

The Opacity of Truth*

Elia Zardini

LOGOS, Logic, Language and Cognition Research Group

Department of Logic, History and Philosophy of Science

University of Barcelona

Northern Institute of Philosophy

Department of Philosophy

University of Aberdeen

`elia.zardini@ub.edu`

August 15, 2014

1 Naivety and Transparency

Say that a *truth bearer* is a kind of entity that, in some sense, is apt for truth (see Zardini [2014f] for some discussion of this sense). Arguably, there is a variety of kinds of truth bearers, some being *linguistic* (like sentences and utterances) and some being

*This paper continues in the tradition of promoting the forces of darkness that I've initiated with Zardini [2008]; [2012]. Earlier versions of the material in the paper have been presented in 2012 at the LOGOS Semantic Paradoxes and Vagueness Seminar (University of Barcelona); in 2013, at the LOGOS Philosophy of Logic Seminar; in 2014, at a research seminar at the University of Valencia. I'd like to thank all these audiences for very stimulating comments and discussions. Special thanks go to JC Beall, Philipp Blum, Enric Casaban, Hartry Field, Joost Joosten, Nasim Mahoozi, José Martínez, Sergi Oms, David Ripley, Sven Rosenkranz, Gonçalo Santos, Jordi Valor, Timothy Williamson and several anonymous referees. I'm also grateful to the editors Massimiliano Carrara and Julien Murzi for inviting me to contribute to this special issue and for their support and patience throughout the process. At different stages during the writing of the paper, I've benefitted from an AHRC Postdoctoral Research Fellowship and from the FP7 Marie Curie Intra-European Research Fellowship 301493 on *A Non-Contractive Theory of Naive Semantic Properties: Logical Developments and Metaphysical Foundations* (NTNSP), as well as from partial funds from the project CONSOLIDER-INGENIO 2010 CSD2009-00056 of the Spanish Ministry of Science and Innovation on *Philosophy of Perspectival Thoughts and Facts* (PERSP), from the FP7 Marie Curie Initial Training Network 238128 on *Perspectival Thoughts and Facts* (PETAF), from the project FFI2011-25626 of the Spanish Ministry of Science and Innovation on *Reference, Self-Reference and Empirical Data* and from the project FFI2012-35026 of the Spanish Ministry of Economy and Competition on *The Makings of Truth: Nature, Extent, and Applications of Truthmaking*.

non-linguistic (like propositions and beliefs). Neutrally, denote with $\ulcorner\varphi\urcorner$ a truth bearer involving the sentence φ . While, as just stated, the overall view on truth bearers I'm adopting is a *pluralist* one, I'll often abandon neutrality and take as operative truth bearers the kind of truth bearers that is most appropriate to the discussion at hand.

Say that truth is *naive* iff it obeys both the *ascent* principle:

(A) φ entails ' $\ulcorner\varphi\urcorner$ is true'

and the *descent* principle:

(D) ' $\ulcorner\varphi\urcorner$ is true' entails φ .

Say that truth is *transparent* iff it obeys the *intersubstitutability* principle:

(I) If Γ_0, Δ_0 differs from Γ_1, Δ_1 at most in replacing a subformula φ with ' $\ulcorner\varphi\urcorner$ is true', then Γ_0 entails Δ_0 iff Γ_1 entails Δ_1 .

As is well-known from the *semantic paradoxes*, under minimal assumptions (which I'll henceforth make) about the expressiveness of the operative truth bearers, in *classical* logic truth can be neither naive nor transparent. Moreover, henceforth assuming the principle of *reflexivity* for logical consequence (φ entails φ), transparency obviously implies naivety, but the converse implication needn't hold, as it requires more controversial principles that, although plausible and classically valid, are rejected at least by some non-classical logics proposed for solving the semantic paradoxes.¹ While truth is naive if transparent, it may be naive without being transparent.

In the terminology of Zardini [2008], (A), (D) and (I) are *correlation* principles, i.e. principles variously connecting the truth of $\ulcorner P\urcorner$ with its being the case that P . In this respect, transparency, and *a fortiori* naivety, is an appealing *formal* principle about truth, whose apparent force should be equally recognised by very different theories about

¹For example, in the *supervaluationist* theory of truth of McGee [1991], (A) and (D) hold but (I) fails, and so the traditional principle of *intersubstitutability of logical equivalents* fails. Somewhat surprisingly, in many theories in which (I) too holds, this is not because it follows from (A) and (D) plus intersubstitutability of logical equivalents, since the latter principle still fails in those theories. For example, in many theories accepting (I) but rejecting the Liar sentence (see for instance Field [2008]), the latter is logically equivalent with but not intersubstitutable with $0 \neq 0$, while, in many theories accepting (I) and accepting the Liar sentence (see for instance Beall [2009]), the latter is logically equivalent with but not intersubstitutable with $0 = 0$. I think that the envisaged failures of intersubstitutability of logical equivalents raise some serious issues about the notion of logical equivalence (and of entailment) at work in all these theories: if φ and ψ are *not intersubstitutable*, they must *differ in logical strength* (by, for example, φ being in some respect logically stronger than ψ), and shouldn't such a difference be reflected in a *lack of entailment* from ψ to φ ? It is thus important to note that acceptance of (I) does *not* necessarily imply the implausible rejection of intersubstitutability of logical equivalents (see Zardini [2011]; [2013a]; [2013b]; [2014a]; [2014c]; [2014d]; [2014g] for an example of a consistent theory accepting (I) as well as intersubstitutability of logical equivalents).

the *nature* of truth, such as, for example, a broadly *correspondentist* theory and a broadly *deflationist* theory. As for correspondentism, transparency can naturally be understood as saying that the truth of $\ulcorner P \urcorner$ is in a very strong sense equivalent with its being the case that P , and that is something at least to some extent suggested by the correspondentist identification of the truth of $\ulcorner P \urcorner$ with $\ulcorner P \urcorner$'s *corresponding with the relevant facts*. As for deflationism, something like transparency seems to be directly required if truth is adequately to fulfil the *expressive functions* that, according to the deflationist, constitute its *raison d'être*.

In fact, an important strand in the contemporary debate on the semantic paradoxes, which I'll call '*quasi-deflationism*', has taken on board precisely the idea that what truth must do is, possibly among other things, fulfil the expressive functions identified by the deflationist, and has understood that idea in such a way as to make it serve as a basis for *an argument against non-naive or non-transparent theories of truth*, claiming that such features of these theories are incompatible with truth fulfilling said functions. Notice that *a quasi-deflationist need not be a deflationist*: she only needs to maintain that truth must have the *formal properties* required to fulfil the expressive functions she emphasises along with the deflationist, but, contrary to the deflationist, she need not deny that truth also has an *underlying nature*, and she need not even deny that it is actually *in virtue of* having such nature that truth fulfils the expressive functions she emphasises. Having noted that, obviously a deflationist must be a quasi-deflationist, and, arguably, from a non-deflationist perspective quasi-deflationism loses some of its appeal, at least to the extent that, from that perspective, the supposition that truth does not fulfil the relevant expressive functions does no longer entail that truth is deprived of its very *raison d'être*.²

Plausible as they may seem from a wide variety of perspectives, transparency, and even naivety, are however not without their problems. In the rest of this paper, I'll go through some of the main kinds of cases in which those principles arguably fail. Most of these kinds of cases are already known in some areas of the philosophy of *language* (David [2005] contains a state-of-the-art list and a perceptive discussion of some of its items); the main contribution of the paper is to connect these failures with the contemporary debate in the philosophy of *logic* on the semantic paradoxes, and, in particular, to argue that, *once such failures are taken into account, the quasi-deflationist argument against non-naive or non-transparent theories of truth is robbed of all its force*. More specifically, I'll exemplify and discuss this point in section 2 with respect to truth-value gaps; in section 3, with respect to truth-bearer contingency; in section 4, with respect to truth-condition context dependence. I'll close in section 5 by arguing that, although the quasi-deflationist argument fails, other truth-theoretic principles that are not correlation principles do force revision of classical logic.

²The most developed version of quasi-deflationism is Field [2008], on whose claims I'll thus mostly focus (Field is of course also one of the most prominent advocates of deflationism). Another important position favourable to quasi-deflationism is the one represented by Beall [2009]. Although this paper is largely devoted to a criticism of the quasi-deflationist argument against non-naive or non-transparent theories of truth, I cannot emphasise enough how much my own thinking about truth has been shaped by reading and discussing with these two authors.

2 Truth-Value Gaps

Let's assume throughout this section that the operative truth bearers are *sentences*, and let's start with a kind of failure of transparency that need not be also a failure of naivety (but which, as we'll variously see, might be, and might in particular be a kind of failure of (A)). Say that $\lceil \varphi \rceil$ is *false* iff $\lceil \text{It is not the case that } \varphi \rceil$ is true. Then, some sentences are plausibly *neither true nor false* (and, in this sense, constitute a *truth-value gap*). The literature is replete with candidate types of gappy sentences: sentences about *borderline cases* ('Harry is bald'), sentences suffering from *semantic indecision* ('Newtonian mass is relativistic mass'), *undecidable* sentences ($\aleph_1 = 2^{\aleph_0}$), sentences suffering from *reference failure* ('Santa Claus doesn't exist'), sentences about *the future* ('There will be a sea-battle on 07/10/2171'), *conditional* sentences ('If Italy played the final of the 2010 World Cup, they won'), *evaluative* sentences ('Abortion is wrong') etc. It is certainly not my task here to argue for any of these views; let's work however under the plausible assumption that at least some of the candidates are indeed types of sentences that are gappy (for concreteness, I'll take sentences about borderline cases and sentences suffering from reference failure to be such types).

Let me emphasise that, although *plausible*, as will be shown in the next paragraph the assumption *cannot but be denied* if it is assumed that (I) unrestrictedly holds (as acknowledged for example by Field [1994b]). I do find it problematic to let an *extremely controversial* thesis in the *philosophy of logic* dictate the rejection of a wide array of *independently plausible* views in the *philosophy of language* and *beyond* (especially if the support for the thesis in turn wholly consists of considerations of the tenor of those concerning the expressive functions of truth),³ but I won't delve into this. For my aim in this section is not to refute the *extreme* position that *assumes that (I) unrestrictedly holds* and that *on these grounds rejects the claim that at least some types of sentences are gappy*; it is rather *to undermine the quasi-deflationist argument*, an argument which, on a more *moderate* position, *might have been thought to retain its essential force even if it is acknowledged that some types of sentences may be gappy*, via the postulation that *paradoxical sentences are not tokens of such types* (so that the argument might have been thought to proceed by establishing that (I) holds at least as *restricted* to paradoxical sentences). It is precisely such a moderate—and *prima facie* much more attractive—position that I'll argue to be untenable.

Suppose then that Harry is borderline bald, so that 'Harry is bald' is gappy. Then, neither 'Harry is bald' nor 'Harry is not bald' is true. But, by (I), if 'Harry is bald' is not true, then Harry is not bald, and so, again by (I), 'Harry is not bald' is true. Contradiction. Thus, (I) fails (essentially, the point goes back at least as far as Dummett

³To take a somewhat extreme example, one might argue that neither 'EZ will lift his left arm' nor 'It is not the case that EZ will lift his left arm' are true (that is, true *now*) because, if either were, its present truth would rob me of the *freedom*, which I apparently have, to lift my left arm. I think this argument deserves much deeper considerations than those concerning the expressive functions of truth. As I'm mostly focussing on Field's works, I should add that, in his specific case, those considerations are indeed grounded in a more fundamental *deflationist* outlook, which is in turn motivated by a deeply rooted *scepticism about semantic properties*.

[1959], pp. 145–146).^{4,5}

One expressive function for truth to fulfil emphasised by quasi-deflationists is that of

⁴Notice that, while the problem cannot be generated by relying purely on naivety, it can be generated without relying on the full power of transparency. For what is really needed to generate the problem is the principle of *contraposed ascent*:

(CA) ‘ $\ulcorner \varphi \urcorner$ is not true’ entails ‘It is not the case that φ ’,

which would do the work done by the first application of (I) in the argument in the text, whereas the second such application is immediately licenced by (A). (Notice that, just as truth-value *gaps* make problems for (CA), truth-value *gluts* may be thought to make problems for its converse, as argued by Priest [2006], pp. 78–80.) Of course, given the property of *contraposition for logical consequence* (if φ entails ψ , ‘It is not the case that ψ ’ entails ‘It is not the case that φ ’), naivety (in particular, (A)) implies (CA). In turn, contraposition for logical consequence is arguably a valid principle even in the presence of the semantic paradoxes. I’ve argued this much already in Zardini [2014c], but I’d now wish to provide a further argument relying on *intersubstitutability of logical equivalents* (which I’ve in turn defended in fn 1). Suppose that φ entails ψ . Then, for at least some broadly conjunction-like connective \wedge , φ is equivalent with $\varphi \wedge \psi$. But, by anyone’s lights, ‘It is not the case that ψ ’ entails ‘It is not the case that $\varphi \wedge \psi$ ’, and so, by intersubstitutability of logical equivalents, ‘It is not the case that ψ ’ entails ‘It is not the case that φ ’. Having noted that, contraposition for logical consequence is unfortunately not acceptable in many non-classical theories of truth (although it unsurprisingly is in the theory developed in my works referenced in fn 1), and so I’ll continue to keep (CA) separate from (A).

⁵Beall [2002] argues that the contradiction is only apparent as one can understand the negation connected with falsity as *choice* negation (henceforth subscripted with ‘ch’) and the negation denying truth and falsity as *exclusion* negation (henceforth subscripted with ‘exc’). I find Beall’s proposal very interesting, but also multiply problematic. Firstly, some candidate types of gappy sentences are such that, if indeed gappy, they would *also include gappy sentences governed by negation_{exc}* (for example, there are borderline cases of non_{exc}-baldness), and, for such cases, Beall’s proposal is equally affected by the argument in the text (given that ‘It is not_{ch} the case that it is not_{exc} the case that φ ’ is intersubstitutable with ‘It is not_{exc} the case that it is not_{exc} the case that φ ’). Secondly, I’ll discuss below in the text how the kinds of truth-value gaps exemplified by certain sentences are *compatible with acceptance* of those sentences (for example, someone thinking that ‘Santa Claus doesn’t exist’ is not true because it suffers from reference failure may well accept that sentence as reporting the non-existence of Santa Claus). But the argument in the text can still infer from the gappiness of any such sentence φ to ‘It is not_{exc} the case that φ ’, and so one would still be stuck with accepting a contradiction (‘ φ and it is not_{exc} the case that φ ’), just as the original argument in the text would have it. Thirdly, I take it that, on a typical truth-value-gap approach, it is supposed to be *false* (rather than *neither_{exc} true nor_{exc} false*) that φ is both gappy and either true or false. (After all, every sentence is supposed to have *exactly one* status, and gappiness is supposed to be a *third* status: it follows from the first assumption that, if φ is gappy and true (false), gappiness is identical with truth (falsity), and so gappiness and truth are one and so not_{ch} two, which is false given the second assumption.) But, if one accepts this, and accepts that φ is gappy, one should presumably infer that it is false that φ is either true or false—that is, infer that φ is neither_{ch} true nor_{ch} false, which reinstates the original argument in the text. Fourthly, Beall’s proposal is *variously at odds with typical arguments in favour of truth-value gaps*. On the one hand, such arguments proceed by identifying a *necessary condition for truth* (for example, that speakers’ practices determine a sufficient condition for ‘bald’ to apply to Harry and Harry satisfies the condition) and contending that *this is not_{ch} met by the relevant type of sentences*: if at all sound, such arguments establish, by *contraposition for implication* (‘If φ , then ψ ’ entails ‘If it is not_{ch} (not_{exc}) the case that ψ , then it is not_{ch} (not_{exc}) the case that φ ’), that the relevant type of sentences are not_{ch} true over and above being not_{exc} true. On the other hand, such arguments proceed by identifying a *broadly semantic feature of the relevant type*

*expressing rejection of a theory*⁶ (see for example Field [2008], pp. 138–139).⁷ If T contains *infinitely many* sentences, or if one *does not know* exactly which sentences belong to T , quasi-deflationists claim that one can still reject T by accepting ‘Some member of T is not true’. Before applying the point of the last paragraph to this claim, it’ll be helpful to see a couple of aspects in which the claim, as made in our dialectical context, is independently puzzling.

Firstly, just as many theorists rest content with the intuitive notion of rejecting a *single sentence* φ without requiring it to be grounded in acceptance of ‘ φ is not true’ or in anything else, it would seem equally legitimate to rest content with the intuitive notion of rejecting φ and ψ *taken together*—that is, rejecting $\{\varphi, \psi\}$ —without requiring it to be grounded in acceptance of ‘It is not the case that both ‘ φ ’ and ‘ ψ ’ are true’ or in anything else.⁸ But, if the notion of rejecting a set of *two* sentences is legitimate, it would seem that the notion of rejecting a set of sentences of *arbitrary cardinality* is also legitimate, and that it is so even if the rejecting subject does not know exactly which sentences belong

of sentences (for example, suffering from reference failure) and contending that *such feature is the root of the gappiness of those sentences*. But the feature in question is typically *not preserved under ascent* (for example, while ‘Santa Claus doesn’t exist’ suffers from reference failure, ‘‘Santa Claus doesn’t exist’ is true’ does not), and so such arguments provide no support for thinking that gappiness is preserved under ascent, contrary to what is required by Beall’s proposal. Thanks to an anonymous referee for recommending a discussion of Beall’s proposal.

⁶Throughout, an attitude *expresses* (or *is associated with*) an attitude iff they have the same correctness conditions. Thus, as I understand them, the expressive functions of truth concern its enabling one virtually to entertain a problematic attitude by non-virtually entertaining another, less problematic attitude, rather than its enabling one to voice one’s attitudes in conversation: as I understand them, the expressive functions of truth are functions of *thought* rather than *communication*. Thanks to an anonymous referee for urging these clarifications.

⁷Field [2008], pp. 138–139 mainly uses ‘*disagree*’ and its relatives, but, focussing purely for simplicity on the case of a single sentence, I take it that, in the relevant sense of disagreeing with a *sentence* φ (rather than with a *person*), what he means is *rejection* of φ (which in turn in this debate is understood as a fairly *sui generis* attitude consisting, roughly, in *ruling out the truth of* φ). A possibly alternative interpretation would be that what he means is *acceptance* of ‘*It is not the case that* φ ’. But, in a framework (such as Field’s) in which these two are indeed alternative interpretations (as I explain below in the text), the latter is arguably a very *unnatural* and rather *uninteresting* understanding of disagreement, since, under such understanding, one should not disagree with ‘Harry is bald’ if Harry is borderline bald. Indeed, Field [2008], p. 208 himself remarks that it is more natural to understand disagreement as rejection. For this reason, I assume throughout such understanding. Thanks to an anonymous referee for pointing out the need for this clarification.

⁸One might object that that is not equally legitimate, since, while acceptance of $\{\varphi, \psi\}$ can be *reduced to acceptances of individual sentences* (namely, to acceptance of φ and acceptance of ψ), rejection of $\{\varphi, \psi\}$ cannot be *reduced to rejections of individual sentences* (namely, to rejection of φ or rejection of ψ). But, even granting for the sake of argument such a dubious reducibility constraint on legitimacy, in my view the objection crucially overlooks the fact that, as I explain in fn 15, there are *two different kinds* of acceptance and rejection, and that the *weaker* kind of acceptance—which is the one that *is* reducible to acceptances of individual sentences—has as corresponding *exclusive* kind of rejection the stronger one that *is* reducible to (conditional) rejections of individual sentences (in fact, just as the weaker kind of acceptance is reducible to a *sum* of acceptances of individual sentences, the stronger kind of rejection is reducible to a *sum* of (conditional) rejections of individual sentences). Thanks to Sergi Oms for discussion of this point.

to that set.⁹ If so, even if T contains infinitely many sentences, and even if one does not know exactly which sentences belong to T , one can still reject T by... simply rejecting T rather than by doing anything else!

Admittedly, there is some attractiveness in grounding (or at least associating) rejection of a theory or even of a single sentence in acceptance of a sentence; indeed, there is some attractiveness in grounding (or at least associating) acceptance of a theory in acceptance of a sentence. In a less interesting respect, that is *conversationally* attractive, as it allows one to *convey* much more simply one's states of mind (see Field [2008], p. 96 for a vivid discussion of how cumbersome it is to convey rejection). In a more interesting respect, that is *theoretically* attractive, as the states of mind in question are supposed to *represent ways things are*, and it is extremely plausible that, for every way things could be, there is a proposition corresponding to it—a proposition that is at least expressed in some context by some sentence of some language, so that every such state of mind is expressed by acceptance of a certain sentence.

It is the latter, more interesting respect of theoretical attractiveness that I think constitutes a compelling rationale for accepting the requirement that rejection of a theory or even of a single sentence, and indeed acceptance of a theory, be at least associated with acceptance of a sentence. Notice however that that rationale is actually incompatible with the theories of truth that are typically supposed to be eventually justified by considerations concerning the expressive functions of truth. For such theories are transparent theories. Now, taking on board the rationale just mentioned, for every φ there is some (possibly metalinguistic) context $\mathcal{B}_\varphi(\dots\dots)$ such that rejection of φ is expressed by $\mathcal{B}_\varphi(\dots\varphi\dots)$. Moreover, $\mathcal{B}_\varphi(\dots\varphi\dots)$ both *is inconsistent with φ* and *follows from φ 's inconsistency* (since rejection of φ rules out φ and is licenced by φ 's inconsistency). But no standard (i.e., for the cognoscenti, *non-substructural*) transparent theory can allow for a notion with those two features (see Zardini [2014c]; [2014d] for more details).¹⁰

Secondly, ironically (but unsurprisingly given the point made in the last paragraph), the theories of truth that are typically supposed to be eventually justified by considerations concerning the expressive functions of truth are theories in which, even taking the case of

⁹It seems to me clear that, quite generally, just as one can accept or reject, say, gifts without knowing exactly which gifts they are, one can straightforwardly accept or reject sentences without knowing exactly which sentences they are (rather than simply expressing such acceptance or rejection by using, for example, truth). For one thing, were this not so virtually no one could straightforwardly accept all the terms and conditions of many contracts, or the whole periodic table, or everything said by any past or future Pope etc. (Notice that, once we naturally extend this point from sentences to propositions, a casualty of the resulting view is arguably the implication from *accepting the proposition that P* to *believing that P* .) However, since the quasi-deflationist framework is not particularly hospitable to such considerations, I'll henceforth set them aside. Thanks to Sergi Oms for raising this issue.

¹⁰An analogous point could have been made about *conditional acceptance*. Taking on board the rationale just mentioned in the text, for every φ and ψ there is some (possibly metalinguistic) context $\mathcal{C}_\varphi(\dots\dots\dots)$ such that conditional acceptance of ψ on φ is expressed by $\mathcal{C}_\varphi(\dots\varphi\dots\psi\dots)$. Moreover, $\mathcal{C}_\varphi(\dots\varphi\dots\psi\dots)$ both *together with φ entails ψ* and *follows from φ 's entailing ψ* (since conditional acceptance of ψ on φ forces ψ given φ and is licenced by φ 's entailing ψ). But no standard (i.e., for the cognoscenti, *non-substructural*) transparent theory can allow for a notion with those two features (see Zardini [2013b]; [2014a]; [2014c]; [2014d] for more details).

a single sentence φ , one cannot always reject φ by simply accepting ‘ φ is not true’. Many such theories *reject* both the Liar sentence (a sentence that says of itself that it is not true) and its negation (since they think that both are inconsistent). But then, if, for example, T merely consists of the Liar sentence, these theories would reject T but would nevertheless not accept ‘Some member of T is not true’, for the latter is under the circumstances tantamount to ‘The Liar sentence is not true’, which is in turn tantamount to the Liar sentence, which the theories in question, far from accepting, reject. In such theories, then, attributions of untruth do not track rejection in that *many sentences are rejected although it is not accepted—worse, it is in fact rejected—that they are not true*. Some other of the theories that are typically supposed to be eventually justified by considerations concerning the expressive functions of truth *accept* both the Liar sentence and its negation (since they think that both are logically true). But then if, for example, T merely consists of the Liar sentence, these theories would not reject T but would nevertheless accept ‘Some member of T is not true’, for the latter is under the circumstances tantamount to ‘The Liar sentence is not true’, which is in turn tantamount to the Liar sentence, which the theories in question, far from rejecting, accept. In such theories, then, attributions of untruth do not track rejection in that *many sentences are accepted not to be true although they are not rejected—worse, they are in fact accepted*. Thus, quite generally, the theories that are typically supposed to be eventually justified by considerations concerning the expressive functions of truth cannot uphold that attributions of untruth are always correlated with rejection.¹¹

To come back to our main thread, quasi-deflationists usually claim that fulfilment of the function of expressing rejection requires truth to satisfy (CA) (see for example Field [2008], p. 205, and Field [2008], pp. 208–209 for Field’s own cursory doubt about this claim). However, certain types of gappy sentences provide counterexamples both to (CA) (as we’ve seen) and to the quasi-deflationist idea that truth only fulfils the function of expressing rejection if (CA) holds (as we’ll now see). If T merely consists of ‘Harry is bald’, one should accept ‘Some member of T is not true’ and, although one need not thereby accept ‘Harry is not bald’ (and one should actually reject it), given what sentence ‘Harry is bald’ is one would indeed thereby reject it. Say that a kind of truth-value gap is *malign* iff, if a sentence exemplifies that kind, one should reject that sentence, and say that a kind of truth-value gap is *benign* otherwise. Thus, for a *malign* kind of truth-value gap such as the one exemplified by sentences about borderline cases, *attributions of untruth express rejection even if (CA) fails*. Notice however that the *way* in which the expression in question *works* is very different from the (CA)-based one imagined by quasi-deflationists. For, in our example, since (CA) fails truth cannot function as a *formal device*, with one attributing untruth as a means to the end of committing oneself to accepting a *common-or-garden-variety truth-free* sentence like ‘Harry is not bald’; rather, truth functions as a *substantial category*, with one attributing untruth as an end in itself, the end of classifying

¹¹Keeping fixed that, at least in the case of the semantic paradoxes, attributions of untruth are always correlated with rejection, since the Liar sentence is tantamount to ‘The Liar sentence is not true’ it surprisingly but straightforwardly follows, under plausible assumptions about acceptance and rejection, that *one should neither accept nor reject the Liar sentence*. While in conflict with many theories of truth, that result is very much congenial to the theory I’ve developed in my works referenced in fn 1.

the sentence ‘Harry is bald’ as exhibiting a certain *feature* which, given what sentence that sentence is, must be regarded as a *defect* that in turn warrants its rejection (just as Harry’s not being bald would more straightforwardly warrant the rejection of ‘Harry is bald’).

These facts have an immediate relevance for the contemporary debate about the semantic paradoxes, for in that debate quasi-deflationists object to certain theories of truth on the grounds that, by rejecting (CA), they would prevent truth from fulfilling the function of expressing rejection. But, as we’ve just seen, there is a variety of reasons, independent of the semantic paradoxes, for rejecting (CA) in its full generality, and indeed for rejecting the objection’s presupposition that truth only fulfils the function of expressing rejection if (CA) holds: sometimes, one may accept ‘ φ is not true’ and thereby reject φ without thereby accepting ‘It is not the case that φ ’.

One might claim that these reasons concern *special* cases, in such a way as to preserve the essential force of the quasi-deflationist argument, which concerns instead the workings of truth in *normal* cases. Set aside that, given the centrality of the considered cases to ordinary and scientific thought and talk, it is actually doubtful whether there is a useful sense in which they are special. The claim is anyways self-defeating, since all parties agree that, in *normal* cases (involving innocent sentences like ‘ $2+2=5$ ’), (CA) holds and is a route *via* which attributions of untruth express rejection. It cannot be emphasised enough that the punch of the quasi-deflationist argument *only comes from insisting that this should continue to be so also in every special case*, from which it follows in particular that it should continue to be so also in the special case involving paradoxical sentences (for which in effect certain theories of truth reject the relevant instances of (CA)). *It is only through this insistence* on (CA) being a route *via* which attributions of untruth express rejection *of any theory whatsoever*, normal or special, that the quasi-deflationist argument has any force at all. But it is *precisely such blind insistence that is belied* by the counterexample offered, which shows that it is not the case that (CA) is a route *via* which attributions of untruth express rejection of any theory whatsoever. And it would seem completely *arbitrary* to insist only that (CA) be a route *via* which attributions of untruth express rejection in the special case involving paradoxical sentences, while allowing that that may no longer be so in the special case involving sentences about borderline cases. Thus, given its very structure, the quasi-deflationist argument has set the odds heavily against itself: it just takes *one single case* in which (CA) is not a route *via* which attributions of untruth express rejection to show that the argument is bankrupt—the issue whether such case is special, normal, related to the semantic paradoxes or what have you is neither here nor there.

There is worse news for quasi-deflationists. For other types of gappy sentences not only provide counterexamples to (CA), they also provide counterexamples to the quasi-deflationist idea that attributions of untruth are always correlated with rejection (an idea which we’ve seen in the fourth last paragraph to be actually already contradicted by standard non-classical theories of truth). If T merely consists of ‘Santa Claus doesn’t exist’, one should accept ‘Some member of T is not true’ although, given what sentence ‘Santa Claus doesn’t exist’ is, one need not thereby reject ‘Santa Claus doesn’t exist’. One

should actually accept it, lest one be barred from *expressing* at the level of acceptance of sentences the *non-metalinguistic* belief that Santa Claus doesn't exist—indeed, since, at least for beings with a certain degree of cognitive sophistication, belief does arguably require being in a relation of acceptance with a corresponding representational vehicle (see also fn 31), lest one be barred from *having* the belief that Santa Claus doesn't exist in the first place (one could still have, and express at the level of acceptance, the *metalinguistic* belief that 'Santa Claus' lacks a referent, but there are many well-known good reasons for thinking that the former belief is not equivalent with the latter). Thus, for a *benign* kind of truth-value gap such as the one exemplified by certain sentences suffering from reference failure,¹² *attributions of untruth do not express rejection and are actually compatible with acceptance.*¹³

¹²'Certain' because, arguably, the kind of truth-value gap exemplified by many other sentences suffering from reference failure (for example, 'Santa Claus is 1.76 cm tall') is rather of the malign variety. It is beyond the scope of this paper to offer an account of the distinction between malign and benign kinds of truth-value gaps, although it may at least be noted that it seems to correlate with the presence or absence of what may to a first approximation be described as *indeterminacy in principle*.

¹³At least given the assumption justified in fn 7, Field [2008], pp. 206–208 disagrees, in effect claiming, about another example equally supposed for the sake of argument to be gappy, that one should reject the sentence in question. As I've briefly argued in the text, such claim is very problematic. In any event, the fact that attributions of untruth are not always correlated with rejection is probably even more clear-cut with other candidate types of gappy sentences: someone thinking on 06/10/2171 that 'There will be a sea-battle on 07/10/2171' is not true because it is not completely settled on 06/10/2171 may well nevertheless accept that sentence as describing what one has overwhelming reasons for thinking will be the case (or at the very least may on those grounds not reject that sentence); someone thinking that 'If Italy played the final of the 2010 World Cup, they won' is not true because it is only endowed with acceptability conditions relative to epistemic states, may well nevertheless accept that sentence as having a consequent that, on one's evidence, is extremely likely given its antecedent (or at the very least may on those grounds not reject that sentence); someone thinking that 'Abortion is wrong' is not true because it expresses an evaluation may well accept that sentence as expressing one's own evaluation (or at the very least may on those grounds not reject that sentence). To take philosophically somewhat less loaded cases, someone thinking that 'If P , then P ' is not true because it contains schematic expressions with no specific content may well accept that sentence as stating the valid law of reflexivity for implication (or at the very least may on those grounds not reject that sentence); someone thinking that 'It might be black' (understood as anaphoric on 'A sheep might come in') is not true because it contains a pronoun with no specific referent may well accept that sentence (under the relevant understanding) as contributing a further specification to an overall correct discourse (or at the very least may on those grounds not reject that sentence). (Since schematic or anaphoric sentences plausibly do not express propositions, there arguably are no beliefs corresponding to their acceptance, thus providing another kind of example, in addition to that discussed in fn 9, of failure of implication from acceptance to belief.) Notice that Field does actually agree that attributions of untruth are not always correlated with rejection, as with non-indicative sentences (and in many other more degenerated cases as well). As I understand it, the idea is that one can still express rejection by a suitably *strengthened* attribution of untruth along the lines of '‘ φ ' is meaningful, indicative and not true'. It is actually very unclear whether something along these lines can by itself adequately deal with the problem posed by *non-indicative* sentences in all its respects, for it would seem that, for example, one can accept *imperative* sentences just as well as *indicative* ones, but acceptance of imperative φ cannot be expressed by either '‘ φ ' is true' (as that is *not true* of *any* imperative sentence) or '‘ φ ' is not true because non-indicative' (as that is *true* of *every* imperative sentence). (Since imperative sentences plausibly do not express propositions, there arguably are no beliefs corresponding to their acceptance, thus providing yet another kind of example, in addition to those discussed in fn 9 and above in this fn, of failure of implication from acceptance to belief. With respect to the last two kinds of examples, one may

Notice that this is compatible with the claim I’ve made in the fourth last paragraph to the effect that untruth of ‘Harry is bald’, “given what sentence that sentence is, must be regarded as a *defect* that in turn warrants its rejection”, for there should be no presumption that truth is an *equally important desideratum for every type of sentence*. Quite the contrary, it is natural to expect that whether sentences *of a certain type* are true or not might affect their acceptability and rejectability in a way in which it does not for sentences *of another type* (just as whether mathematical sentences are in principle provable or not plausibly affects their acceptability and rejectability in a way in which it does not for empirical sentences, so that, for example, by attributing in-principle unprovability to the Continuum Hypothesis one would plausibly thereby reject it, but, by attributing in-principle unprovability to ‘There are dogs’, one need not thereby reject it).¹⁴

These facts have an immediate relevance for the contemporary debate about the semantic paradoxes, for in that debate quasi-deflationists object to certain theories of truth on the grounds that, by rejecting (CA), they would prevent truth from fulfilling the function of expressing rejection. But, as we’ve just seen, there is a variety of reasons, independent of the semantic paradoxes, for rejecting (CA) in its full generality, and indeed for rejecting the objection’s presupposition that truth always fulfils the function of expressing rejection (a point analogous to that made in the third last paragraph applies to the claim that such reasons concern special cases): sometimes, one may accept ‘ φ is not true’ without thereby rejecting φ .

There is more bad news for quasi-deflationists. One expressive function for truth to fulfil emphasised by quasi-deflationists is that of *expressing acceptance of a theory* (see for example Field [2008], pp. 138–139). If T contains *infinitely many* sentences, or if one *does not know* exactly which sentences belong to T , quasi-deflationists claim that one can still accept T by accepting ‘Every member of T is true’.¹⁵

well suspect that the quasi-deflationist argument *illicitly conflates* the *general* kind of acceptance that is at issue in considerations concerning the expressive functions of truth with a *specific* kind of acceptance for whose instances there are corresponding beliefs.) Setting aside (as I generally mean to do in this paper) the problem of non-indicative sentences, what I’d wish to stress here is rather that one cannot *further strengthen* the attribution of untruth along the lines of, say, ‘ φ is meaningful, indicative, not about the future and not true’, for that is not always correlated with rejection of φ now in the other direction (for example, if T consists of ‘There will be peace talks on 07/10/2171’, one should never accept and should actually always reject ‘Some members of T are meaningful, indicative, not about the future and not true’ although, on 06/10/2171, one may reject T as misdescribing what one has overwhelming reasons for thinking will be the case). Thanks to Enric Casaban, José Martínez, Sergi Oms, Gonçalo Santos and Jordi Valor for discussions of some of these issues.

¹⁴In fact, even setting aside the important issue represented by “mixed sentences”, as noted more accurately in fn 12 whether an untrue sentence is acceptable will sometimes be a question *with a finer mesh* than what type that sentence is a token of. Thanks to an anonymous referee for raising the issues discussed in this paragraph.

¹⁵In fact, by the quasi-deflationist’s own lights, by doing so one necessarily accepts *as a whole* the sentences belonging to T over and above accepting *individually each* of them (as Field [2008], p. 139 himself notes). That there is an important difference between these *two kinds of acceptance* is brought out by the *preface paradox* (introduced by Makinson [1965]). But that means that, by the quasi-deflationist’s own lights, truth *only* expresses the former, *stronger* kind of acceptance and not the latter, *weaker* kind. However, in many contexts, the only relevant kind of accepting T is arguably the weaker one (when there

Quasi-deflationists usually claim that fulfilment of the function of expressing acceptance requires truth to satisfy (A) (see for example Field [2008], pp. 139–141). However, as was already implicit in the discussion of the fourth last paragraph, certain benign kinds of truth-value gaps provide counterexamples both to (A) and to the quasi-deflationist idea that acceptance is always expressed by attributions of truth (as we’ll now explicitly see). Santa Claus doesn’t exist, and so, by (A), ‘Santa Claus doesn’t exist’ is true. But, precisely because Santa Claus doesn’t exist, ‘Santa Claus’ fails to refer, and so ‘Santa Claus doesn’t exist’ is gappy and so not true. Contradiction. Thus, (A) fails. Moreover, if T merely consists of ‘Santa Claus doesn’t exist’, one should accept T although one should not accept ‘Every member of T is true’ (and should actually reject it and accept its contradictory). Thus, for such benign kinds of truth-value gaps, *acceptance is not expressed by attributions of truth*.

These facts have an immediate relevance for the contemporary debate about the semantic paradoxes, for in that debate quasi-deflationists object to certain theories of truth on the grounds that, by rejecting (A), they would prevent truth from fulfilling the function of expressing acceptance. But, as we’ve just seen, there is a variety of reasons, independent of the semantic paradoxes, for rejecting (A) in its full generality, and indeed for rejecting the objection’s presupposition that truth always fulfils the function of expressing acceptance (a point analogous to that made in the sixth last paragraph applies to the claim that such reasons concern special cases): sometimes, one may accept φ without accepting ‘ φ is true’.¹⁶

Notice that all this is not to jettison the requirement that rejection of a theory or

is good but non-conclusive evidence for each of the many independent sentences belonging to T), and *truth will not help to express it*. (Notice that to fall back on accepting something along the lines of ‘Every member of T is *likely* to be true’ would be on the contrary *too weak*, for such acceptance does not commit one to accepting individually *each* sentence belonging to T , which is certainly involved even by the weaker kind of accepting T —in fact, such acceptance does not even commit one to accepting individually *any* sentence belonging to T !) An analogous distinction can be drawn for *rejection*: in a *stronger* sense, rejection of T consists in, for some non-empty $U \subseteq T$, rejecting *individually each* sentence belonging to U *conditional* on accepting *individually each* other such sentence (the sense in which the author of a book cannot rationally reject the theory of her book), while in a *weaker* sense it consists in rejecting *as a whole* the sentences belonging to T (the sense in which the author of a book can rationally reject the theory of her book). (Thus, thinking of oneself as one’s body of belief, while, in the weaker sense, one must rationally accept oneself, in the stronger sense one can rationally not accept oneself, and while, in the stronger sense, one cannot rationally reject oneself, in the weaker sense one can rationally reject oneself.) And an analogous point applies concerning the fact that truth does not express the former, stronger kind of rejection. Having said that, I should like to note that, in certain *non-contractive* logics, a distinction can be drawn between the *standard* universal quantifier ‘every’ and a *weaker* universal quantifier ‘any’, and that the latter quantifier may well offer a solution to the expressive problem pointed out in this fn (see Zardini [2014g] for more details). Thanks to an anonymous referee for an observation that led to this fn.

¹⁶Richard [2008], whose general outlook is broadly congenial to the arguments of this section, in effect explicitly notices on p. 44 that some benign kinds of truth-value gaps prevent truth from always fulfilling the function of expressing acceptance, but does not bring this observation to bear on the quasi-deflationist argument (oddly, Richard [2008], p. 68 also claims that a close kin of (A) and (D) holds unrestrictedly, which, as far as I can see, is inconsistent with the existence of benign kinds of truth-value gaps). Thanks to an anonymous referee for alerting me to the relevance of Richard’s work.

even of a single sentence, and indeed acceptance of a theory, be at least associated with acceptance of a sentence, but, at least in the case of benign kinds of truth-value gaps, this would have to be achieved by means other than truth. In my view, a good way of expressing rejection of φ is to use a *Boolean negation* $\neg_{\mathcal{B}}$ and accept $\neg_{\mathcal{B}}\varphi$, and a good way of expressing acceptance of T is to use some *suitably expansive substitutional quantification* and suitable quotation device $[\cdot]$ and accept ‘For every X , if $[X]$ belongs to T , X ’.^{17,18}

It is crucial to note that the scheme I’ve just proposed is *not* something that non-naive or non-transparent theories of truth would be comfortable with, for, in order to do its designed work, the envisaged suitably expansive substitutional quantification would need to be governed by principles that clash with classical logic just as (A), (D) and (I) do (although this statement must admittedly remain more conjectural than other ones, since the paradoxes of quantification of this kind have been less studied than the paradoxes of truth). However, I don’t think that this circumstance implies that the quasi-deflationist argument can simply be *reformulated* in terms of the envisaged suitably expansive substitutional quantification (in fact, the two points I’m going to make apply generally with at least equal force—and indeed with even more force in the particular case of the third item mentioned by the first point—if the quasi-deflationist argument is reformulated in terms of an alternative device for fulfilling expressive functions that is more similar to truth in simply being a first-order predicate).

Firstly, the scheme I’ve proposed is itself *not of unrestricted applicability*: I’ve discussed in fn 15 a kind of case in which it can only be applied under some extremely controversial assumptions (concerning the principle of *contraction*), I’ve discussed in fn 17 a kind of case in which it can only be applied under some dubious assumptions (concerning the existence of *embeddable surrogates* of imperative sentences), I’ll discuss in section 3 a kind of case in which it can only be applied under some unusual assumptions (concerning the *expansiveness* of substitutional quantification) and I’ll discuss in section 4 a kind of case in which it cannot be applied at all. It would rather seem that, given the *variety* and *complexity* of what one can accept or reject, *there is no absolutely universal device for fulfilling expressive functions*, and that, *in different situations, different devices will be appropriate* (including devices of a more *pragmatic* nature than those I explicitly consider in this paper). Thus, since, as explained in the ninth last paragraph,

¹⁷For cases along the lines of some of those discussed in fn 13, the relevant members of T should be systematically massaged into sentences that *embed as expected under* $\neg_{\mathcal{B}}$ (for example, schematic sentences should be massaged into the universally quantified sentences suggested at the end of section 3). It is not clear that this is always possible (for example, in the case of imperative sentences).

¹⁸Given how really expansive the envisaged substitutional quantification will turn out to be supposed to be (see especially fn 30), it is to some extent *misleading* to label it as ‘*substitutional*’, as this suggests that, contrary to *objectual* propositional quantification, its range is in some sense tied to the expressive resources of the language, while it is not (in fact, in this respect, as far as I can see the sentences that come out true under my substitutional reading of the quantifiers are the same as those that come out true under an objectual reading). For a variety of reasons (which I’ll list in decreasing order of importance), for our purposes I still prefer the substitutional way of doing things: it is not subject to the vicissitudes that may prevent meaningful sentences from expressing propositions, it makes the ascent required by a quotation device like $[\cdot]$ smoother and it has a semantics that is more naturally extended to reach the suitable expansiveness.

the quasi-deflationist argument crucially relies on insisting that the relevant device for fulfilling expressive functions should continue to do its work also in every special case, it would still break down.

Secondly, the scheme I’ve proposed is not something that many quasi-deflationists would be comfortable with either, for, as I’ve already noted, *its immediate ground*—the requirement that rejection of a theory or even of a single sentence, and indeed acceptance of a theory, be at least associated with acceptance of a sentence—is *actually incompatible with the theories of truth that are typically supposed to be eventually justified by considerations concerning the expressive functions of truth*—an incompatibility which becomes manifest in the use of *Boolean negation* made by the scheme.¹⁹ In fact, as far as I can see, just about the *only* kind of theory that remains standing after that requirement has been imposed is the kind of theory developed in my works referenced in fn 1 (see in particular Zardini [2011], pp. 512–514; [2014c]; [2014d] for a discussion of how Boolean negation and (I) become both available once the structural principle of *contraction* is restricted). I do find such an argument from “*propositional expression of representational attitudes*” attractive, but I doubt that many other theorists would do so.

3 Truth-Bearer Contingency

Let’s proceed with another kind of failure of transparency that is also a kind of failure of naivety, in particular a kind of failure of (A). In section 2, we assumed that the operative truth bearers are *sentences*. But another prominent kind of entities that are apt for truth are *propositions*. Indeed, for some purposes, propositions are better suited than sentences to be taken as the operative truth bearers. For example, there’s intuitively something true about me saying “I’m hungry”, but, extremely plausibly, what is true is not the sentence ‘I’m hungry’, but the proposition ⟨I’m hungry⟩.²⁰

Some propositions arguably *depend for their existence on the existence of the objects they refer to*: for example, it is arguable that, necessarily, ⟨Socrates is wise⟩ exists only if Socrates exists.²¹ It is certainly not my task here to argue extensively for this view,

¹⁹I should remark in passing that that requirement (properly understood) and the justification that I’ve given for it are *not at all affected* by the possible limits of the scheme I’ve proposed that I’ve emphasised in the last paragraph, so that *that scheme perfectly fits the requirement and optimally embodies the justification that I’ve given for the requirement*. To make this explicit with respect to the two possible limits considered so far, the kind of case discussed in fn 15 does not involve a *single* state of mind in the first place, and so falls outside the scope of the justification that I’ve given for the requirement and outside the scope of the requirement itself understood as concerning only the *stronger* kind of acceptance and the *weaker* kind of rejection; the relevant kind of case discussed in fn 17 does not involve a *representational* state of mind in the first place, and so falls outside the scope of the justification that I’ve given for the requirement and outside the scope of the requirement itself understood as concerning only *indicative* sentences.

²⁰Throughout, an utterance of ⟨ φ ⟩ refers to the proposition expressed at the utterance’s context by φ .

²¹Such propositions are often labelled as ‘*singular*’, but that strikes me as a very misleading label, since the *object-dependence* feature mentioned in the text is universally agreed to be supposed to be exemplified for exactly the same reason by ⟨Socrates and Plato are wise⟩, which has little “singular” about it. Thanks

but one suggestive argument in its favour can at least be formulated even if not properly discussed. The argument starts with the very plausible assumption that \langle Socrates is wise \rangle is partly the proposition it is because it refers to Socrates, and very plausibly infers from it that, necessarily, \langle Socrates is wise \rangle exists only if it refers to Socrates, which, together with the other plausible assumption that, necessarily, Socrates is referred to only if he exists, yields the desired conclusion. Be that as it may with this particular argument, given that the view is appealing for a wide variety of conceptions of propositions (from *structured*—see for example Prior [1957]—to *unstructured*—see for example Stalnaker [2010]), let's work under the plausible assumption that it is correct.

Notice that, if so, an analogous view would seem to be correct for *semantically individuated* sentences: if \langle Socrates is wise \rangle is such that, necessarily, it exists only if Socrates exists, then surely also the sentence 'Socrates is wise', *qua* meaning that Socrates is wise, is such that, necessarily, it exists only if Socrates exists. Indeed, *mutatis mutandis*, the suggestive argument formulated in the last paragraph applies just as well for semantically individuated sentences. And, even more straightforwardly, necessarily the semantically individuated sentence 'Socrates is wise' exists only if it expresses \langle Socrates is wise \rangle , and so, necessarily, if that proposition does not exist (which, necessarily, it does not if Socrates does not exist), since it is plausible that the semantically individuated sentence does then not express it (as it is plausible that, necessarily, a proposition is expressed only if it exists) the semantically individuated sentence does not exist either.^{22,23}

Consider then that 'Socrates is wise or Socrates is not wise' is a logical truth, and so presumably is 'Necessarily, Socrates is wise or Socrates is not wise'.²⁴ Since, by (A), 'Socrates is wise or Socrates is not wise' entails ' \langle Socrates is wise or Socrates is not wise \rangle is true', it then follows, by single-premise closure of necessity under logical consequence, that, necessarily, \langle Socrates is wise or Socrates is not wise \rangle is true. However, possibly, Socrates does not exist, and so, given the object dependence of \langle Socrates is wise or Socrates is not wise \rangle , possibly \langle Socrates is wise or Socrates is not wise \rangle does not exist either. But surely, necessarily, something is true only if it exists, and so, possibly, \langle Socrates is wise or Socrates is not wise \rangle is not true, and so it is not the case that, necessarily, \langle Socrates is wise or Socrates is not wise \rangle is true. Contradiction. Thus, transparency, and also naivety (in particular, (A)), fail (essentially, the point goes back at least as far as Fine [1977], p. 136).

to Nasim Mahoozi for prompting this fn.

²²Since, necessarily, languages are *conventional*, it is very plausible that, necessarily, they are the *product of rational beings* (a sort of *abstract artifact*); if so, it follows that, necessarily, 'Socrates is wise' (just as well as any other sentence, whether semantically individuated or not) exists only if there are rational beings.

²³I should note that the *simple-present tense* in these arguments is meant in the familiar sense of '*at some time in the present, past or future*'. I should also note that I'm not inclined to run analogous arguments for *temporality*: for example, I'm inclined to reject the claim that \langle Socrates is wise \rangle now is partly the proposition it is because (and so now exists only if) it *now refers* to Socrates (it might suffice that it *referred* to Socrates).

²⁴The latter claim could be derived from the former by *necessitation*, but it enjoys independent high plausibility (even more so for the weaker claim that, necessarily, Socrates is wise or Socrates is not wise, which is all is needed by the argument to follow in the text).

Recall from section 2 that one expressive function for truth to fulfil emphasised by quasi-deflationists is that of expressing acceptance of a theory. Observe now that that point is supposed to *extend to suppositional contexts* (understood broadly so as to include *modal* contexts). To take one example, one expressive function for truth to fulfil emphasised by quasi-deflationists is that of *expressing acceptance of the necessity of a theory* (see for example Field [1994a], pp. 265–266; T is *necessary* in the natural sense that, for every X , if $\langle X \rangle$ belongs to T , necessarily X). If T contains *infinitely many* propositions, or if one *does not know* exactly which propositions belong to T , quasi-deflationists claim that one can still accept the necessity of T by accepting ‘Every member of T is necessarily true’.²⁵

Quasi-deflationists usually claim that fulfilment of the function of expressing acceptance of necessity requires truth to satisfy (A) in suppositional contexts (see for example Field [1994a], pp. 265–266). However, truth-bearer contingency provides counterexamples both to (A) in suppositional contexts (as we’ve seen) and to the quasi-deflationist idea that acceptance of necessity is always expressed by attributions of necessary truth (as we’ll now see). If T consists simply of $\langle \text{Socrates is wise or Socrates is not wise} \rangle$, one should accept that T is necessary (for one should accept that, necessarily, Socrates is wise or Socrates is not wise) although one should not accept ‘Every member of T is necessarily true’ (and should actually reject it and accept its contradictory). Thus, given truth-bearer contingency, *acceptance of necessity is not always expressed by attributions of necessary truth*.²⁶ Notice that one cannot weaken the attribution of necessary truth along the lines of ‘Every member of T is such that, necessarily, if it exists, it is true’, for that would not express acceptance of the necessity of T either: for example, if T merely consists of $\langle \text{Socrates exists} \rangle$, one should accept ‘Every member of T is such that, necessarily, if it exists, it is true’ although one should not accept the necessity of T (and should actually reject it and accept its contradictory).²⁷

²⁵The interaction of *quantification* with *modality* unsurprisingly generates issues of *scope*. In this respect, the obvious, scope-inverting alternative acceptance of which might be thought to express acceptance of the necessity of T is ‘Necessarily, every member of T is true’. However, as I explain in fn 26, although it might behave interestingly differently with respect to our dialectic I don’t think that this alternative is ultimately more tenable than the sentence I focus on in the text.

²⁶To come back to the alternative mentioned in fn 25, notice that, in the example in the text, on a standard understanding of the sentence and under a very plausible metaphysical assumption about theories, one should not accept ‘Necessarily, every member of T is true’ either. For, on a standard understanding of the sentence, that is true only if, necessarily, T exists, and, under a very plausible metaphysical assumption about theories, necessarily T exists only if all its members exist. I suppose however that one could so view ‘Every member of T is true’ as to count it as vacuously true at worlds in which no members of T exist. On this scheme, in the example in the text, ‘Necessarily, every member of T is true’ would be true. But it would still not express acceptance of the necessity of T : for example, if T merely consists of $\langle \text{Socrates exists} \rangle$, on this scheme one should accept ‘Necessarily, every member of T is true’ although one should not accept the necessity of T (and should actually reject it and accept its contradictory).

²⁷Curiously, Field [1994a], pp. 250–251 does note truth-bearer contingency while apparently assuming (at least officially) that, contrary to what I’ve just argued in the text, it does not create problems for truth to fulfil the function of expressing acceptance of necessity (see however Field [1994a], p. 264, fn 18 for a hint that Field is probably aware of the complexities generated by truth-bearer contingency).

These facts have an immediate relevance for the contemporary debate about the semantic paradoxes, for in that debate quasi-deflationists object to certain theories of truth on the grounds that, by rejecting (A) in suppositional contexts, they would prevent truth from fulfilling the function of expressing acceptance of necessity. But, as we've just seen, there are reasons, independent of the semantic paradoxes, for rejecting (A) in its full generality in suppositional contexts, and indeed for rejecting the objection's presupposition that truth always fulfils the function of expressing acceptance of necessity (a point analogous to that made in section 2 applies to the claim that such reasons concern special cases): sometimes, one may accept the necessity of φ without accepting ' $\langle\varphi\rangle$ is necessarily true'.

To take another example, one expressive function for truth to fulfil emphasised by quasi-deflationists is that of *expressing supposition of a theory* (see for example Field [2008], pp. 209–210; one *supposes* T in the natural sense that, for every X , if $\langle X\rangle$ belongs to T , one supposes that X). If T contains *infinitely many* propositions, or if one *does not know* exactly which propositions belong to T , quasi-deflationists claim that one can still suppose T by supposing 'Every member of T is true'.

Quasi-deflationists usually claim that fulfilment of the function of expressing supposition requires truth to satisfy (A) in suppositional contexts (see for example Field [2008], p. 210). However, truth-bearer contingency provides counterexamples both to (A) in suppositional contexts (as we've seen) and to the quasi-deflationist idea that supposition is always expressed by suppositional attributions of truth (as we'll now see). If T consists simply of \langle Socrates is not wise \rangle , one may suppose T (for one may suppose that Socrates is not wise) although one need not suppose 'Every member of T is true' (for a situation in which Socrates does not exist, and so in which \langle Socrates is not wise \rangle does not exist either and so is not true, is a situation compatible with the supposition that Socrates is not wise). Thus, given truth-bearer contingency, *supposition is not always expressed by suppositional attributions of truth*. Notice that one cannot weaken the suppositional attribution of truth along the lines of 'Every member of T is such that, if it exists, it is true', for that would not always express supposition of T either: for example, if T merely consists of \langle Socrates exists \rangle , one may suppose 'Every member of T is such that, if it exists, it is true' although one need not thereby suppose T .

These facts have an immediate relevance for the contemporary debate about the semantic paradoxes, for in that debate quasi-deflationists object to certain theories of truth on the grounds that, by rejecting (A), they would prevent truth from fulfilling the function of expressing supposition. But, as we've just seen, there are reasons, independent of the semantic paradoxes, for rejecting (A) in its full generality in suppositional contexts, and indeed for rejecting the objection's presupposition that truth always fulfils the function of expressing supposition (a point analogous to that made in section 2 applies to the claim that such reasons concern special cases): sometimes, one may suppose φ without supposing ' $\langle\varphi\rangle$ is true'.²⁸

²⁸It is well-known that (A), (D) and (I) fail in *hyper-intensional* contexts such as *propositional attitudes* and *implications*: the nominalist believes that snow is white without believing that \langle Snow is white \rangle is true; if the nominalist were right, snow would be white without \langle Snow is white \rangle being true. Sometimes,

I'd now like to push things even further, for I think that, with some ingenuity, using the same considerations about truth-bearer contingency (A) can also be seen to be problematic in *non-suppositional* contexts. Choosing a particular egg cell e and sperm cell s from Xanthippe and Socrates respectively (plus whatever other details may be required in view of the *essentiality of origins*), that determines a unique person that would have been Xanthippe and Socrates' daughter. Then, possibly, for some x , x is the unique person es generated by e and s and, in that possibility, it is the case that, actually, es does not exist.²⁹ From this possibility and the logical properties of the interaction between possibility and actuality, using the suitably expansive substitutional quantification mentioned at the end of section 2 we know that, actually, for some X , X but $\langle X \rangle$ is not true (as $\langle X \rangle$ does not even exist).³⁰ More informally, we know that, actually, *something is the case that is not reflected in the truth of a proposition*. And that is at least in great tension with (A).

Now, one expressive function for truth to fulfil emphasised by quasi-deflationists is that of *expressing acceptance of the absolute generality of a principle* (see for example Field [2008], p. 220). If P is a principle with an associated schematic formulation $\text{sch}(P)$, quasi-deflationists claim that one can accept the absolute generality of P by accepting 'Every instance of $\text{sch}(P)$ is true'.

Quasi-deflationists usually claim that fulfilment of the function of expressing accep-

(A), (D) and (I) even come with an explicit proviso concerning such contexts (see e.g. Field [1994a], p. 251). The proviso is of course completely legitimate and does not in the least detract from the validity of those principles. However, *the broad kinds of expressive functions emphasised by quasi-deflationists also concern these contexts*. Thus, although the failures of (A), (D) and (I) in hyper-intensional contexts do not tell against *their validity* (contrary to their other failures discussed in this paper), they do tell against *the quasi-deflationist argument in favour of them*.

²⁹In conformity with the argument in favour of truth-bearer contingency set out at the beginning of this section, I don't assume that ' es ' actually refers to anything. But I do assume, at least in the official formulation of the claim made in the text, that this does not prevent es from functioning in a natural way when embedded in suppositional contexts—in particular, that, when embedded in a suppositional context, ' es ' can be used to pick out the unique witness of an existential claim occurring at the same or less deep level of embedding. I recognise however that this may be controversial, and observe that the gist of the claim made in the text and that of similar claims remains unaltered if we so reformulate them as to replace 'Possibly, for some x , x is the unique person es generated by e and s and ... es ...' with 'Possibly, for some x , x is the unique person generated by e and s and ... x ...'.

³⁰For this to make sense, substitutional quantification must be *really* expansive: specifically, it must be the case that a particular substitutional quantification is true if (and only if), possibly, there is an expansion of the language such that some sentence in the relevant substitution class is true at the relevant world of evaluation (in our case, the relevant sentence could be something like ' es does not exist'—which, individuated semantically, could exist even if it actually doesn't—and the relevant world is the actual world). Notice that the previous clause only requires, as usual, truth *at* a world rather truth *in* a world (and that, in our case, there is no problem for something like ' es does not exist' to be true at the actual world even if it is not true in it). One might wonder whether the envisaged kind of expansive substitutional quantification is legitimate, but it's hard to see why it shouldn't be: the explanation just given seems perfectly *coherent* (although it could certainly use some more precision), and we clearly seem to have a notion of *something being the case* under which, for example, actually something is the case with respect to the specific non-existence of the unique person that would have been generated by e and s .

tance of absolute generality requires truth to satisfy (A) (see for example Field [2008], p. 220). However, truth-bearer contingency provides both examples problematic for (A) (as we've seen) and counterexamples to the quasi-deflationist idea that general attributions of truth always express acceptance of absolute generality (as we'll now see). Suppose that one thinks that the law of excluded middle (LEM) fails only in some cases involving non-existing objects. Then, one should accept that every instance of $\text{sch}(\text{LEM})$ is true although one should not accept the absolute generality of LEM (and should actually reject it and accept that LEM is not absolutely general). Thus, given truth-bearer contingency, *general attributions of truth do not always express acceptance of absolute generality and are actually compatible with rejection of absolute generality.*³¹

These facts have an immediate relevance for the contemporary debate about the semantic paradoxes, for in that debate quasi-deflationists object to certain theories of truth on the grounds that, by rejecting (A), they would prevent truth from fulfilling the function of expressing acceptance of absolute generality. But, as we've just seen, there are reasons, independent of the semantic paradoxes, for rejecting (A) in its full (here, more accurately, absolute) generality, and indeed for rejecting the objection's presupposition that truth always fulfils the function of expressing acceptance of absolute generality (a point analogous to that made in section 2 applies to the claim that such reasons concern special cases): sometimes, one may accept 'Every instance of $\text{sch}(P)$ is true' without thereby accepting the absolute generality of P .

All this is not to jettison the requirement that acceptance of the necessity of a theory, or supposition of a theory, or acceptance of the absolute generality of a principle be at least associated with acceptance or supposition of a sentence,³² but, in the case of truth-bearer contingency, this would have to be achieved by means other than truth. In my view, a good way of doing all this is to use the suitably expansive substitutional quantification mentioned at the end of section 2 and in this section, and, for acceptance of the necessity of T , accept 'For every X , if $\langle X \rangle$ belongs to T , necessarily X '; for supposition of T , suppose 'For every X , if, actually, $\langle X \rangle$ belongs to T , X '; for acceptance of the absolute generality of a principle P , accept the result of replacing the schematic letters of $\text{sch}(P)$ with substitutional variables and binding these with universal quantifiers (so as, to take

³¹In many of the cases discussed in section 2 and in this section, what emerges is that our thought is *diaphanous* in the sense that, in accepting (for certain purposes) φ , we *simply think that things are such-and-such (full stop), typically without any direct bearing on φ itself*, whose truth might then be prevented by factors that do not prevent things from being such-and-such. For example, in accepting 'Santa Claus doesn't exist', we simply think that Santa Claus doesn't exist (full stop), without any direct bearing on 'Santa Claus doesn't exist' itself, whose truth might then be prevented by factors that do not prevent things from being such that Santa Claus doesn't exist. In all these cases, it is precisely the *diaphanousness* of our *thought* that, by enabling us to appreciate the *gaps* between *how things are* and *which truth bearers are true*, enables us to appreciate the *opacity* of truth.

³²Notice that the justification that I've given in section 2 for a broadly related requirement does not apply in the case of supposition of a theory, for that justification relies on the assumption that the state of mind in question is supposed to represent a way things *are*, which supposition clearly isn't. However, the obvious modification of that justification relying on the assumption that the state of mind in question is supposed to represent a way things *could be* does apply to the case of supposition of a theory, and yields as conclusion that every supposition is expressed by *supposition* of a certain sentence.

the example of LEM, to accept ‘For every X , either X or it is not the case that X ’.³³

4 Truth-Condition Context Dependence

Let’s close with another kind of failure of transparency that is also a kind of failure of naivety, in particular a kind of failure of both (A) and (D). In sections 2 and 3, we assumed that the operative truth bearers are *sentences* or *propositions* respectively. But another prominent kind of entities that are apt for truth are *utterances* (understood, roughly, as *use-specific sentence tokens*). Indeed, for some purposes, utterances are better suited than sentences or propositions to be taken as the operative truth bearers. For example, there’s intuitively something true about me saying “I’m not hungry” in yesterday’s entry of my diary, but, plausibly, what is true is neither the sentence ‘I’m not hungry’ nor the proposition \langle I’m not hungry \rangle ,^{34,35} but the utterance consisting in the inscription in the

³³Heck [2004], pp. 322–329 also takes up the challenge of Field [1994a], pp. 265–266 concerning how to express acceptance of the necessity of T . Keeping fixed a non-deflationist conception of truth, he grants Field that it would not be adequate to accept ‘Every *sentence* member of T is necessarily true’, for that would involve *issues about the representational properties of sentences* that are intuitively foreign to the issue whether T is necessary. Heck’s proposal is rather to accept ‘Every *proposition* expressed by any member of T is necessarily true’. *Keeping fixed a non-deflationist conception of truth as applied to propositions too*, it is not clear to me why acceptance of the latter sentence does not incur in the analogous problem of involving *issues about the representational properties of propositions* that are intuitively foreign to the issue whether T is necessary. Be that as it may, the proposal is subject to the problem from truth-bearer contingency that I’ve explored in this section. I should add though that Heck also says things (as on p. 327) suggesting to me that he wouldn’t be hostile to developing further his proposal along lines similar to those of my own proposal in the text.

³⁴For, at least under the assumption of *temporalism* (very roughly, the doctrine that propositions typically change their truth value over time), that proposition is true (that is, true *now*) iff I’m now not hungry, which however I am. Granted, even under temporalism, some remotely related proposition is true, like \langle I was not hungry (on 10/07/2013) \rangle . However, firstly, that proposition does not seem *relevant* to the intuition that there’s something true about me saying “I’m not hungry” in yesterday’s entry of my diary, as it never gets expressed by anyone in the envisaged circumstances; secondly, at least under the further assumption of *tensism* (very roughly, the doctrine that propositions are tensed and that their truth value is sensitive to such tense), that proposition was false yesterday, while the intuition in question is also an intuition to the effect that there’s something that *was true yesterday and is true today* about me saying “I’m not hungry” in yesterday’s entry of my diary.

³⁵In general, I resist the automatic inference from the premise that instances of kind k can only be F *as considered under a certain aspect* to the conclusion that instances of k cannot be, in the relevant cases, *the sole bearers of F ness* (for example, from the premise that people can only vote for people when considered as electors and candidates respectively for a particular post to the conclusion that people cannot be, in the relevant cases, the sole bearers of the relation of voting). In fact, in Zardini [2014b] I’ve resisted precisely that inference in the similar case of the question of the bearers of *logical consequence*. But I think that, in the case of the question of the bearers of *truth* discussed in the text, several auxiliary considerations are available to licence the inference. Firstly, contrary to the cases which motivate the idea that the inference might fail, it is not at all intuitive to think that ‘I’m not hungry’ or \langle I’m not hungry \rangle *used by me yesterday* are true (in contrast to its being intuitive to think that they *were* true). Secondly, contrary to the cases which motivate the idea that the inference might fail, ‘I’m not hungry’ and \langle I’m not hungry \rangle are *unqualifiedly not true* (and so *not unqualifiedly true*), while the intuition in question is also an intuition to the effect that there’s *nothing unqualifiedly not true* (and, indeed, *something unqualifiedly*

diary.³⁶

Some sentences are undoubtedly *context dependent*, in that utterances of them in different contexts can differ in truth conditions and so in truth value. For example, suppose that Rajoy is hungry while Merkel is not. Then, Rajoy’s utterance of ‘I’m hungry’ is true iff Rajoy is hungry and so is true, while Merkel’s utterance of ‘I’m hungry’ is true iff Merkel is hungry and so is false.

Given that there is *no utterance that is uniquely associated with a sentence*, there is an issue as to how to understand (A), (D) and (I) if the operative truth bearers are utterances.³⁷ The simplest understanding takes ‘ $\ulcorner\varphi\urcorner$ ’ as it occurs in those principles as standing in a systematically ambiguous way for *absolutely any* utterance of the value of ‘ φ ’. However, I’m a philosopher, but Rajoy’s utterance of ‘I’m a philosopher’ is false. Thus, on the understanding in question, transparency, and also naivety (in particular, (A)), fail.³⁸ Moreover, Merkel’s utterance of ‘I’m not a philosopher’ is true, but I’m a philosopher. Thus, again, on the understanding in question, transparency, and also naivety (this time, in particular, (D)), which had remained untouched by the previous shadows cast by opacity), fail.

It’s relatively easy to come up with more restrictive understandings of (A), (D) and (I)

true) about me saying “I’m not hungry” in yesterday’s entry of my diary. Thirdly, what’s intuitively true about me saying “I’m not hungry” in yesterday’s entry of my diary is *the record I produced yesterday consisting in certain marks on paper*: something that can be burnt, that contains an apostrophe, that is in English, that in two thousand years will be exhibited in an archeology museum etc. All those are not properties that ‘I’m not hungry’ or ⟨I’m not hungry⟩ can plausibly exemplify, even as used by me yesterday.

³⁶In other work (for example, Zardini [2008]), I’ve rather used ‘utterance’ to refer to *assertoric speech acts* and focus on those as the operative truth bearers. But utterances understood along those lines would be unsuitable to account for the intuition discussed in the text, for the only assertoric speech act made in the envisaged circumstances is the act performed by inscribing ‘I’m not hungry’ in my diary: that act only existed yesterday, and so, given the plausible assumption that something is true only if it exists, it is not true (that is, true *now*), and so cannot account for the intuition that there’s something true about me saying “I’m not hungry” in yesterday’s entry of my diary.

³⁷True, given truth-condition context dependence, it is also the case that there is *no proposition that is uniquely associated with a sentence* (for example, uttered by Rajoy ‘I’m hungry’ expresses something along the lines of ⟨Rajoy is hungry⟩, while uttered by Merkel ‘I’m hungry’ expresses something along the lines of ⟨Merkel is hungry⟩). But the dialectic of section 3 did not require considering context-dependent sentences, and many sentences that are not context dependent do have a proposition that is uniquely associated with them. Indeed, even if the relevant sentences in (A), (D) and (I) are context dependent, one could take ‘ $\ulcorner\varphi\urcorner$ ’ as it occurs in an utterance of any of those principles as standing for the proposition expressed in the context of that utterance by any utterance of the value of ‘ φ ’. (Notice that such understanding would not have the crippling effects that a similar move would have in the case of utterances (which would take ‘ $\ulcorner\varphi\urcorner$ ’ as it occurs in an utterance of any of those principles as standing for the utterance of the value of ‘ φ ’ that is correlated with it in the principle), as it still yields versions of (A), (D) and (I) that govern attributions of truth also to propositions expressed by utterances made at a context different from the context in which (A), (D) and (I) are uttered (as long as those propositions are also expressed by some utterance (that could be) made in the latter context).)

³⁸Utterances, even more clearly than other kinds of truth bearers, are contingent. Therefore, similarly to what was observed in section 3, (A) fails in a possible world w (including the actual world), if, as it will so often be the case, φ is true at w but there are no utterances of φ in w . Points similar to those made in section 3 thus apply if utterances are assumed to be the operative truth bearers.

that avoid the problems raised in the last paragraph.³⁹ For example, one could stipulate that ‘ $\ulcorner \varphi \urcorner$ ’ as it occurs in those principles stands in a systematically ambiguous way for any utterance of the value of ‘ φ ’ *that is made in the same context as the relevant utterance of (A), (D) and (I)*. Notice however that, while such a restriction (as well as the even more draconian restriction mentioned in fn 37) would indeed avoid the problems raised in the last paragraph, it (as well as the even more draconian restriction mentioned in fn 37) does so only at the cost—unaffordable by the quasi-deflationist—of preventing truth from fulfilling the expressive functions we’ve been reviewing *in the quasi-deflationist’s way*, since, while those functions clearly require truth to apply also to utterances made in a context different from the context in which (A), (D) and (I) are uttered, the quasi-deflationist’s way to fulfil them goes *via* (A), (D) and (I).

The situation would improve for the more restrictive understanding of (A), (D) and (I) discussed in the last paragraph if, for every utterance u_0 made in a context different from the context in which (A), (D) and (I) are uttered, there were an utterance u_1 made in the latter context such that u_0 is true iff u_1 is (under a suitably strong reading of ‘iff’).⁴⁰ Since people are neither as *knowledgeable* (as quasi-deflationists themselves are fond of stressing) nor as *talkative* as the thought under consideration apparently presupposes (I might utter ‘Everything Rajoy says is true’ without uttering anything else—either because I don’t know what exactly Rajoy says or because, even if I know that, I can’t be bothered to utter anything else—and so be left in my context without a suitable choice for ‘ u_1 ’), that thought only makes sense if it dubiously appeals to *truth bearers other than (actual) utterances*, but let that pass. For even granting that the thought under consideration may thus work for at least *some* choices for ‘ u_0 ’, it still does not for *many* other choices for ‘ u_0 ’: for example, if, in the solitude of his office, Rajoy points to one of his hairs and

³⁹A failed attempt is to replace ‘ $\ulcorner \varphi \urcorner$ ’ with a simple variable ‘ u ’ and conditionalise (A), (D) and (I) with the clause ‘if u expresses $\langle \varphi \rangle$ ’ (see for example Williamson [1998]). For, as *per* the beginning of this section, yesterday’s entry in my diary expresses $\langle \text{I’m not hungry} \rangle$, but it is true although I’m hungry (see Zardini [2008] for further discussion of this attempt in the slightly different framework mentioned in fn 36).

⁴⁰For example, merely *material* implication will not do, for, in that sense, if, for instance, u_0 is an utterance of ‘Rajoy promotes regressive policies’ and u_1 is an utterance of ‘Merkel promotes regressive policies’, u_0 is true iff u_1 is (in which way I’ll leave it to you to decide), but, clearly, one may accept ‘ u_0 is true’ without accepting ‘ u_1 is true’.

utters ‘This is grey’, there is in my context no suitable choice for ‘ u_1 ’.^{41,42} Since fulfilment of the expressive functions emphasised by the quasi-deflationist requires truth to apply also to utterances made in a context different from the context in which (A), (D) and (I) are uttered, and since I don’t know of any restricted understanding of (A), (D) and (I)

⁴¹To ward off a likely rejoinder, an utterance of something along the lines of ‘The hair referred to by that utterance of ‘this’ by Rajoy is grey’ will not work for at least two reasons. Firstly, Rajoy’s utterances might be just as inextricable as his hairs are, in which case I won’t be able to use ‘that utterance of ‘this’ by Rajoy’ to *single out the relevant utterance* (and so won’t be able to use ‘the hair referred to by that utterance of ‘this’ by Rajoy’ to single out the relevant hair). Secondly, the device envisaged by the rejoinder is anyways inadmissible in the quasi-deflationist framework. For it would seem that the quasi-deflationist argument about *truth*, if it is any good, should be extendable to similar semantic properties like *reference*, in which case the rejoinder would simply push the problem to the one of getting ‘the hair referred to by that utterance of ‘this’ by Rajoy’ to behave as expected (that is, roughly, to be intersubstitutable with a singular term standing for the relevant hair) *in virtue of some sort of correlation principle about reference*. If this problem were dodged, and so if it were in effect conceded that ‘the hair referred to by that utterance of ‘this’ by Rajoy’ behaves as expected even in the absence of a suitable correlation principle, that would open the flood gates to opacity: if a *singular term* like ‘the object referred to by τ ’ behaves as expected even in the absence of a suitable correlation principle, given that the quasi-deflationist should have no objection to a *predicate* like ‘satisfies τ ’ behaving as expected (whether in virtue of some sort of correlation principle about satisfaction or not) one could accept any utterance u of any *atomic sentence* simply by accepting, roughly, ‘The objects referred to by the singular terms of u as they are used in u satisfy the predicate of u as it is used in u ’, *without need of any correlation principle about truth* (with this extending in the usual way to cover utterances of *compound sentences*). (Notice that any fuss about satisfaction still needed to be naive or transparent would be quite out of place, since *any predication can be rephrased as an identification*: for example, ‘Rajoy admires Merkel’ can be rephrased as ‘Rajoy is the Rajoy admirer of Merkel’. Reference and identity can ground truth without need of satisfaction.)

⁴²At this point of the dialectic, the problem has really become the more general one—stretching beyond the domain of truth-condition context dependence—of how, using correlation principles, to account for the fact that attributions of truth to *utterances one cannot understand* express acceptance of the relevant utterances just as well as attributions of truth to utterances that one does understand. Field [2001], pp. 147–151 proposes an interesting new strategy for solving that problem from a deflationist perspective. Adapted to the present dialectic, Field’s idea is that, even if one cannot understand an utterance u of φ , by attributing truth to u one *in some sense borrows φ and uses it* (or, if φ is context dependent, uses some suitable relative of it) *deferentially to u* , in such a way that one in some sense accepts ‘ φ iff u is true’. Interesting as it is, I think that there is a dilemma for this broad kind of strategy at least when it is supposed to help in the present dialectic (and I hasten to note that Field does not say that it does). On the one hand, if the strategy is understood so that φ , as one uses it, has a *virtually completely indeterminate content* (which I think is how Field intended the proposal), then, since one accepts ‘ φ iff u is true’, the content of ‘ u is true’, as one uses that sentence, will be forced to be virtually completely indeterminate too. But then one will not accept the theory expressed by u by accepting ‘ u is true’, for the content of the latter, as one uses it, will be virtually completely indeterminate, and so will fall short of entailing virtually any interesting theory. On the other hand, if the strategy is understood so that φ , as one uses it, has the *same content as it has in u* , then the underlying ultraliberal conception of how one can use sentences as having certain contents, in addition to being totally implausible (at least under standard assumptions concerning the relation between use of sentences and grasp of their contents), will suffice all by itself for fulfilling the expressive functions emphasised by quasi-deflationists. For that conception amounts in effect to the idea that it is sufficient for using a sentence as having a certain content that one be in a fairly remote relation to that content (plus, of course, that one intend to use that sentence as having that content). Thus, if T contains infinitely many sentences, or if one does not know exactly which sentences belong to T , whereas one was supposed to accept T by accepting ‘Every member of T is true’, it will now be sufficient that one simply intend to use φ so that it has the content of T (to which one is surely at least

under which these principles can effectively govern such applications, I'll thus henceforth assume the unrestricted understanding of (A), (D) and (I) discussed in the second last paragraph.

Focussing for illustrative purposes on (D), as we've already seen in section 2 one expressive function for truth to fulfil emphasised by quasi-deflationists is that of *expressing acceptance of a theory*, and quasi-deflationists usually claim that fulfilment of the function of expressing acceptance requires truth to satisfy (D) (see for example Field [2008], p. 205). However, truth-condition context dependence provides counterexamples both to (D) (as we've seen) and to the quasi-deflationist idea that truth only fulfils the function of expressing acceptance if (D) holds (as we'll now see). If T merely consists of Rajoy's utterance of 'I'm hungry', one should accept 'Every member of T is true' and, although one need not thereby accept that one is hungry (and may actually reject it), one would indeed thereby accept Rajoy's utterance. Thus, given truth-condition context dependence, *attributions of truth express acceptance even if (D) fails*. Notice however that the *way* in which the expression in question *works* is very different from the (D)-based one imagined by quasi-deflationists. For, in our example, since (D) fails truth cannot function as a *formal device*, with one attributing truth as a means to the end of committing oneself to accepting a *common-or-garden-variety truth-free* sentence like 'I'm hungry';⁴³ rather, truth functions as a *substantial category*, with one attributing truth as an end in itself, the end of classifying Rajoy's utterance as exhibiting a certain *feature* that in turn warrants its acceptance (just as one's being hungry more straightforwardly warrants the acceptance of one's utterance of 'I'm hungry').^{44,45}

in a fairly remote relation). This discussion actually helps to bring out yet another expressive function for truth to fulfil (one which, unsurprisingly, is not emphasised by quasi-deflationists since it has little to do with (A), (D), (I) or any other correlation principle): that of *expressing attitudes towards what one cannot understand* (when such attitude is acceptance, the *etymologies* of 'truth' and *veritas*, in stark contrast to the Heideggerian fanfare about ἀλήθεια, become both very suggestive). (Notice that, if what one cannot understand are sentences or propositions, in view of the limitations on truth explored in sections 2 and 3 the suitably expansive substitutional quantification fulfils such function even better (since, for every sentence or proposition one cannot understand, there is presumably an expansion of the language that contains a sentence synonymous with the sentence one cannot understand or a sentence expressing the proposition one cannot understand). However, if what one cannot understand are utterances, I'll point out at the end of this section that, in an important respect, truth fulfils such function better than the suitably expansive substitutional quantification. Thanks to an anonymous referee for a question that prompted this last comment.)

⁴³True, in our particular example it might be thought that one does have the end of committing oneself to accepting a different common-or-garden-variety truth-free sentence: 'Rajoy is hungry' (or a similar sentence). But, as the discussion in the last paragraph makes clear, there are many cases in which that cannot be the case given that it is totally plausible that there are no relevant common-or-garden-variety truth-free sentences one can understand, and so no relevant common-or-garden-variety truth-free sentences one can, in the relevant sense, commit to accepting.

⁴⁴Ironically, in the case of truth-condition context dependence, *it is exactly because all of (A), (D) and (I) fail* that truth fulfils the expressive functions emphasised by the quasi-deflationist.

⁴⁵(A), (D) and (I) are sometimes appealed to in justifying the mantra that *logical consequence requires necessary truth preservation* (see for example Field [2008], pp. 42–43, 284–286 for a perceptive reconstruction of such justification; Field himself rejects the justification and indeed the mantra for the reason I adumbrate in fns 10 and 49). In Zardini [2012], I've criticised the justification and its underlying concep-

These facts have an immediate relevance for the contemporary debate about the semantic paradoxes, for in that debate quasi-deflationists object to certain theories of truth on the grounds that, by rejecting (D), they would prevent truth from fulfilling the function of expressing acceptance (see for example Field [2008], pp. 148–149). But, as we’ve just seen, there are reasons, independent of the semantic paradoxes, for rejecting (D) in its full generality, and indeed for rejecting the objection’s presupposition that truth only fulfils the function of expressing acceptance if (D) holds (a point analogous to that made in section 2 applies to the claim that such reasons concern special cases): if u is an utterance of φ , sometimes one may accept ‘ u is true’ and thereby accept u without thereby accepting φ .⁴⁶

Notice that, as the first counterexample in section 2, also this last counterexample is not a counterexample to the claim that truth fulfils a certain expressive function—it’s only a counterexample to the claim that truth only fulfils that function if a certain correlation principle holds. And it is indeed good that it is not a counterexample to the former claim, for the alternative device for fulfilling expressive functions that I’ve proposed in sections 2 and 3 (the suitably expansive substitutional quantification) would not seem to work in the case of truth-condition context dependence. For, to take the example introduced at the beginning of this section, neither ‘For some X , yesterday’s entry in my diary is an utterance of $[X]$, and X ’ nor ‘For some X , yesterday’s entry in my diary expresses $\langle X \rangle$, and X ’ would work because of the reasons I’ve given there.

5 Opacity and the Semantic Paradoxes

I’ve argued that a prominent argument against certain theories of truth fails: truth is opaque, and this sometimes even leads to its not fulfilling certain expressive functions. However, in closing, I should emphasise that, even setting aside the (essentially non-truth-theoretic) argument from propositional expression of representational attitudes (in favour of a very specific revision of classical logic) that I’ve constructed in counterpoint to the quasi-deflationist argument in section 2, the arguments I’ve given are *not at all* supposed to provide an element in an overall defence of theories that solve the semantic paradoxes by rejecting the paradoxical instances of (A), (D) or (I): even if truth is in general opaque, those instances, or at least the relevant inferences licenced by those instances, might still be compelling. The arguments I’ve given are only supposed to show that those instances should not be accepted *simply in virtue of their being instances of* (A), (D) and (I), both

tion of necessary truth preservation mainly relying on considerations concerning context dependence (of a different kind though from those I’m exploiting in this section). To the (restricted) extent that logical consequence does require necessary truth preservation, the real grounds for such requirement are opaque. Notice that the requirement is actually of great relevance for the debate on the semantic paradoxes (see fn 49).

⁴⁶Notice that, while I’ve variously argued against each of (A), (D) and (I), and against several putative connections between acceptance/rejection of T and acceptance of ‘Every member of T is true’/‘Some member of T is not true’, no consideration in this paper tells against the idea that, if one accepts ‘Every member of T is true’, one thereby accepts T . Thanks to an anonymous referee for pointing this out.

because *the quasi-deflationist rationale for those principles is flawed* (since sometimes truth could fulfil certain expressive functions even if—or actually exactly because—the relevant instances of (A), (D) and (I) failed to hold) and because *the principles are anyways untenable in their full generality* (which sometimes even leads to truth not fulfilling certain expressive functions). We’ve seen that the quasi-deflationist argument tries to force the paradoxical instances of (A), (D) or (I) by arguing in effect, rather heroically, for the unrestricted holding of those principles.⁴⁷ But, quite generally, certain *instances* of a principle might hold, or at least certain *inferences licenced by those instances* might be acceptable, even if the *principle* does not unrestrictedly hold. Accordingly, while I’ve argued that considerations concerning the expressive functions of truth do not offer *finer-grained reasons for accepting the relevant inferences licenced by the paradoxical instances of (A), (D) or (I)*, there might be such reasons *in other areas of the philosophy of truth*. If we do insist on solving the semantic paradoxes by preserving those inferences and by revising classical logic instead, we should make sure that we do so for the right reasons.

Without aiming at an exhaustive list of the best candidates for being such reasons, let me close by briefly sketching a couple of them that I think are particularly salient.⁴⁸ Consider the *Liar sentence* l identical with ‘ l is not true’. A first, fast-track argument in favour of revising classical logic relies on ideas characteristic of the *correspondentist* theory of truth (contrary to the quasi-deflationist argument, which can rather be associated with the *deflationist* theory); the argument clings onto the last flickers of correlation principles lingering after the onset of opacity. Notice first that the l -instance of (A) and the l -instance of (D) truth-theoretically suffice for carrying through a standard paradoxical reasoning concerning l . (Assume for *reductio* that l —that is, ‘ l is not true’—is true. Then, by the l -instance of (D), l is not true. By *reductio*, l is not true. By the l -instance of (A), l is true. Contradiction.) But it is *prima facie* compelling that l represents things as being such that l is not true, and, since it is *prima facie* compelling that l ’s being true consists in things being how l represents them to be (i.e. in l ’s not being true), it is *prima facie* compelling that l ’s not being true entails and is entailed by l ’s being true. Granted, those three *kinds of claims* might have counterexamples for several types of sentences, as I’ve been at pains of stressing. *But l would not seem to be a token of any such type*, and so the compellingness of these three *particular claims* stands undefeated (in striking *contrast*

⁴⁷Setting aside those cases in which truth itself does not fulfil certain expressive functions, I’ve argued in Zardini [2014e] that a suitable development of the *hierarchical* theory of truth might well provide all the resources needed in the remaining cases. If so, even if some sort of revamped quasi-deflationist argument could pull off the magic trick of justifying a restriction that leaves out all the counterexamples offered in this paper and nevertheless leaves in paradoxical sentences, such argument would still break down at the further step assuming that truth only fulfils the relevant expressive functions concerning *paradoxical sentences* if the *paradoxical instances* of (A), (D) and (I) hold. (With its replacement of the *property of truth* with a hierarchy of *truth-like properties*, the hierarchical theory may also not be overly impressed by the two arguments in favour of revising classical logic that I’ll run below in the text. For what it’s worth, I’ve argued in Zardini [2014e] that the hierarchical theory cannot account for a kind of generalising use of truth which is different from the expressive one and which consists, roughly, in using truth to *attribute a unifying feature to all instances of certain kinds of sentences*, and, by doing so, to be able to speak for the first time about *general principles and notions*.)

⁴⁸Thanks to an anonymous referee for recommending this discussion.

to what happened with the quasi-deflationist argument, where a single counterexample to (A) or (D), *even if unrelated to the semantic paradoxes*, was sufficient for sinking the whole argument).

A second, more thoughtful argument in favour of revising classical logic more abstractly relies on truth's *interaction with logical operations*, on its *structural features* and on its *connections with other properties* (thus rising above specific theories on truth's *nature*, contrary to the quasi-deflationist argument and the argument in the last paragraph, which can rather be associated with the deflationist theory and with the correspondentist theory respectively); the argument operates entirely in the penumbra of opacity turning away from the glitter of correlation principles. A first compelling principle about truth concerns its interaction with *negation* in the form of *consistency*:

(CONS) For every truth bearers t and t^* such that t^* is the negation of t , if t^* is true, t is not true.

(CONS) is *opacity-friendly*, in that it is not at all affected by any of the considerations in favour of opacity reviewed in this paper—(CONS) may hold, and the most natural reasons in its favour be cogent, even if truth is opaque through and through (unsurprisingly so, since (CONS) is *not a correlation principle in the first place*). ((CONS) is also of course transparency-friendly; indeed (CONS) is compelling from very different perspectives on the nature of truth—correspondentist, deflationist, coherentist etc.)

A second compelling principle about truth concerns its structural feature of *iteration*:

(ITER) For every truth bearer t , if t is true, $\lceil t \text{ is true} \rceil$ is true

(given a suitable understanding of quantification into corner quotes). Notice that (ITER) is *opacity-friendly*—(ITER) may hold, and the most natural reasons in its favour be cogent, even if truth is opaque through and through (unsurprisingly so, since, although it may be regarded as a tightly restricted correlation principle, (ITER) is more naturally regarded as a *structural principle* concerning iterations of 'true', along the lines of the **S4**-axiom for lary operators—a principle to the effect that truth as a certain "flat" structure). ((ITER) is also of course transparency-friendly; indeed (ITER) is compelling from very different perspectives on the nature of truth—correspondentist, deflationist, coherentist etc.)

A third compelling principle about truth concerns its connection with *proof* in the form of *soundness*:

(SOUND) For every closed truth bearer t containing only standard logical and semantic vocabulary, if t is provable, t is true.

Notice that (SOUND) is *opacity-friendly*—(SOUND) may hold, and the most natural reasons in its favour be cogent, even if truth is opaque through and through (unsurprisingly so, since (SOUND) is *not a correlation principle in the first place*). ((SOUND) is

also of course transparency-friendly; indeed (SOUND) is compelling from very different perspectives on the nature of truth—correspondentist, deflationist, coherentist etc.)

Now, (CONS) yields that l —that is, ‘ l is not true’—is true only if ‘ l is true’ is not true, from which, by contraposition for implication, the l -instance of (D) follows by (ITER). By *reductio*, such instance yields l , given which (SOUND) suffices for licencing the inference that would otherwise be licenced by the l -instance of (A), yielding ‘ l is true’. Contradiction. Opaque truth is in conflict with classical logic just as transparent truth is.⁴⁹

References

- JC Beall. Deflationism and gaps: untying ‘not’s in the debate. *Analysis*, 62:299–305, 2002.
- JC Beall. *Spandrels of Truth*. Oxford University Press, Oxford, 2009.
- Marian David. Some T-biconditionals. In Bradley Armour-Garb and JC Beall, editors, *Deflationary Truth*, pages 382–419. Open Court, Chicago, 2005.
- Michael Dummett. Truth. *Proceedings of the Aristotelian Society*, 59:141–162, 1959.
- Hartry Field. Deflationist views of meaning and content. *Mind*, 103:249–285, 1994a.
- Hartry Field. Disquotational truth and factually defective discourse. *The Philosophical Review*, 103:405–452, 1994b.
- Hartry Field. Postscript to “Deflationist views of meaning and content”. In *Truth and the Absence of Fact*, pages 141–156. Oxford University Press, Oxford, 2001.
- Hartry Field. Solving the paradoxes, escaping revenge. In JC Beall, editor, *Revenge of the Liar*, pages 78–144. Oxford University Press, Oxford, 2007.

⁴⁹I find this second argument insightful, among other things, because it is easily *generalisable* to cover paradoxes concerning *properties other than truth* for which some sort of quasi-deflationist argument would seem to be a non-starter and for which there are no plausible instances of correlation principles that suffice for generating a conflict with classical logic. Such are the properties of *necessity, knowledge, justified belief* etc. In particular, those properties completely fail to satisfy ascent principles (and some of them even descent principles). But they do arguably satisfy consistency, iteration and soundness principles. Opaque necessity, knowledge, justified belief etc. are in conflict with classical logic just as truth is. The second argument is also easily *generalisable* to cover paradoxes concerning *sentences other than Liar sentences*, such as for example *Curry sentences* (in which case the relevant interaction of truth is with *implication* in the form of the *preservation* principle to the effect that, if an implication is true, then, if its antecedent is true, so is its consequent). In fact, these generalisations suggest that the general principle yielding as special cases all the relevant interaction and soundness principles is an abstract principle of *closure under logical consequence* suitably understood. Closure and iteration suffice for paradox. And, since closure in turn suffices for iteration “at level ω ” (see e.g. Field [2007], p. 115, fn 41), *in the end closure alone suffices for paradox* (in fact, as I’ve argued in Zardini [2013b]; [2014a], it suffices for forcing not only a revision of *classical logic*, but also a revision of the *traditional structural properties* preserved by many non-classical logics, thus pointing in the direction of the theory of truth developed in my works referenced in fn 1).

- Hartry Field. *Saving Truth from Paradox*. Oxford University Press, Oxford, 2008.
- Kit Fine. Prior on the construction of possible worlds and instants. In *Worlds, Times and Selves*, pages 116–168. Duckworth, London, 1977.
- Richard Heck. Truth and disquotation. *Synthese*, 142:317–352, 2004.
- David Makinson. The paradox of the preface. *Analysis*, 25:205–207, 1965.
- Vann McGee. *Truth, Vagueness, and Paradox*. Hackett, Indianapolis, 1991.
- Graham Priest. *In Contradiction*. Oxford University Press, Oxford, 2nd edition, 2006.
- Arthur Prior. *Time and Modality*. Clarendon Press, Oxford, 1957.
- Mark Richard. *When Truth Gives Out*. Oxford University Press, Oxford, 2008.
- Robert Stalnaker. Merely possible propositions. In Bob Hale and Aviv Hoffmann, editors, *Modality: Metaphysics, Logic, and Epistemology*, pages 21–34. Oxford University Press, Oxford, 2010.
- Timothy Williamson. Indefinite extensibility. *Grazer Philosophische Studien*, 55:1–24, 1998.
- Elia Zardini. Truth and what is said. *Philosophical Perspectives*, 22:545–574, 2008.
- Elia Zardini. Truth without contra(di)ction. *The Review of Symbolic Logic*, 4:498–535, 2011.
- Elia Zardini. Truth preservation in context and in its place. In Catarina Dutilh-Novaes and Ole Hjortland, editors, *Insolubles and Consequences*, pages 249–271. College Publications, London, 2012.
- Elia Zardini. It is not the case that [P and ‘It is not the case that P ’ is true] nor is it the case that [P and ‘ P ’ is not true]. *Thought*, 1:309–319, 2013a.
- Elia Zardini. Naive *modus ponens*. *Journal of Philosophical Logic*, 42:575–593, 2013b.
- Elia Zardini. Naive truth and naive logical properties. *The Review of Symbolic Logic*, 7: 351–384, 2014a.
- Elia Zardini. Context and consequence. An intercontextual substructural logic. *Synthese*, 2014b. Forthcoming.
- Elia Zardini. Getting one for two, or the contractors’ bad deal. Towards a unified solution to the semantic paradoxes. In Theodora Achourioti, Kentaro Fujimoto, Henri Galinon, and José Martínez, editors, *Unifying the Philosophy of Truth*. Springer, Dordrecht, 2014c. Forthcoming.

Elia Zardini. Restriction by non-contraction. *Notre Dame Journal of Formal Logic*, 2014d. Forthcoming.

Elia Zardini. The general missing from the hierarchy. In Fabio Bacchini, Stefano Caputo, and Massimo Dell'Utri, editors, *New Frontiers in Truth*. Cambridge Scholars Publishing, Cambridge, 2014e. Forthcoming.

Elia Zardini. The role of utterances in Bradwardine's theory of truth. *Recherches de théologie et philosophie médiévales*, 2014f. Forthcoming.

Elia Zardini. \forall and ω . ms, 2014g.