AUSTRALASIAN JOURNAL OF PHILOSOPHY
https://doi.org/10.1080/00048402.2021.1933114

Against Disquotation

Andrew Bacon and Jeremy Goodman

University of Southern California

ABSTRACT
We argue against the disquotational meaning schema—'φ' means that φ—and suggest that rejecting it is the key to resolving further intensional paradoxes about the limits of thought.

ARTICLE HISTORY Received 4 March 2020; Revised 23 April 2021

KEYWORDS meaning; truth; paradox; disquotation

1. Against T

There is something extremely compelling about the following schema:[1]

(T) 'φ' is true if and only if φ.

But now consider the instance of T obtained by replacing the schematic sentence letter φ with the sentence L = 'L is not true':[2]

(1) 'L is not true' is true if and only if L is not true.

By Leibniz's Law, we have

(2) L is true if and only if L is not true.

And (2) can be reduced to absurdity in classical propositional logic.

[1] Instances of schemas like T are obtained by substituting, for φ, any declarative English sentence, understood generously to include sentences with logical notation, semantic jargon, and names introduced by stipulation. For ease of exposition, for now we will treat the bearers of truth/meaning (in English) as sentence-types and identify these with strings of symbols. Later, we will show that a more realistic treatment, where the bearers of truth/meaning are sentences in context, or particular uses of sentences, does not threaten our mode of argument.

[2] Can we legitimately stipulate that L = 'L is not true'? One might think that the most we can stipulate is that 'L' refer to 'L is not true.' This would imply the above identity, assuming (i) that 'L' refers to L, (ii) that 'L' refers to at most one thing, and (iii) classical quantificational reasoning. But since we argue, in section 2, that we must reject either disquotational principles for meaning, the uniqueness of meanings, or classical quantificational reasoning about meanings, one might object that, by our lights, the combination of (i)–(iii) is dubious. We have three replies. First, (i)–(iii) are extremely plausible, independently of any general principles about disquotation, unique reference, and quantification. Second, (ii) and (iii) seem especially secure, and the falsity of (i) would allow us to argue against disquotational principles (our aim) by other means. Third, and most importantly, our argument does not rest on the good standing of this way of introducing self-referential sentences; it would be straightforward to use, instead, alternatives techniques like Gödelian diagonalization or Kripkean contingent liar sentences.

At this point, some will give up classical logic. Our view is that this is the wrong reaction: the benefits are not worth the costs. Here is not the place to defend this position at any length, but let us look briefly at some of the relevant considerations. The costs of rejecting classical logic are fairly widely appreciated. For example, the most common strategy for doing so involves giving up the law of excluded middle.[3] Less widely appreciated is that the benefits of rejecting classical logic are not as great as they are often taken to be. For, although we can hold on to the letter of T, it is not clear that we can accept it with the interpretation on which it was originally compelling. For example, if we reject the law of excluded middle, then we cannot accept T on the interpretation of 'if and only if' as the material biconditional (defined in terms of conjunction, negation, and disjunction), since, on that interpretation, T entails the law of excluded middle.[4] In fact, we cannot accept T for any conditional $\rightarrow$ that obeys *modus ponens* (from $\varphi \rightarrow \psi$ and φ, infer ψ) and *contraction* (from $\varphi \rightarrow (\varphi \rightarrow \psi)$, infer $\varphi \rightarrow \psi$), for reasons to do with the Curry paradox.[5] We find these principles no less compelling than T, and so we are not convinced that, when understood in terms of a conditional that fails to satisfy them, T is faithful enough to its pre-theoretical motivations to warrant rejecting classical logic in order to salvage it.[6]

None of this is to suggest that non-classical approaches to the semantic paradoxes aren't worthy of serious consideration. But in this paper we will restrict our attention to classical approaches.

T classically follows from the following two schemas, which most philosophers find no less compelling than T:[7]

---

[3] Paraconsistent strategies keep the law of excluded middle but reject disjunctive syllogism.

[4] Suppose that $\varphi \supset true(\text{'}\varphi\text{'})$ and $true(\text{'}\varphi\text{'}) \supset \varphi$. Since even those who reject the law of excluded middle accept the transitivity of $\supset$, it follows that $\varphi \supset \varphi$, which is equivalent to the law of excluded middle, given the definition of $\varphi \supset \psi$ as $\neg\varphi \vee \psi$.

Paraconsistent theories like those of Priest [1979] (according to which not everything follows from a contradiction) can accept T, formulated with the material biconditional, at the cost of losing the transitivity of $\supset$. But since, in such systems, (i) $\supset$ fails to obey *modus ponens*, and (ii) one can prove the *negations* of instances of T, proponents of such theories are under strong pressure to interpret the conditional in T as something other than the material conditional (which is what they in fact do).

[5] Consider the sentence $C = \text{'}true(C) \rightarrow \bot\text{'}$, where $\bot$ is a contradiction. We have $true(C) \rightarrow (true(C) \rightarrow \bot)$ (by T, left-to-right); hence, $true(C) \rightarrow \bot$ (by contraction); we also have $(true(C) \rightarrow \bot) \rightarrow true(C)$ (by T, right-to-left); and hence $\bot$ (by two applications of *modus ponens*).

[6] In addition to being compelling on its own, contraction can be derived from the following even more seemingly obvious principles about conditionals:

(ID) $\varphi \rightarrow \varphi$
(PMP) $(\varphi \wedge \varphi \rightarrow \psi) \rightarrow \psi$
(MP) From $\varphi \rightarrow \psi$ and φ, infer ψ.
(CC) From $\varphi \rightarrow \psi$ and $\varphi \rightarrow \chi$, infer $\varphi \rightarrow (\psi \wedge \chi)$.
(Tran) From $\varphi \rightarrow \psi$ and $\psi \rightarrow \chi$, infer $\varphi \rightarrow \chi$.

Contraction can also be motivated by considerations linking restricted quantification to conditionals. Consider the argument:

Every philosopher admires every logician. $(\forall x(Px \rightarrow \forall y(Ly \rightarrow Axy)))$
Every logician is a philosopher. $(\forall x(Lx \rightarrow Px))$
Therefore: every logician admires themselves. $(\forall x(Lx \rightarrow Axx))$

On views like those of Field [2014], which link the validity of principles involving restricted quantification to principles involving quantified conditionals, the above argument is valid only if so too is this argument form: from $\varphi \rightarrow (\psi \rightarrow \chi)$ and $\psi \rightarrow \varphi$, infer $\varphi \rightarrow \chi$. Given (ID), contraction follows (by letting φ be ψ). (This strengthens the challenges raised by Bacon [2013a] and Field [2014], by avoiding potentially tendentious assumptions about the proper regimentation of relative clauses.)

[7] See Andjelkovic and Williamson [2000].

(M) 'φ' means that φ.

(MT) If S means that φ, then S is true if and only if φ.

Since we reject T, we must reject either M or MT.

We are a bit surprised at the extent to which T is emphasized in the literature on the semantic paradoxes whereas M is relatively neglected. After all, the notion of meaning is widely (although not universally) considered to be the more explanatorily fundamental notion for semantic theorizing.[8] In this paper, we explore the prospects of maintaining M. Although we will ultimately argue that it is untenable, the situation is subtle. Unlike T, there is no way to derive a contradiction from M without further non-logical assumptions. Of course, MT is one such assumption. But, as we will see, there are interesting views about the connection between truth and meaning that invalidate MT and are consistent with M. We will begin by describing two such views. To assess these views, we begin by considering how the meanings of sentences relate to the norms and practices connecting the use of those sentences to facts about extra-linguistic reality. We then argue that M is destabilized by reflection on these connections between meaning and use, and show that our argument extends to variants of M that accommodate context-sensitivity. We conclude by considering some ramifications of our argument: in particular, we suggest that rejecting M is the key to resolving other intensional paradoxes that are not explicitly disquotational.

2. Uniqueness and Instantiation

Let '$M(S, \varphi)$' abbreviate 'S means that φ'. M then becomes the following schema:

(M) $M(\text{`}\varphi\text{'}, \varphi)$

Unlike T, M is consistent considered on its own. However, it is inconsistent with the following two attractive principles:

INSTANTIATION
$\forall p\varphi \rightarrow \varphi[\psi/p]$ —where $\varphi[\psi/p]$ is the result of replacing every free occurrence of p in φ with the sentence ψ

UNIQUENESS
$\forall p \forall q((M(S, p) \wedge M(S, q)) \rightarrow (p \leftrightarrow q))$

Here the bound variable p takes the place of a sentence, unlike variables in first-order languages (which take the place of singular terms).[9] Despite this difference, INSTANTIATION is no less compelling than its more familiar first-order counterpart $\forall x\varphi \rightarrow \varphi[a/x]$ (for a an individual constant and x an individual variable), which is an axiom schema of first-order logic. Note that, given the duality of $\forall$ and $\exists$, INSTANTIATION is equivalent to the perhaps even more intuitively compelling schema $\varphi[\psi/p] \rightarrow \exists p\varphi$. The simplest instances of this schema, for $\varphi = p$, is $\psi \rightarrow \exists pp$: if ψ, then something is the case.[10] Although we will pronounce sentences like $\exists pp$ as 'something is the case', this is merely a gloss, and the sentence that it abbreviates nowhere contains any expression corresponding to 'is the case'.

[8] Here we agree with Soames [1992] against Davidson [1967].

[9] See Prior [1971] for a defence of the intelligibility of this kind of quantification into sentence position. Note that we are not here endorsing Prior's further nominalist position that there are no such abstract objects as propositions, and readers who prefer to theorize in terms of an ontology of propositions can modify the discussion below accordingly (see note 15).

[10] Here and elsewhere, we omit quotation marks where there is no danger of confusion.

To derive a contradiction from M, INSTANTIATION, and UNIQUENESS, consider the sentence $L^* =$ '$\forall p(M(L^*, p) \rightarrow \neg p)$'. Suppose that, as M requires, L^* means that everything it means is not the case. Either (i) everything it means is not the case, or (ii) something it means is the case. It can't be the former, since (i) itself is meant by L^*, and so (by INSTANTIATION) would have to not be the case if it were the case, which is a contradiction. So, there is something L^* means that is the case. And since it means (i), which is not the case, it both means something that is the case and means something that is not the case (by INSTANTIATION), contradicting UNIQUENESS. A formal version of this argument is given below, in a footnote.[11]

This argument is diagnostically valuable, since it shows how M leads to inconsistency, given plausible principles about meaning and generality, without taking a detour through the notion of truth (as the argument from M and MT to a contradiction via T does). It also suggests two different ways of thinking about the semantic profile of L^* that are consistent with the relevant instance of M. By rejecting UNIQUENESS, we can hold that L^* means more than one thing, and the things it means are not all materially equivalent. Or, by rejecting INSTANTIATION, we can hold that there is nothing that L^* means, despite its meaning that everything L^* means is not the case: we cannot existentially generalize on 'everything L^* means is not the case.' (Recall the equivalence of universal instantiation and existential generalization, noted above.)

To gain a feeling for these two views, let us reconsider MT in light of them. We know from the previous section that anyone who accepts M must reject MT. In particular, they must reject the following instance: if L means that L is not true (as M requires), then L is true if and only if L is not true (a contradiction). Although this result does not depend on what we take truth to be, it is instructive to consider the details of how MT fails if we assume the not-unnatural view that to be true is to mean something that is the case. By an argument parallel to that of the previous paragraph, we can conclude from M and INSTANTIATION that L means something that is the case and means something that isn't the case, in which case L is true (despite also meaning something that is not the case). If we instead accept UNIQUENESS and deny that L means anything, then L isn't true (despite meaning that it isn't true).

The view that to be true is to mean something that is the case is not the only possible hypothesis about the connection between meaning and truth. For example, one might instead think that for a sentence to be true is for it to mean something that is the case and not mean anything that isn't the case. This hypothesis leaves the situation unchanged in many respects: given M, INSTANTIATION still allows us to argue that L

[11] M, INSTANTIATION, and UNIQUENESS are classically inconsistent, given the existence of L^*:

1. $M(\text{`}\forall p(M(L^*, p) \rightarrow \neg p)\text{'}, \forall p(M(L^*, p) \rightarrow \neg p))$ [M]
2. $L^* = \text{`}\forall p(M(L^*, p) \rightarrow \neg p)\text{'}$ [fact about the identity of L^*]
3. $M(L^*, \forall p(M(L^*, p) \rightarrow \neg p))$ [1, 2, Leibniz's Law]
4. $\forall p(M(L^*, p) \rightarrow \neg p) \rightarrow (M(L^*, \forall p(M(L^*, p) \rightarrow \neg p)) \rightarrow \neg\forall p(M(L^*, p) \rightarrow \neg p))$ [INSTANTIATION]
5. $\neg\forall p(M(L^*, p) \rightarrow \neg p)$ [3, 4]
6. $\forall p\forall q((M(L^*, p) \wedge M(L^*, q)) \rightarrow (p \leftrightarrow q))$ [UNIQUENESS]
7. $\forall q(M(L^*, \forall p(M(L^*, p) \rightarrow \neg p) \wedge M(L^*, q) \rightarrow (\forall p(M(L^*, p) \rightarrow \neg p) \leftrightarrow q))$ [6, INSTANTIATION]
8. $\forall q(M(L^*, \forall p(M(L^*, p) \rightarrow \neg p)) \wedge \neg\forall p(M(L^*, p) \rightarrow \neg p))$ [3,5, vacuous quantification]
9. $\forall q(M(L^*, q) \rightarrow \neg q)$ [7, 8, deduction under quantifiers]
10. $\perp$ [5, 9, relettering bound variables]

Note again that, while our gloss on this argument used the phrase 'is the case', this is merely an artefact of our way of using English words to pronounce logical notation, since no corresponding expression figures in the above derivation.

both means something that is the case and means something that is not the case, and there is still the alternative of rejecting INSTANTIATION in order to save UNIQUENESS by denying that L means anything (while maintaining that L means that it isn't true). The main change is that L now comes out as untrue whichever strategy is adopted.[12]

Both rejecting UNIQUENESS and rejecting INSTANTIATION have impressive pedigrees. For views that reject UNIQUENESS, see Slater [1986], Read [2002] (who traces the view back to Thomas Bradwardine in the fourteenth century), Restall [2008], and Dorr [2020]. Rejecting INSTANTIATION can take a number of forms. The one that we have been exploring takes its cue from views according to which names of fictional characters obey a positive free logic. Just as, on such views, 'Pegasus' refers to Pegasus even though there isn't anything to which it refers, likewise we can say that '$\forall p(M(L^*, p) \rightarrow \neg p)$' means that $\forall p(M(L^*, p) \rightarrow \neg p)$ even though there isn't anything that it means.[13] One version of this proposal, modelled on Russell's [1908] prohibition on impredicativity, weakens INSTANTIATION to the following principle:[14]

PREDICATIVE INSTANTIATION
$\forall p\varphi \rightarrow \varphi[\psi/p]$, where ψ involves no quantification into sentence position

A very different way of rejecting INSTANTIATION would be to reject the intelligibility of quantification into sentence position. But even those wary of such quantification need some way of generalizing about what sentences mean. The standard way of doing so is to adopt a first-order theory of propositions. Having done so, parallel decision-points arise.[15] Since quantification into sentence position simplifies our discussion without

[12] One might think that those who reject UNIQUENESS face a further question of which (if either) of these two proposals about the connection between meaning and truth is the right one. But, reflecting on the situation, we don't find this to be a productive question. We think that those who reject UNIQUENESS would do best to bar the word 'true' as a predicate of sentences when doing semantic theorizing and instead theorize directly about which sentences mean which things and which of those things are the case. Note that this recommendation concerns only 'true' as a predicate of sentences, and not its use as a predicate of propositions or of speech acts or the use of the sentential operator 'it is true that'.

[13] See Bacon [2013b]. Bacon also holds that there is no such thing as Pegasus ($\neg\exists x(x = \text{Pegasus})$). Allowing ourselves an analogue $\approx$ of identity in sentence position (see Rayo [2013] and Dorr [2016]), proponents of the view under consideration might likewise hold that $\neg\exists p(p \approx \forall p(M(L^*, p) \rightarrow \neg p))$. But this is not the only option. A different model for INSTANTIATION-failures are cases of quantification into opaque contexts. According to the orthodox treatment in Kaplan [1968], for all x and y, if $x = y$, then the ancients knew that x is visible at night if and only if they knew that y is visible at night; yet, while Hesperus = Phosphorus and the ancients knew that Hesperus is visible at night, they didn't know that Phosphorus is visible at night. Universal instantiation fails, despite there being such a thing as Hesperus/Phosphorus. Opacity-induced INSTANTIATION-failures are invoked by Bacon [2021] as a response to the intensional paradoxes discussed in section 7 and by Goodman [2017] as a Fregean response to Russell's paradox of structured propositions.

[14] Note that Russell's view was more complicated, involving a hierarchy of such principles, each corresponding to a different quantifier. Kaplan [1995], Tucker and Thomason [2011], Kripke [2011], and Tucker [2018] express sympathy for some version of this approach; see also Bacon et al. [2016].

[15] In particular, the following three schemas are inconsistent in classical first-order logic enriched with a term-forming functor 'the proposition that … ':

(E) 'φ' expresses the proposition that φ.
(PT) The proposition that φ is true if and only if φ.
(U) If S expresses x and S expresses y, then x is true if and only if y is true.

The argument is as before, with $L^{**} =$ ' $\exists x(L^{**}$expresses $x \wedge x$ is not true)' in place of L^*. Note that, unlike T, PT is perfectly consistent: for example, it holds if propositions are identified with the sets of possible worlds in which they are true. Since, on that view, propositions are not structured like sentences, no analogue of Gödel's diagonal lemma can be used to argue against PT. Moreover, Russell [1903: sec. 500] showed that the claim that propositions are structured like sentences generates paradoxes independently of PT; see Goodman [2017] and references therein for discussion.

loss of generality, transposing our argument into the idiom of a first-order theory of propositions is left as an exercise for the reader who is suspicious of such quantification.

Before moving on, let us briefly compare the INSTANTIATION-rejecting position sketched above, which accepts M and so rejects MT, with the perhaps more familiar position that accepts MT and INSTANTIATION and rejects the L-instance of M on the ground that L doesn't mean anything. The puzzle for the more familiar view is that, according to it, L is not true, yet in stating this fact its proponents utter L itself. Why, by their own lights, are they doing this, if the sentence doesn't mean anything? We don't want to get into the various ways in which proponents of such views might respond to this challenge (for example, by postulating some sort of context-sensitivity in L). Instead, we want to emphasize that proponents of the INSTANTIATION-denying view that accepts M have an answer to the challenge of explaining why they utter L, one that is not available to the M-denying theorist who asserts L: they can say that they utter L to assert that L isn't true, and they succeed because L means that L isn't true (despite there being nothing that it means).[16] The availability of this kind of explanation makes the view especially intriguing.[17]

3. Meaning and Assertion

We are social creatures. We share our knowledge by saying things to one another, and this involves the use of language. Perhaps the most central constraint on semantic theorizing is the following principle, connecting the use of sentences to communication:

(MA) If A (sincerely) utters S and S means that φ on that occasion of use, then A thereby asserts that φ.

(We will reserve 'assert' and its cognates for indirect speech reports—such as in 'John asserted that it was raining', which could be true because John uttered a German sentence—and 'utters' and its cognates for direct speech reports, as in 'John uttered "It was raining"' (which is true only if John uses a particular English sentence), understood broadly to include the use of sentences in writing.)

Since meaning is a natural social phenomenon, MA ought to be our starting point in assessing schemas like M. By contrast, the dominant way of theorizing about semantic

[16] This feature of the view highlights that we are understanding M so that substituting L for φ yields an instance of it even if L doesn't mean anything. (This instance might mean something even if L doesn't.) The view that, for any φ that means something, the corresponding instance of M is true, but the L-instance of M is not true because L doesn't mean anything, is beyond the scope of this paper. Note, however, that weakening M to "if 'φ' means something, then 'φ' is true if and only if φ', combined with MT, is vulnerable to the revenge argument from Bacon [2015]: the theory, if consistent, will prove that some of its theorems don't mean anything.

[17] There is a very different kind of INSTANTIATION-denying response, unlike the one that we have been considering so far, which does not involve uttering L (or accepting any sentence while simultaneously holding that the sentence means nothing that is the case). Assuming (as above) that to be true is to mean something that is the case, this response holds that, in addition to meaning that L isn't true, L means something, and everything that L means is materially equivalent to L being true: $M(L, \neg true(L)) \wedge \exists p M(L, p) \wedge \forall p(M(L, p) \rightarrow (p \leftrightarrow true(L)))$. This is undoubtedly a strange view. Unlike the view discussed in the main text, it cannot be motivated on the ground that 'means that' is opaque (see note 13), since any analogue of identity in sentence position surely implies material equivalence (see Bacon and Russell [2019] and Caie et al. [2020]); so, $\forall p(p \approx \neg true(L) \rightarrow (p \leftrightarrow \neg true(L))$, and hence, according to the view under consideration, $\neg\exists p(p \approx \neg true(L) \wedge M(L, p))$. The view also, arguably, fails to capture the intended force behind the idea of unique meaning, since it requires rejecting the schema STRONG UNIQUENESS: $\forall p_1 \ldots \forall p_n((M(S, \varphi) \wedge M(S, \psi)) \rightarrow \varphi \leftrightarrow \psi)$. It is unclear whether the fact that the proposal allows that every sentence means something compensates for these oddities.

antinomies is to propose disquotational schemas like T and M as if they were candidate *a priori* axioms. To our mind, this dominant methodology is inappropriate to the contingent and empirical subject matter of natural language semantics—a point to which we will return later. For now, it suffices to emphasize that everyone should recognize the centrality of MA to a theory of linguistic meaning, no matter what one thinks about the status of disquotational schemas.[18]

4. Norms on Assertion

One shouldn't speak falsely. How this platitude is best understood is controversial, but it should not be controversial that there is some important sense in which our linguistic behaviour is governed by it. We shall regiment it as follows:

(PA) *A* properly asserts that φ only if φ.[19]

Not much will turn on the exact details of how proper assertion is understood; for present purposes, we can leave the precise normative force of propriety fairly open. We are also not claiming that PA is in any sense 'constitutive' of assertion: for example, perhaps fully proper assertion requires *knowledge* of what one asserts—in which case, PA would follow from the fact that one can know something only if it is the case (see Williamson [2000]).

As discussed above in connection to MA, the primary function of uttering declarative sentences is to make assertions. We can thereby assess utterances for propriety in a way that is parasitic on the propriety of the assertions that they constitute. In particular:

(PU) If, in sincerely uttering *S*, *A* thereby asserts that φ, then *S* is properly uttered only if *A* properly asserts that φ.

Combining MA, PA, and PU, we arrive at the following schema, linking meaning with the propriety of utterances:

PROPRIETY. If *S* means that φ, then *S* is properly uttered only if φ.

This principle is highly plausible on its face, independently of its derivability from MA, PA, and PU. It will be the main premise of our argument against M.

5. Against M

Those who accept M and PROPRIETY are committed to the following disquotational schema governing the propriety of utterances:

(DP) 'φ' is properly uttered only if φ.

[18] This conception of meaning (linking it to what we use sentences to assert) is not uncontroversial: it is denied, for example, by Chomsky [1995] and Pietroski [2017b]. But these authors also deny that sentences' meanings determine truth conditions, and so reject M by rejecting the terms in which it is formulated: see Pietroski [2017a, forthcoming]. Setting aside 'meaning', these authors do not deny the obvious fact that we use sentences to make assertions. In section 8, we argue against disquotational schemas about what we use sentences to assert.

[19] This principle, governing only assertions that are actually made, should not be confused with the principle that it is assertable that φ only if φ. That principle is arguably false: for example, even if no assertions are ever made, that at least one assertion is made may still be assertable, since, had it been asserted, it would have been the case.

But accepting DP puts those theorists in an awkward position.

Consider the instance of DP obtained by replacing both occurrences of φ with the sentence α = 'α is not properly uttered':

(3) 'α is not properly uttered' is properly uttered only if α is not properly uttered.

By Leibniz's law, we obtain

(4) α is properly uttered only if α is not properly uttered,

which is classically equivalent to

(5) α is not properly uttered.

But 5 is α itself. This reveals the combination of PROPRIETY and M (and classical logic) to be untenable.

Here is one way to dramatize the problem. If it were proper to accept both PROPRIETY and M, then it would be proper to utter α, since α is an immediate consequence of those schemas (as demonstrated by the above derivation) and there is nothing improper in drawing out the immediate consequences of things that you properly accept. We have also just shown that if α means that α is not properly uttered, then α is not properly uttered. So, given the acceptability of PROPRIETY (which we will not question), M is not acceptable if α means that α is not properly uttered. Since those sympathetic to M think that if it is acceptable then α means that α is not properly uttered, their combination of commitments is unstable.[20]

A second problem concerns not the acceptability of the schema M, but merely the claim that α means that α is not properly uttered. A consequence of this claim is that α is not properly uttered. This seems like an interesting fact, worth expressing, and one that readers of this paper ought to be able communicate to their friends. But how are they to do so? The first sentence that comes to mind for them to use is 'α is not properly uttered', but, since this sentence is α, it cannot be properly uttered. One could instead utter some distinct but trivially equivalent sentence, such as 'It is not the case that α is properly uttered.' But it doesn't feel as though such a gimmick should be necessary for properly expressing this fact. We can turn this feeling into an argument by noting that substituting 'trivially equivalent to something properly uttered' in the principles invoked above does little to diminish their appeal and would prevent the gimmick from working. Although not a formal inconsistency, this awkwardness is a strong reason to reconsider the view that α means that α is not properly uttered.

6. A Belief Norm

Some philosophers of a more internalist bent will object to factive norms on assertion, such as PA. They will prefer norms couched in terms of what agents (perhaps justifiably) believe, such as the following one:

[20] One could conceivably maintain that M—and hence its instance 'α means that α is not properly uttered'—is acceptable, while, in light of our argument, denying that α means that α is not properly uttered. Such a view would be bizarre, but it is crucial to realize that it is not inconsistent. It could even be reasonably believed by a non-native English speaker who, without knowing what the English word 'means' means, believed on the basis of testimony that M was an acceptable schema. But we will set aside this view for the remainder of the paper, since we don't imagine that anyone will be inclined to maintain it.

(PA_B) A properly asserts that φ only if A believes that φ.

Friends of M might hope that replacing PA with PA_B will block the argument of the previous section. But this hope is mistaken.

By reasoning parallel to the previous section's, it is clear that MA, PA_B, PU, and M classically entail the following:

(DP_B) If A properly utters 'φ', then A believes that φ.

Our argument now turns on the sentence β = 'A does not believe that A properly utters β.' To simplify our argument, we introduce the following abbreviations:

$p :=$ A properly utters β,
$B\varphi :=$ A believes that φ.

Next, consider the instance of DP_B obtained by replacing the schematic sentence letter φ with β:

(6) $p \rightarrow B\neg Bp$.

We now argue as follows:

(7) $B(p \rightarrow B\neg Bp)$
(8) $B(p \rightarrow B\neg Bp) \rightarrow (Bp \rightarrow BB\neg Bp)$
(9) So, $Bp \rightarrow BB\neg Bp$ (from 7, 8)
(10) $BB\neg Bp \rightarrow B\neg Bp$
(11) $B\neg Bp \rightarrow \neg Bp$
(12) So, $Bp \rightarrow \neg Bp$ (from 9, 10, 11)
(13) So, $\neg Bp$ (from 12)

Premises (7), (8), (10), and (11) are justified as follows, on the hypothesis that A is a proponent of M who accepts PA_B. Since they are committed to DP_B, and hence to its instance (6), we may assume that they believe it, which gives us (7). Assuming that their beliefs are closed under *modus ponens* gets us (8). (10) is justified by the thought that such a person needn't be mistaken about what she believes, and (11) by the thought that such a person needn't be mistaken about what she doesn't believe—at least in this instance.

Since (13) is β itself, we have a version of the tension from the previous section. Suppose that A, our internalist proponent of M, is following along with the above argument. In accordance with DP_B, they think that they properly utter β only if they believe that they do not believe that they properly utter β. And, in following along, they end up uttering β. Yet it follows from things they accept that they don't believe that they have properly uttered it. Since they needn't be bizarrely alienated from, or pessimistic about, their own utterances, the natural conclusion is that they are mistaken in having the relevant theoretical commitments.

This completes our argument against M.

7. Intensional Paradoxes

There are a number of so-called 'intensional' paradoxes that, although they have a similar flavour to the semantic paradoxes, depart from them in not explicitly appealing to disquotational schemas or to properties of linguistic expressions. Here is not the place to discuss these paradoxes in detail. But we do want to suggest that, at least in

many cases, the feeling of paradox is rooted in a tacit appeal to a disquotational principle like M. As in the case of the explicitly semantic paradoxes, we want to suggest that the solution is to give up M.

Consider the paradox of someone who asserts that they are asserting something that is not the case. By reasoning parallel to that of section 2, classical logic and INSTANTIATION allow us to conclude that any such person both asserts something that is the case and asserts something that is not the case. That such a conclusion follows by logic alone is shocking indeed, especially since, unlike the semantic paradoxes, there is no disquotational premise used in the derivation.[21] They also have nothing in particular to do with assertion; the argument goes through just as well for what we think, seem to assert, hope, write down, etc.[22]

Why, on reflection, do we find this conclusion so puzzling? After all, we don't think that there is anything puzzling about the impossibility of a Russellian barber, and the derivation of the above theorem involves formally quite similar reasoning. The reason, we think, is that there seem to be *counterexamples*. Clearly, someone who is unaware of their newly owned massive inheritance could reasonably utter the sentence 'The richest person in the building is asserting something that is not the case' in an attempt to complain about their dishonest wealthy roommate. Surely (the thought goes), they would thereby assert that the richest person in the building is asserting something that is not the case, without thereby asserting anything materially inequivalent to that!

But suppose that we reject M. Let us grant that our hypothetical heir asserts whatever 'The richest person in the building is asserting something that is not the case' means. Without M, why think that it means that the richest person in the building is asserting something that is not the case? Just as *L* provides an example of a sentence containing semantic vocabulary that cannot satisfy T, likewise INSTANTATION and the analogue of UNIQUENESS for our protagonist's utterance entail that their utterance is a counterexample to M, presumably owing to its intensional vocabulary.

This diagnosis generalizes. The reason why it strikes us as possible that various things could be uniquely asserted/feared/etc. is usually that we imagine someone uttering or subvocalizing a sentence and judging that it could mean the same as the English sentence that appears in the derivation of the surprising limitative result. But, without M, we will not be able to disquote the sentence in question, and so will not find any inconsistency between the limitative result and the underlying facts of the thought experiment.

A similar line of thought helps to dissolve the air of paradox surrounding the undefinability of truth. Tarski proved that we cannot introduce into any language a predicate that will thereupon be satisfied by all and only the true sentences of that

[21] See Prior [1961]. The argument is also structurally different from the liar paradox in section 1. It does not involve self-reference, and it isn't doing, for what we assert, anything like what Gödel's diagonal lemma does for sentences (namely, generating a 'liar' sentence), since the diagonal lemma relies on the fact that sentences are structured so that it makes sense to substitute expressions for variables. By contrast, Prior's argument does not assume that the things we assert (whether understood in terms of quantification into sentence position, as he does, or in terms of an ontology of propositions) have any structure, and it is consistent with the coarse-grained view that these are individuated modally (cf. note 15).

[22] There are also set theoretic and plural analogues of the paradox. For the former, see Kripke [2011]. On the latter, imagine someone who (non-distributively) thinks about the plurality of all and only the people who are (non-distributively) thinking about a plurality of people that they are not among. It follows, by Priorian reasoning, that any such person is thinking about two pluralities of people, only one of which he is a member of.

language (assuming, that is, that the language has the expressive power to talk about its own sentences). This might feel constraining. What's to stop us? Surely we can have in mind the property of being true in the relevant language; in which case, shouldn't we be able to stipulate that this property be expressed by some new word that we introduce into that language? In reply: what makes us so sure that we can single out in thought, say, the property of being true in German? Perhaps, by using the predicate 'is true in German', we can single out whatever property is expressed in English by 'is true in German'. But, without a disquotational principle for expressing, we cannot conclude that this will be the property of being true in German. We recognize that this thoroughgoing rejection of disquotational reasoning is dizzying at first.[23] But in our view it offers the most satisfying solution, not only to the liar paradox, but also to many other paradoxes about what can be said and thought.

8. Context-Sensitivity

Almost every sentence of natural language is context-sensitive, in the sense of being able, consistently with the conventions of the language, to mean different things on different occasions of use. For example, 'I am hungry' means something about you when uttered by you, but not when uttered by someone else. This shows that the bearers of meaning aren't sentences, but instead are particular utterances of sentences (or perhaps pairs of sentences and contexts). It would clearly be a disaster to try to disquote the utterances of people in other contexts, who mean different things by their words than you do. For example, speeches like 'John's utterance of "I am hungry" means that I am hungry' are not in general true unless the person referred to by 'John' is the one making the speech.

Considerations of context-sensitivity might seem to call into question altogether the appropriateness of disquotational principles. But that reaction is too quick, since, when it is John (the referent of 'John') himself who makes the disquotational speech above, the speech regains its intuitive plausibility. More generally, while disquotational attributions of meanings to utterances are not always acceptable, they are at least *prima facie* acceptable in the special case where the attribution is made in the same context as the utterance itself.[24] In this section, we will consider how our arguments apply to disquotational principles modified to take account of context-sensitivity.

Relativizing to contexts (understood, roughly, as circumstances in which utterances are made) in the way suggested above does nothing to help the disquotationalist about truth, since the sentence L_i = 'L_i is not true in this context' renders the modified schema T_i (below) classically inconsistent, for the same reasons that L renders T inconsistent.

(T_i) 'φ' is true in this context if and only if φ.

[23] Bacon [2018] defends a version of this view, on which instances of M are typically false. By contrast, instances of T can typically be true (since there is no contradiction in the schema: if 'φ' means that ψ, then typically $\varphi \leftrightarrow \psi$), as may instances of '"a" refers to a' (although the Berry paradox suggests that not all of its instances are true).

It might appear that disquotational reasoning is needed to license our practice of using words introduced by stipulative definition as interchangeable with their definitions—for example, in speeches like 'Let a *flagel* be a flat bagel. So, the flagels are all and only the flat bagels.' This would be so if the first sentence is understood as equivalent to the stipulation that 'flagel' expresses the property of being a flat bagel. But, given our rejection of disquotation, we think that it should be understood differently, as a way of using (not mentioning) 'flagel', with 'Let' and emphasis on 'flagel' signalling that we are using a new word to assert truly what can thereafter be asserted, unceremoniously, with 'To be a flagel is to be a flat bagel'; see Dorr [2016: sec. 6].

[24] See Field [2017] for further discussion.

But, just as M is more resistant than T is to paradox, it is natural to wonder about the analogous modification of the M schema:

(M_i) 'φ' means that φ in this context.

Unlike M, this principle doesn't allow us to derive DP from MA, PA, and PU. But it does allow us to derive the following schema:

(DP_i) 'φ' is properly uttered in this context only if φ.

Does the retreat from DP to DP_i afford proponents of M_i a way to resist the analogue of our argument against M?

We think not. For we can still reason as follows, where α_i = 'α_i is not properly uttered in c' and c is the context of the first line of the following derivation. By DP_i, we have

(3_i) 'α_i is not properly uttered in c' is properly uttered in this context only if α_i is not properly uttered in c.

By Leibniz's law (substituting c and α_i), we obtain

(4_i) α_i is properly uttered in c only if α_i is not properly uttered in c,

which is classically equivalent to

(5_i) α_i is not properly uttered in c.

The proponent of M_i has found themselves in the same uncomfortable position as the proponent of M. As discussed in section 5, this can be dramatized in several ways.

5_i is α_i itself, and c is the context in which we just uttered it—namely, the context of the above derivation. Suppose that our imagined proponent of M_i follows along with our derivation above in the same context as us. In so doing, they end up uttering α_i in c, having just derived it from a premise that they accept. But we have just seen that α_i is not properly uttered in c, given the assumption that α_i means in c that α_i is not properly uttered in c (an assumption they will accept). So, if the imagined proponent of M_i is right, then they have done something improper. Therefore, their overall combination of commitments is untenable.

A natural strategy for resisting this argument is to deny that c is the context in which they end up asserting α_i. In starting out in context c, and running through steps 3_i–5_i, at some point we switched to a different context c', in which α_i was properly uttered. This proposal raises a number of questions to which we cannot think of a principled answer (for example, what is the mechanism driving the context shift in these kinds of situation?). And, regardless of questions about mechanism, there are other reasons to think that this manoeuvre is unsatisfactory, as we will now argue.

Any proponent of M_i faces an analogue of the expressive worry from section 5—namely, to properly express the claim, to which they are committed, that α_i is not properly uttered in c. The present proposal is that we have already done this by uttering α_i as part of the above derivation, because that utterance was made in a context c' distinct from c in which it is properly uttered. Note that, in making this response, our theorist is committed to the following:

(i) α_i means in c that α_i is not properly uttered in c.
(ii) α_i means in c' that α_i is not properly uttered in c.
(iii) α_i is properly uttered in c' but not in c.

But, given (i) and (ii), how is it that α_i is not properly uttered in c but is properly uttered in c'? If we would have said the same thing by uttering α_i in either context, and we are just as knowledgeable in both, it is hard to see why we can properly utter it in c' but not c.[25] The most straightforward explanation would be that, by uttering α_i in c we would have asserted additional unassertable propositions that we did not assert by uttering α_i in c'. This proposal is at odds with orthodox contextualist treatments of the liar, since such accounts subscribe to the assumption that sentences express at most one proposition in each context. This shows that the most natural contextualist strategy for maintaining M_i involves rejecting the principle UNIQUENESS from section 2.[26]

Moreover, there is a stronger objection to the contextualist strategy, one that doesn't turn on the question of UNIQUENESS. Consider the following principle:

> If A would not be properly uttered if made in a context c and is a logical consequence of a sentence B, then B would also not be properly uttered if made in c.

This principle is plausible on its face, and it fits with the orthodox conception of contexts, according to which they determine meanings for sentences in such a way that, if A is a logical consequence of B, and A means, relative to a given context, something that is not the case, then B also means, relative to that context, something that is not the case. And it allows us to argue that the instance 3_i of DP_i was not properly uttered in c (that is, in the beginning of the above derivation). For suppose that it was properly uttered in c. Since the relevant identities ('c = this context' and 'α_i = "α_i is not properly uttered in c") clearly could be properly uttered in c, so too could be their conjunction with 3_i, of which α_i is a logical consequence. But, according to the position under consideration, α_i would not have been properly uttered in c.[27] So, by the above principle, 3_i was not properly uttered when made in c. This is a *reductio* of our assumption that 3_i was properly uttered in c. Since 3_i is an instance of DP_i, this undermines the propriety of that schema, and hence of M_i, from which that schema was derivable, given what we claim should be uncontroversial principles connecting meaning and assertion (MA, PA, and PU).

Note that even if one were to deny one of the premises of this argument in order to save the propriety of M_i, doing so would strip that disquotational schema of the theoretical importance that its proponents tend to claim for it. This is because it would confine the proper acceptance of its instances to contexts in which even elementary logical deductions cannot be carried out, and so in which sustained semantic theorizing is not possible.

[25] We are not claiming that, for any reasonable sense of 'properly utterable', if a sentence means the same thing in two contexts then it is properly utterable in both or in neither. For example, notions that take into account issues of politeness won't be like this; but that is not the notion with which we are operating, and, in any case, there is no such contrast between c and c' as regards α_i. Another way that this could happen would be if the sentence expresses a proposition that is true at the time of one context but false at the time of the other context, or that we know at only one of those times; but, again, this is not happening in the case of α_i, c, and c'.

[26] It is less clear how denying INSTANTIATION might afford the contextualist an explanation of the contrast in propriety of α_i between c and c'.

[27] We have been sliding between what was properly uttered and what would be properly uttered in a given context. We don't think that anything of substance turns on this, although, as noted in note 19, there are special cases where it is important to distinguish principles about proper assertion from ones about proper assertability.

Before concluding, we should mention a generalization of this mode of argument aimed at philosophers who are suspicious of the ideology in M_i—namely, of 'contexts' such that we can ask what an arbitrary sentence means relative to a given context. We appealed to that notion above in order to have a concrete way of framing the hypothesis that the inference from 3_i to 5_i might involve a kind of equivocation, so that we could then argue against that hypothesis. But if we are willing to assume that we can perform trivial inferences without equivocating, then our general argument can be reformulated in a way that avoids altogether the notion of meaning and contexts.[28] The reformulated argument targets the following disquotational schema:

(S) If 'φ' is used in this argument, then it is used to say nothing false only if φ.

Let γ = 'If γ is used in this argument, then it is used to say something false.' Now consider the argument:

(3_S) If 'If γ is used in this argument, then it is used to say something false' is used in this argument, then it is used to say nothing false only if, if γ is used in this argument, then it is used to say something false.

(4_S) If γ is used in this argument, then it is used to say nothing false only if, if γ is used in this argument, then it is used to say something false.

(5_S) If γ is used in this argument, then it is used to say something false.

Since 5_S is γ itself, and is used in the above argument, our conclusion implies that it was used to say something false, and hence, by (PU), was not properly uttered. Since it was trivially deduced from 3_S, it follows that 3_S was also not properly uttered. Since 3_S is an instance of S, that schema, too, must be rejected.

9. Conclusion

The situation regarding the disquotational meaning schema M is more subtle than the situation regarding the disquotational truth schema T. Although we think that both should be rejected, and in particular that one should not accept the instances of M involving α, we have not argued that one should accept the negation of that instance of the schema. This is in contrast to the case of T, for which we do accept the negation of one of its instances: we accept the negation of its instance involving the liar sentence, since it is a theorem of classical logic. Nothing that we have said entails the negation of any particular instance of M. Of course, if we were to accept MT, then we would have to reject the instance of M involving the liar sentence. But MT is most clearly motivated by the combination of UNIQUENESS and INSTANTIATION, neither of which was an assumption by our argument against M. By appealing to norms linking meaning and assertion, our argument bypassed those principles—which is a good thing, since in our view UNIQUENESS and INSTANTIATION have not earned the orthodox status that they arguably enjoy.[29]

[28] This argument is aimed both at those like Chomsky and Pietroski, who are sceptical about assigning truth-condition-determining meanings to sentences even relative to contexts, and Dorr [2020], who is suspicious about the ideology of contexts as a framework for theorizing about the flexibility in what sentences can be used to assert, but who does not deny that we can hold fixed the interpretation of our words in the course of a short argument.

[29] Thanks to Cian Dorr, Hartry Field, and Harvey Lederman for comments on earlier versions of this paper.

Disclosure Statement

No potential conflict of interest was reported by the author.

References

Andjelković, Miroslava and Timothy Williamson 2000. Truth, Falsity and Borderline Cases, *Philosophical Topics* 28/1: 211–44.

Bacon, Andrew 2013a. A New Conditional for Naive Truth Theory, *Notre Dame Journal of Formal Logic* 54/1: 87–104.

Bacon, Andrew 2013b. Quantificational Logic and Empty Names, *Philosophers' Imprint* 13/24: 1–21.

Bacon, Andrew 2015. Can the Classical Logician Avoid the Revenge Paradoxes? *The Philosophical Review* 124/3: 299–352.

Bacon, Andrew 2018. Radical Anti-Disquotationalism, *Philosophical Perspectives* 31/1: 41–107.

Bacon, Andrew 2021. Opacity and Paradox, in *Modes of Truth: The Unified Approach to Truth, Modality, and Paradox*, ed. Carlo Nicolai and Johannes Stern, New York: Routledge: 231–65.

Bacon, Andrew and Jeffrey Sanford Russell 2019. The Logic of Opacity, *Philosophy and Phenomenological Research* 99/1: 81–114.

Bacon, Andrew, John Hawthorne, and Gabriel Uzquiano 2016. Higher-Order Free Logic and the Prior-Kaplan Paradox, *Canadian Journal of Philosophy* 46/4-5: 493–541.

Caie, Michael, Jeremy Goodman, and Harvey Lederman 2020. Classical Opacity, *Philosophy and Phenomenological Research* 101/3: 524–66.

Chomsky, Noam 1995. Language and Nature, *Mind* 104/413: 1–61.

Davidson, Donald 1967. Truth and Meaning, *Synthese* 17/3: 304–23.

Dorr, Cian 2016. To Be F Is to Be G, *Philosophical Perspectives* 30/1: 39–134.

Dorr, Cian 2020. Plural Signification and Semantic Plasticity. *Carl Hempel Lectures*, delivered September 2020 at Princeton University.

Field, Hartry 2014. Naive Truth and Restricted Quantification: Saving Truth a Whole Lot Better, *Review of Symbolic Logic* 7/1: 1–45.

Field, Hartry 2017. Egocentric Content, *Noûs* 51/3: 521–46.

Goodman, Jeremy 2017. Reality Is Not Structured. *Analysis* 77/1: 43–53.

Kaplan, David 1968. Quantifying in, *Synthese* 19: 178–214.

Kaplan, David 1995. A Problem in Possible-World Semantics, in *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*, ed. Walter Sinnott-Armstrong, Diana Raffman, and Nicholas Asher, Cambridge: Cambridge University Press: 41–52.

Kripke, Saul A 2011. A Puzzle about Time and Thought, in *Philosophical Troubles: Collected Papers: Vol. 1*, New York: Oxford University Press: 373–9.

Pietroski, Paul 2017a. I-Languages and T-Languages, in *Reflections on the Liar*, ed. Bradley Armour-Garb, New York: Oxford University Press: 141–90.

Pietroski, Paul 2017b. Semantic Internalism, in *The Cambridge Companion to Chomsky*, 2nd edn, ed. James McGilvray, Cambridge: Cambridge University Press: 196–216.

Pietroski, Paul forthcoming. Fostering Liars, *Topoi*.

Priest, Graham 1979. The Logic of Paradox, *Journal of Philosophical Logic* 8/1: 219–41.

Prior, Arthur N. 1961. On a Family of Paradoxes, *Notre Dame Journal of Formal Logic* 2/1: 16–32.

Prior, A.N. 1971. Platonism and Quantification, in his *Objects of Thought*, Oxford: Clarendon Press: 31–47.

Rayo, Agustín 2013. *The Construction of Logical Space*, Oxford: Oxford University Press.

Read, Stephen 2002. The Liar Paradox from John Buridan back to Thomas Bradwardine, *Vivarium* 40/2: 189–218.

Restall, Greg 2008. Modal Models for Bradwardine's Theory of Truth, *Review of Symbolic Logic* 1/2: 225–40.

Bertrand, Russell 1903. *The Principles of Mathematics*, London: W.W. Norton & Co.

Bertrand, Russell 1908. Mathematical Logic as Based on the Theory of Types, *American Journal of Mathematics* 30/3: 222–62.

Slater, B.H. 1986. Prior's Analytics, *Analysis* 46/2: 76–81.

Soames, Scott 1992. Truth, Meaning, and Understanding, *Philosophical Studies* 65/1–2: 17–35.

Tucker, Dustin 2018. Paradoxes and Restricted Quantification: A Non-Hierarchical Approach, *Thought* 7/3: 190–9.
Tucker, Dustin and Richmond H. Thomason 2011. Paradoxes of Intensionality, *Review of Symbolic Logic* 4/3: 394–411.
Williamson, Timothy 2000. *Knowledge and Its Limits*, Oxford: Clarendon Press.