Note also that similar qualifications don't apply to (WIS$^\forall$), which essentially involves second-order quantification.

[27]Thanks to Hartry Field for probing questions that led to the present regimentation of the argument, and to Timothy Williamson for helpful feedback on it.

[28]If you still think that 's' is indexical because, being contingent, picks out different objects in different contexts, substitute for it the corresponding quantifier, which presumably isn't. If you think that λ_3 has to be tensed and that tense contributes indexicality, replace the indexical expressions with non-indexical definite descriptions, and then apply to such descriptions what the previous sentence says about 's'.

[29]Assume the antecedent of (WISI). Then:

Left-to-right. An instance of the rule:

(R$_0$) $\exists!^2 X(\Phi(X), \Psi(X)), \Phi(\varphi) \vdash \Psi(\varphi)$.

Right-to-left. An instance of the rule:

(R$_1$) $\exists!^2 X(\Phi(X), \top), \Phi(\varphi), \Psi(\varphi) \vdash \exists!^2 X(\Phi(X), \Psi(X))$.

Thanks to Julie Hunter for discussion of this step.

[30]The argument thus generalizes to every attempt at solving the semantic paradoxes by claiming that the utterance in question fails to say what it seems to say. Some prominent contemporary defenders of this approach are Ryle [1950]; Bar-Hillel [1957]; Prior [1961]; Kneale [1971]; Mackie [1973], pp. ???; Goldstein [1992]; Smiley [1993]; Williamson [1998]; Sorensen [2001], pp. ???. The view seems to go back to Chrysippus (see Bocheński [1970], pp. ???).

References

Miroslava Andjelković and Timothy Williamson. Truth, falsity, and borderline cases. *Philosophical Topics*, 28:211–244, 2000.

Yehoshua Bar-Hillel. New light on the liar. *Analysis*, 18:1–6, 1957.

Jozef Bocheński. *A History of Formal Logic*. Chelsea, New York, 2nd edition, 1970.

Herman Cappelen and John Hawthorne. *Relativism and Monadic Truth*. Oxford University Press, Oxford, 2009. URL http://folk.uio.no/hermanc/docs/RelativismAndMonadicTruthFinal.pdf.

Michael Dummett. Existence, possibility and time. In Georg Meggle and Andreas Mundt, editors, *Analyomen 2*, volume I, pages 43–67. De Gruyter, Berlin, 1997.

Hartry Field. Deflationist views of meaning and content. *Mind*, 103:249–285, 1994.

Laurence Goldstein. 'This statement is not true' is not true. *Analysis*, 52: 1–5, 1992.

Allen Hazen. Contra Buridanum. *Canadian Journal of Philosophy*, 17:875–880, 1987.

Ian Hinckfuss. Pro Buridano; contra Hazenum. *Canadian Journal of Philosophy*, 21:389–398, 1991.

David Kaplan. Demonstratives. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes from Kaplan*, pages 481–563. Oxford University Press, Oxford, 1989.

William Kneale. Russell's paradox and some others. *The British Journal for the Philosophy of Science*, 22:321–338, 1971.

David Lewis. Attitudes *De Dicto* and *De Se*. *The Philosophical Review*, 88:
  513–543, 1979.

David Lewis. Index, context, and content. In Stig Kanger and Sven Öhman,
  editors, *Philosophy and Grammar*, pages 79–100. Reidel, Dordrecht, 1980.

John MacFarlane. Making sense of relative truth. *Proceedings of the Aris-
  totelian Society*, 105:321–339, 2005.

John Mackie. *Truth, Probability and Paradox*. Clarendon Press, Oxford,
  1973.

Tim Maudlin. *Truth and Paradox*. Oxford University Press, Oxford, 2004.

Philip Percival. Absolute truth. *Proceedings of the Aristotelian Society*, 53:
  189–213, 1994.

Stefano Predelli. *Contexts*. Oxford University Press, Oxford, 2005.

Arthur Prior. On a family of paradoxes. *Notre Dame Journal of Formal
  Logic*, 2:16–32, 1961.

Stephen Read. The liar paradox from John Buridan back to Thomas Brad-
  wardine. *Vivarium*, 40:189–218, 2002.

François Recanati. *Literal Meaning*. Cambridge University Press, Cambridge,
  2004.

François Recanati. *Perspectival Thought*. Oxford University Press, Oxford,
  2007.

Gilbert Ryle. Heterologicality. *Analysis*, 11:61–69, 1950.

Timothy Smiley. Can contradictions be true? *Proceedings of the Aristotelian
  Society Supplementary Volume*, 67:17–33, 1993.

Roy Sorensen. *Vagueness and Contradiction*. Oxford University Press, Ox-
  ford, 2001.

Charles Travis. Pragmatics. In Bob Hale and Crispin Wright, editors, *A
  Companion to the Philosophy of Language*, pages 87–107. Blackwell, Ox-
  ford, 1997.

Timothy Williamson. Vagueness and ignorance. *Proceedings of the Aris-
  totelian Society Supplementary Volume*, 62:145–162, 1992.

Timothy Williamson. Indefinite extensibility. *Grazer Philosophische Studien*,
  55:1–24, 1998.