

Prolegomena to Any Consistent Theory of the Liar

Casper Storm Hansen

Draft of October 19, 2018

Abstract: This paper provides a solution to the liar paradox, in the sense of an answer to those who believe that the liar is both true and not true, based on the premise that language is conventional. According to David Lewis's theory of conventions, for a sentence to have truth conditions is for the language community in question to have a convention regarding the circumstances in which the sentence is appropriately assertible (in a certain sense). The power to institute language conventions does not come with the power to make a state of affairs both obtain and not obtain, and therefore the liar is not both conventionally appropriately assertible and not conventionally appropriately assertible. Whether it is one or the other is an empirical question that depends on contingent details about our conventions. I draw on Thomas Nagel's ideas about a *view from nowhere* to explain why our language psychology can make it seem that the liar ought to be true if and only if it is not.

I will propose a solution to the liar paradox. Solving this paradox is a feat that has been attempted many times before, but almost always with a focus on providing a detailed semantics for a self-referential language in which the liar sentence, in one way or another, avoids being both true and not true. I believe that such a semantics is neither necessary nor sufficient to this purpose. It is not necessary, for a paradox is a reason to believe in both halves of a contradiction, and a solution to a paradox is anything that undermines that reason; determining *which half* is incorrect is not necessary. And it is not sufficient, because in the absence of a deeper understanding of the liar phenomenon, *any* specific consistent theory of which sentences are true will seem ad hoc and implausible: a fact masterfully exploited by, in particular, Graham Priest (2002; 2006a; 2006b) to argue for dialetheism.

A scientist who gives a plausible account of the phenomenon of light without invoking the existence of a luminiferous aether is justified in rejecting the existence of the aether. Similarly, if a plausible account of the liar phenomenon is given without invoking the existence of sentences that are both true and not true, we would be justified in rejecting dialetheism (setting aside, as I will, any other reasons for believing in dialetheism to which the present solution cannot be extended). If the scientist, in the course of giving her account, begs the question against the aether theorist by assuming that the aether does not exist, that cannot be held against her if she eventually manages to account for all the data. Similarly, I will unapologetically beg the question against the dialetheist and assume that no state of affairs can

both obtain and fail to obtain. By doing so, I incur a debt which is repaid when it is shown that the liar phenomenon can be understood without recourse to that hypothesis.

To be more specific, to solve the liar paradox is to show that the truth condition of the liar, “The liar is not true”, is different from its naïve truth condition. The naïve truth condition of the liar is that the liar is not true; and if that were its actual truth condition, a contradiction would result. Separate analyses of the two concepts *truth conditions* and *naïve truth conditions* can reveal that they do indeed have distinct extensions in the case of the liar. Accordingly, we need to focus on understanding those two concepts. Crucial insights relevant to such an understanding can be found in the existing literature, but we have to look outside of the literature that explicitly deals with paradox: specifically, to David Lewis’s (1969) theory of conventions for an understanding of truth conditions, and to Thomas Nagel’s (1986) idea about a *view from nowhere*, to understand why we are naïve about them when we encounter the liar.

Thus, I will not present the kind of thesis that has been central to most previous attempts to solve the paradox, namely a thesis about what the actual truth condition of the liar is. It turns out to be sufficient to the purpose of accounting for the liar phenomenon without the hypothesis of dialetheism to understand what it is for a condition to be the truth condition of the liar; it is not essential to know *which* condition actually is. This means that this prolegomena is consistent with very many of the prior detailed theories about the liar, for instance the theory of Kripke (1975), which I will use as an example later. It is even consistent with a weak form of dialetheism, according to which two sentences can be true even though they contradict each other (according to their naïve truth conditions)—that form of dialetheism is a consistent theory about the world, even though it implies an inconsistency of sorts in a language used in the world. I will merely defend the claim that the state of affairs of a given sentence being true cannot both obtain and not obtain.

At the end of this paper, readers may still find each *individual* consistent semantics implausible as a candidate for the semantics of natural language, but my aim is to convince them that one of those semantics must nevertheless be correct (although it may be one that has not been described yet).

The key parts of this paper with respect to the analysis of truth conditions and naïve truth conditions are sections 1 and 5, respectively. The intermediate sections add details to the picture I am trying to draw. The previously published text in the literature about the liar paradox that is closest to my position is Chihara (1979); I compare the two in section 6, before concluding in the final section.

Before diving into the details, it may be helpful to give a rough indication of the location of this contribution in the philosophical landscape. I take myself to be aligned, at least in spirit and at least with respect to the more moderate aspects of his work, with the late Wittgenstein (1953) and opposed to Plato and Frege (1918). While it is never explicit, the liar paradox is almost always approached as if it was a matter of necessity what the truth

value of the liar sentence is, and that if we could just sneak a peek into the platonic-fregean world of abstract propositions, we would be able to see how the truth values are distributed over them. A central point of this paper is that the truth value of the liar sentence is contingent and a matter of human conventions that have been created to solve practical problems of communication. None of the classical texts offering solutions to the liar contains language to the effect that the author is merely making a claim about the actual world and that things could have been different. And it is rarely considered whether Kripke might be right about English, while Tarski managed to capture some principles that hold of French. I see dialetheism as the culmination of this implicit platonism: the “hardness” of logic, due to its otherworldly character, precludes any failures of truth conditions and naïve truth conditions being identical, and therefore the liar *must* be both true and not true.

1 Conventional truth conditions

The solution starts from the premise that the truth conditions of indicative sentences are conventional. To fix ideas, I will take my point of departure in David Lewis’s (1969) theory of conventions, of which I will therefore give a short summary.

Lewis understands conventions as solutions to coordination problems. An example of a simple coordination problem is that two people need to meet each other once a day, but it does not matter where. The first person will want to go where the second person goes, and the second person will want to go to where the first person goes. If they succeed in finding a place to meet regularly, and meet there because they expect the other person to show up there, they have instituted a convention. In general, Lewis analyzes conventions as arbitrary but self-perpetuating solutions to coordination problems: arbitrary in the sense that there are multiple solutions that, to all the involved agents, are roughly equally beneficial; and self-perpetuating because the choice of one of these arbitrary solutions at an earlier occurrence of the problem will tend to make agents choose the same solution in later occurrences.

In addition to characterizing conventions, Lewis explains how they can be initiated. Returning to the example, the two people do not have to arrive at their convention through explicit agreement; maybe they are unable to do so. They may meet where they do because they both take a chance showing up there, perhaps because they both expect that the other person is most likely to show up there, perhaps because they both expect the other person to consider it most likely that he himself will show up there, etc., through some finite number of higher-order beliefs. The meeting-place may not be intrinsically better than the alternatives; it may simply be that this place has some salient feature that induces the expectations, the expectations about expectations, and so on. Having met there once, the place becomes more salient for the purpose of subsequent meetings, and thus the different orders of expectations become stronger. A convention is in place.

A central insight of Lewis's is that his analysis extends to languages: they are conventional and can arise in basically the same way. Agreeing on the use of a language in a community is a solution to a coordination problem regarding communication. In the case of an indicative sentence, the convention consists of an agreement among a majority of the language users on the conditions under which it is appropriate to assert the sentence. Once we have a convention in place for an indicative sentence, I can, by uttering the sentence, inform you that the world is such that those conditions are satisfied. Thus, according to Lewis, for a sentence to be true is simply for the actual world to be among those possible worlds for which it has been conventionally agreed (in the actual world) that it is appropriate to utter the sentence in question—appropriate in an idealized sense where we disregard matters of relevance (the sentence may be true but of no interest to the conversation partners), the possibility that the agent has incorrect beliefs, the possibility that in the given situation it is morally obligatory to lie (e.g., to mislead a murderer), etc.¹

A language community has the power to institute conventions and to decide in which circumstances a sentence is true, i.e., conventionally appropriately assertible; but that does not give the community the power to make a state of affairs both obtain and not obtain. If they try to, they fail. If, for instance, they explicitly agree that “it is appropriate to assert of an object that it is *kratosk* if it is more than 1 metre high, and it is inappropriate to assert of an object that it is *kratosk* if it is less than 2 metres high”, then they have not made the sentence “*x* is *kratosk*” both true and not true in cases where *x* is 1.5 metres high. What is most likely to happen in that situation is that their “agreement” results in confusion and fails to govern their linguistic behavior, meaning they have failed to create a convention for the appropriateness of asserting “*x* is *kratosk*” in those circumstances. It is also possible that they will all go on to act upon, say, the first half of the agreement, so that the agreement, although not followed, nevertheless causes a convention to be instituted. But it is not possible that their agreeing has the effect that they all act in compliance with a contradictory convention when confronted with an object 1.5 metres high.

Similarly, a language community does not have the power to make a state of affairs both obtain and not obtain by assigning conventional truth conditions to “The liar is not true”. In no language community can that sentence be both conventionally appropriately assertible and not conventionally appropriately assertible at the same time. Thus, in our language community,

¹As mentioned, my use of one particular theory of conventions is just to fix ideas; what I have to say will be independent of very many of the specific details of that theory. For instance, many dissenters have put forward examples that show that Lewis's definition of convention is either too broad or too narrow; see, e.g., Burge (1975) and Miller (2001). These examples do not affect my conclusions. I also do not depend on the concrete claims that are made above about how agents manage to arrive at a particular solution; only that there is *some* mechanism different from explicit agreement that allows them to do so, for example the one explained in Skyrms (1998). I do depend on the primary purpose of language conventions being coordination with respect to communication, but they do not have to be solutions to coordination problems in the specific game-theoretic sense that Lewis relies on. Alternatives that could be substituted for Lewis's account include Gilbert's (1989).

the liar is not both true and not true.

2 Possible language conventions

We have now, with Lewis, accounted for the *property* of truth. It is a property that a sentence has if there are conventional conditions for the circumstances under which it is appropriately assertible, and those conditions as satisfied. But what about the truth *predicate*? It is certainly a predicate that has been introduced into language with the intention of being able to describe the situation of something having the property of being true.² The idea was to have a sentence like “ ϕ is true” have the property of being true iff ϕ has the property of being true and to have a sentence like “ ϕ is not true” have the property of being true iff ϕ does not have the property of being true. However, intentions are not always fulfilled, and in this case the liar sentence prevents this intention from being fulfilled. The truth conditions for a sentence ϕ are the conditions under which ϕ has the property of being true, and not necessarily the conditions under which a sentence in which the truth predicate is applied to a term referring to ϕ has the property of being true; nor are they necessarily the opposite of the conditions under which a sentence in which the negation of the truth predicate is applied to a term referring to ϕ has the property of being true. The members of a language community can have the intention of making it so, but since that intention is inconsistent, they cannot succeed.

The obvious next question to ask is what the actual truth conditions for sentences of those forms then are, in full generality. Unfortunately, when an intention is frustrated, there is rarely a way to predict what happens instead merely from knowledge of the intention. If someone intends to build an immovable object and create an unstoppable force, I know that his intention will be not be fulfilled, but there are several possibilities for what will happen instead, namely that he builds an immovable object but fails to create an unstoppable force; that he creates an unstoppable force but not a immovable object; and that he manages neither. Similarly, I do not know what the actual truth conditions of “ ϕ is true” and “ ϕ is not true” are in all cases. But I can describe several possibilities, and that is what I will do in this section. To argue for dialetheism from the existence of an intention to create an immovable object and an unstoppable force, one would have to argue that it is more plausible that both are created than that one of the three above-mentioned alternatives ensue. Similarly, to make dialetheism plausible by appealing to the existence of the liar, one would have to argue that it is more plausible that language users manage to make a state of affairs both obtain and fail to obtain than that events unfold as in one of the possible worlds described below (or as in one of innumerable other consistent alternatives).

²*Pace* the deflationists who are so impressed by the fact that the truth predicate can be used to make blind generalisations that they attempt to reduce the semantics of the truth predicate to this utility, and claim that there is no property of truth. See, e.g., Beall (2009, 2), who writes “Our device ‘true’ [...] was not introduced to name any feature of the world”. After a language, initially without a truth predicate, had been introduced into the world, there was a new feature of the world to be named, namely the property of being true.

The first possible world I will describe will be designated the “Simple World”. In the Simple World, language develops gradually. In the initial, primitive stage, there are a few fixed expressions, each associated with its own individual convention governing its use. But very quickly, the users of the language develop conventions that provide general rules for how words contribute to truth conditions of sentences. According to the language users themselves, one of these rules is that a sentence of the form “ ϕ is true” is true iff “ ϕ ” denotes a true sentence. The language users similarly claim to have an equally general rule for a negation construction that turns any true sentence into a non-true sentence and vice versa. Since they also have very flexible rules for how to form names, they are able to construct a liar sentence, i.e., a sentence that seems to them to say of itself that it is non-true, and for which it follows from the alleged general rules that it is true if and only if it is not true.

That the sentence actually is true iff it is not is impossible, so the language users’ beliefs about the rules of their own language are mistaken. Fortunately for them, social conventions do not function in the same stringent way as mathematical definitions and axioms. Thus, the fact that the rules they believe in and try to follow cannot be fully satisfied does not have the effect that the entire system breaks down as it (in a manner of speaking) does when a mathematician formulates a system of inconsistent axioms. Rather, the conventions they have actually succeeded in instituting are a close approximation to those they think they have instituted: almost every time they have a need for their truth predicate or their negation, the convention works for the specific case in the way that they think it always works. For instance, when someone says “‘ $2+2 = 3$ ’ is not true”, her listeners understand what truth conditions the speaker intended for that sentence to have, and that she is trying to communicate that they are satisfied. Only in the case of their liar sentence and some similar sentences that are quite special and rarely arise in normal conversation do the rules not apply. Their conventions do not allow those who hear their liar asserted to infer what is claimed to be the case. (The distinction between “good” and “bad” cases is not sharp. When someone says, for example, “I always speak the truth”, most listeners will have a pretty good idea of what the speaker intended to convey, but some very logically minded listeners will notice the self-reference and insist that it is not clear exactly what that sentence means.) So in fact, but unbeknownst to them, the people of Simple World do not have a convention that covers the appropriateness of asserting their liar sentence. The same applies to their negation of their liar sentence (i.e., the sentence that, according to the rules that hold in most cases, would be true iff their liar is not). As being a true sentence is to be a conventionally appropriately assertible sentence, this means that their liar sentence and its negation do not have the property of being true.

In all the other possible worlds I will describe, there *are* conventions governing the truth conditions of their liar sentence (but apart from this, and the existence of conventions governing the truth conditions of other “pathological” sentences, these worlds are similar to Simple World). In Revision

World,³ the convention is that their liar is appropriately assertible every second day and not appropriately assertible on the remaining days. In Dialetheism World,⁴ the convention is that their liar, along with its negation, is always appropriately assertible (again, this is different from the strong dialetheism I am arguing against: that the same sentence can be both appropriately assertible and not appropriately assertible). In Kripke World,⁵ the opposite convention is in force: both the liar and its negation are not appropriately assertible. Thus, the Kripke World resembles the Simple World, in that those two sentences are not true; but for different reasons. In the Kripke World, both are not true because there is a convention that says so, whereas in the Simple World, the reason is that there is no convention for the two sentences.

I will add more details to the descriptions of these possible worlds in sections 3, 4, and 6, but for now they are elaborate enough to serve the purpose explained above. That is, it should be clear that these are ways a human language community could have developed. That undermines the claim that there are sentences that must be both true and not true.

I will try to narrow down the location of the actual world a little more below, because the details of how our language actually functions are of philosophical interest, even though they could easily have been different. It is important to emphasize, however, that I am less certain about the hypothesis I put forward concerning how the actual world happens to be than I am about the more central claims of this paper, and that the solution to the liar paradox is not undermined if I get some of the details about these contingent facts wrong. The reason I think it is of philosophical interest to locate the actual world with more precision is that the solution so far amounts to pointing out that the truth conditions of the liar cannot be what we naïvely think they are. There is something very obvious and trivial to that observation, and that gives rise to a puzzle that needs to be addressed: if the solution is so trivial, why has the paradox seemed so hard? What is it that makes it seem like the liar ought to be true if and only if it is not true? To answer that question, we need to understand the psychology of naïve truth conditions. I will get to that in section 5. The next section contains a more gentle introduction of the concept of naïve truth conditions, which will be sufficient for an assessment, in section 4, of the relative pros and cons of living in Simple World, Revision World, Dialetheism World, and Kripke World.

3 Naïve truth conditions

Simple World, Revision World, Dialetheism World, and Kripke World all have something in common, namely that—with one small exception to be explained in the next paragraph—there is no difference between how the *ordinary* language users of those worlds use language, or what they know

³Inspired by Gupta (1982) and Belnap and Gupta (1993).

⁴Inspired by Priest (2006b).

⁵Inspired by Kripke (1975).

and believe, explicitly and implicitly, about language. In particular, most ordinary language users in any of the worlds will say, if asked, that a sentence of the form “ ϕ is true” is true iff ϕ has the property of being true, and that negation-constructions flip the truth value of the negated sentence. They will therefore also be ready to declare that the truth condition of “The liar is not true” is that the liar is not true (at least until they notice the problem with that declaration). We can call that “truth condition” its *naïve truth condition*. In general, by “naïve truth conditions” we mean the truth conditions that a person who is naïve with respect to self-referential (and other non-well-founded) sentences, but who knows everything else about the language’s semantics, would think a given sentence has.

The one small exception is that in Revision World, Dialetheism World, and Kripke World, unlike in Simple World, ordinary language users consider logicians to be authorities on the meanings of logical connectives, the truth predicate, and self-referential constructions and are willing to defer to them on these matters, just as ordinary language users in every one of those worlds consider biologists authorities on the meanings of terms for plant species and are willing to defer to them in that regard. And because of the linguistic division of labour that makes experts’ conventions about specialised terminology the conventions of the entire language community,⁶ this means that the truth conditions and the truth values of the liar sentences can be different in the four possible worlds, even though the naïve truth condition is the same. So even though the typical inhabitant of, say, Dialetheism World has the same immediate reaction to the liar as the typical inhabitant of Kripke World, the liar is conventionally appropriately assertible in the former but not in the latter, as the ordinary language users will be willing to admit once they have consulted their respective expert logicians.

In Revision World, Dialetheism World, and Kripke World, the logic experts noticed at some point in history that the naïve truth conditions were not sufficient by themselves to give actual truth conditions to the liar and similar “ungrounded” sentences,⁷ and that there was therefore gaps in the conventions of their societies, in the form of indicative sentences that lacked truth conditions.⁸ So they decided to fill those gaps by instituting conventions for the truth conditions of those sentences. They did not start from scratch; they just added some rules that still revolved around the naïve truth conditions. For instance, the logicians of Revision World decided that when naïve truth conditions make reference to the truth values of other sentences, the

⁶See Putnam (1975, 227–229).

⁷I do not intend the use of the term “ungrounded” to provide a *precise* characterization of the sentences for which the naïve truth conditions are sufficient by themselves to provide truth conditions. One can debate whether or not a sentence such as “The liar is true or the liar is not true” is ungrounded; and similarly, one can imagine possible worlds in which the naïve truth conditions of that sentence suffice by themselves to provide truth conditions, and other possible worlds in which they do not. (And one might instead describe the former kind of worlds as some wherein the naïve truth conditions do not suffice by themselves, but some extra convention has been instituted to nevertheless make them identical to the truth conditions of the sentence.) The vagueness of some of the vocabulary I use simply reflects that there are many possible (very similar) language conventions.

⁸I am using the term “indicative sentence” for any sentence that has naïve truth conditions associated with it, even if it does not have actual truth conditions.

truth values that must be used to determine whether those naïve truth conditions are satisfied are the truth values of the previous day. So, in Revision World, the naïve truth condition of the liar is that the liar is not true, while the truth condition is that the liar was not true yesterday.⁹

Notably, this means that in Simple World they do have *a* convention that concerns the liar, just not a convention for the truth conditions of the liar. The existence of a convention that provides naïve truth conditions for the liar can be used to account for an otherwise perplexing psychological fact of Simple World: that, even though the sentence has no truth conditions, it seems to the language users there to be meaningful, and it seems to them that it makes sense to reason from it as an assumption. The naïve truth conditions deliver that. For most of the sentences that a typical language user encounters in the course of normal communication, the naïve truth conditions suffice by themselves to actually make those sentences (fully) meaningful and allow for correct reasoning using the usual rules of inference; so, when they do not suffice by themselves (and a dialetheian or Kripkean or revision or some other auxiliary convention would be needed), they give the illusion of doing that. The distinction between naïve truth conditions and truth conditions thus gives us a first hint as to why the paradox seems so hard. As mentioned, I will approach this question head-on in section 5.

4 Pragmatic assessment of possible conventions

What are the pragmatic consequences of the different conventions that have been considered above? Given that ordinary language users do not know about the difference between truth conditions and naïve truth conditions, it would, for each indicative sentence, be beneficial if that sentence's truth conditions were satisfied (i.e., the sentence were true) iff its naïve truth conditions were satisfied. When this biconditional is satisfied, language users can communicate in the way they (naïvely) assume they can. When the left-to-right direction is satisfied, hearing the sentence in question being asserted gives a listener exactly the information he or she thinks it gives (assuming the utterer is honest and well-informed). And when the right-to-left direction is satisfied, asserting a sentence conveys exactly the information the utterer thinks it conveys.

In Kripke World, where one of the theories described by Kripke (1975) is adopted as a convention,¹⁰ the expert logicians have, in a manner of

⁹Above, in the *kratosk* example, I mentioned that the actual effect of the inconsistent agreement might be that a consistent convention is instituted (say, because of a shared psychological propensity to prioritize what one is told earlier over what one is told later). Similarly, we might have instituted a convention similar to the one in Revision World, Dialetheism World, or Kripke World in the actual world, even though no explicit, consistent decision has been made by a council of logicians commanding deference from the language community at large. I merely use this fiction to make the distinction between truth conditions and naïve truth conditions vivid, before characterizing the latter with more precision below.

¹⁰For present purposes it does not matter which one, except that what I say here does not apply to the supervaluation version that quantifies over only maximally consistent interpretations.

speaking, erred on the side of making too few sentences true. That is, the left-to-right direction is satisfied for all sentences. The consequence of the failure of the right-to-left direction for “ungrounded” sentences is a certain expressive weakness. For instance, their liar sentence cannot be used to communicate that their liar does not have the property of being true, even though it not being true is its naïve truth condition.

In an attempt to avoid this kind of expressive weakness, the logicians of Dialetheism World have devised a convention that ensures that the right-to-left direction is satisfied for all indicative sentences. But in the presence of a liar, both directions cannot be universally satisfied at the same time, so there are failures in the opposite direction. This means that the ordinary language users in Dialetheism World cannot trust true sentence to always convey the information they think they convey.

In Revision World, the biconditional can fail in both directions. On days when the liar is true, it fails in the left-to-right direction for that sentence, and on the remaining days it fails in the other direction. However, when the revision sequence has entered into a loop,¹¹ both directions are satisfied for the stably true sentences, and the set of stably true sentences is a superset of the set of sentences that are true according to the Kripkean convention.¹²

The biconditional is a version of the T-schema, and in these examples it fails.¹³ Such failures are not violations of some deep logico-metaphysical necessity that forces dialetheism; they just result in some communication problems.¹⁴

Let me reiterate that I have not here compared contenders for the title of solution to the liar paradox. That is, I have not compared epistemically possible solutions. I have compared metaphysically possible conventions. The solution resides, so to speak, on a higher level: it consists in realizing that certain consistent languages really are possible conventions in spite of their shortcomings, while a language containing sentences that are both true and non-true is not. When considered as a proposed solution with a claim to platonic necessity, the shortcomings that each of these languages has may be seen as a reason to reject that solution—and since they all have such shortcomings, one may reject them all and end up in the embrace of dialetheism. But when considered as a possible convention, the shortcomings

¹¹See Gupta (1982, 45).

¹²Gupta’s (1982) revision theory can also be used as inspiration for another convention, namely one where a sentence is true iff it is stably true according to that theory, instead of changing truth value on a daily basis. This convention satisfies the left-to-right direction for all sentences, like the Kripkean convention, and of course the set of stably true sentences is still a superset of Kripke’s true sentences. (Disclaimer: neither this convention nor the other Gupta-inspired convention should be interpreted as reflecting Gupta’s actual view.)

¹³Another version is this: a sentence is true iff its truth conditions are satisfied. It never fails. The term “T-schema” may be ambiguous between these two versions and that is the reason why I have refrained from using it up until now, even though using it might have made comparisons between this paper and the existing literature easier by introducing the former as arguing for the well-known thesis that there are exceptions to the T-schema.

¹⁴In addition to being a problem for communication between one person and another, the liar may constitute a problem for thought. I am not sure how severe that problem is. One factor that would be relevant to determining its severity is the extent to which thought takes place in a compositional language (see below).

have no such force: there is nothing unusual about there being no perfect convention available.¹⁵

I have on a couple of occasions encountered the reaction to this conventionalist solution that conventions are irrelevant and that the problem for the liar sentence has nothing to do with the fact that the meaning of the sentence is conventional. I can see why people would react that way. Conventions are a complex phenomenon with many aspects, and most of those aspects are indeed irrelevant to the solution (while perhaps the most salient aspects of conventions for those who react in that way). For instance, it is not important to the solution that conventions are created by *groups* of people. Nor is it important that conventions govern recurring situations. It is not even essential that conventions have alternatives (although I think that helps with making my point more vivid). What is essential is that conventions are created by humans with limited power. Our power to create a convention for the conditions under which sentences are appropriately assertible is not a power that allows us to make the state of affairs of the liar being appropriately assertible both obtain and not obtain. If we try to, we simply fail and end up with an imperfect language. The dialetheists claim that if the liar is not true, it must also be true, but they fail to explain who or what the enforcer of this *must* is and how he/she/they/it can make it so hard that it can defy a metaphysical law, because mere mortals certainly cannot. It would seem that, lurking in the background, there is an assumption of a platonic-fregean realm of propositions that are so perfect that they can ignore the most basic rules restricting everything else—and if I am wrong about that, then we are owed an alternative explanation.

Given certain restrictions, every possible convention will result in some communication problems of the kind described.¹⁶ The kind of conventions we have considered all (1) keep in place the sentences of English and their naïve truth conditions, (2) allow the truth conditions of a sentence to be identical to its naïve truth conditions when that sentence is grounded,¹⁷ and (3) only provide an auxiliary convention for what the truth conditions are to be otherwise. So, should we get rid of one or more of those restrictions, and make more radical changes to our language to overcome these communication problems? From a pragmatic perspective, I think the answer is “no”, because I suspect that the liar problem is a necessary side-effect of the compositionality that makes language efficient and easy to learn; and that any solution aimed at eradicating it altogether would involve sacrifices that far outstrip the gains. However, we could take a less pragmatic and more

¹⁵If someone were to insist, against my advice, on calling, e.g., the language described by Kripke a “solution” to the paradox, they should characterize my position as one of pluralism with respect to liar paradox solutions.

¹⁶A similar point is made by Maudlin (2004), who also uses the concept of a sentence being conventionally appropriately assertible (in his parlance it can be “appropriate [in a ‘sui generis’ sense] to assert” a sentence according to “rules” we “lay down”—see page 95). However, Maudlin discusses different ways of changing the conventions about what is appropriately assertible against the background of a conception of truth that is independent of conventions. Thus, his full theory is based on a distinction without a difference, but I agree with much of what he says about appropriate assertibility.

¹⁷With this qualification, concerning Revision World: after a sufficient number of days have passed.

principled view of the matter and focus on in-principle expressive strength, disregarding the usefulness of languages to actual human beings. Then, the question to be asked would be whether it is possible to construct a language convention such that *for any possible state of affairs, there is a sentence that is true iff the state of affairs obtains*. Self-reference would certainly be a challenge for any attempt to achieve such a convention, for among the states of affairs that need to be expressible in the language are the states of affairs of the language’s various sentences being true. But it is not clear to me that the revenge phenomenon implies the impossibility of this.¹⁸ First, because the $\forall\exists$ -structure of the italicised clause above renders the goal more modest than what is typically aimed at by those who have constructed languages that can accommodate the liar. And second, because the existence of a liar sentence is not a necessary feature of sophisticated languages, not even ones that contain sentences about their own semantics.

We could call the problem of finding such a convention the *liar expressibility problem*. It is closely related to the liar paradox, but not identical. I am offering a solution to the latter, but the former I have only diagnosed.¹⁹

5 The view from nowhere

What the naïve truth conditions *are* (i.e., the extension of the concept) is well known: they are the conditions that are (1) used in each individual valuation in the supervaluation versions of Kripke’s theory, (2) applied to the facts at one stage of a revision sequence to determine the next stage, (3) used in a Tarskian language after having been restricted to non-semantic facts and semantic facts concerning languages at lower levels of his hierarchy, (4) appealed to in any claim that some theory about the semantic paradoxes comes with a revenge paradox, and (5) the reason for the liar paradox in the first place, because they are assumed to be the liar’s actual truth conditions. But how and why did *those* conditions end up being the naïve truth conditions? In other words, why do we think, when we are naïve, that those conditions are the truth conditions, when it is clear not only that they aren’t

¹⁸I set aside the option that there might be uncountably many different states of affairs.

¹⁹In addition to the characterisation in terms of the biconditional, a different characterisation of *how* the various conventions fail may be illuminating. They all squeeze the set of sentences that have their naïve truth conditions satisfied between two approximations: a subset of that set, and a superset of that set. In Kripke World, the subset is the set of true sentences, and the superset is the complement of the set of false sentences. In Dialetheism World, the subset is the set of true-only sentences, and the superset is the complement of the set of false-only sentences. In Revision World, they are the set of stably true sentences and the complement of the set of stably false sentences, respectively. But note that in all cases, the complement is relative to the universe of indicative sentences of English, and not to the more limited universe of sentences that are admitted as meaningful according to the formal language in question. So a Tarskian (1944) convention also fits in, for each level of the hierarchy: the subset is the set of true sentences, and the superset is the complement of the false sentences. (And the recent proposal by Scharp (2013) does too: the subset is the set of descending true sentences, and the superset is the set of ascending true sentences. This becomes most clear in section 8.3.) Because the set of sentences that have their naïve truth conditions satisfied is convention-relative, it is impossible to “squeeze” so much that the subset and the superset meet and coincide with the target set.

in the case of the liar, but that they couldn't be? That is the puzzle raised by our solution to the paradox, which has to be worked out to make that solution fully satisfying. To do that, we will need a bit of inspiration from Thomas Nagel and his book *The View from Nowhere* (1986).

The relevancy of that book to the liar phenomenon is quite straightforward. True indicative sentences are supposed to represent the state of the world. The totality of true sentences of a given language is supposed to be the most comprehensive and objective representation of the world that the language has to offer. If the language contains self-referential sentences, then the objective representation is also supposed to cover the part of the world that is the language itself. And Nagel's book is all about attempts to objectively represent totalities that include the representer and the means of representation.

Let me first give a brief synopsis of this book about the human pursuit of objectivity.²⁰ Human beings start out with a highly subjective conception of the world, i.e., a conception that is largely shaped by the time and place the person happens to occupy in the world, the specifically human perceptual apparatus, and the human way of life. However, we have gradually managed to attain a view on the world that is somewhat removed from that initial, limited vantage point. For example, a subjective human description of something in the world will often contain color terms; but we have found out that a more objective description can be given in terms of electromagnetic radiation. That is a description which would also make sense to other intelligent species that are quite different from ours, and it can form part of an explanation of our subjective color experiences.

The limit of the transcendental impulse towards an objective understanding of the world is referred to by Nagel as the *view from nowhere*. It is an understanding of the world completely independent of worldly facts about the subject itself, qua observer, that distort and limit how the subject views the world. The word "view" should be understood in a very wide sense, as Nagel applies the concept of a view from nowhere not only to the most obvious area of philosophy, epistemology, but also to philosophy of mind, free will, and ethics.

It is a common human mistake to assume that we command a view from nowhere when we do not. One example is a naïve realism according to which the world has only the kind of properties we can directly perceive (e.g., colors) and not others (e.g., electromagnetic radiation). A second example is the assumption of free will, in its most naïve form, when it comes to deliberating which of several possible actions to take. We pretend that we can take "action from nowhere". In order not to make the deliberation seem absurd, we have to avoid seeing the *I* that deliberates and is about to act as being merely a part of the natural order. We need (the illusion of) autonomy, and that involves seeing ourselves as agents who can influence the causal stream of events from a place that is nowhere in it.²¹

²⁰Be warned that this summary is a bad substitute for the book itself. A full understanding of the content of this section may therefore require reading the book.

²¹Priest (2002) similarly discusses the liar and related phenomena in terms of the dual

The same kind of naïvety has led to the semantic paradoxes. Our naïve beliefs about what the truth conditions of the sentences of English are presuppose that we can take a view from nowhere with our language; that we do not have to take into account that the language itself is part of the world to be described by it; that the effects of instituting a language convention with which to describe the world can be kept isolated from the world. Let me try to explain. Language conventions have an “input” and an “output”: the input consists of the facts relevant to whether the truth conditions of a given sentence are satisfied, and the output is the fact that the sentence is true or not true (as the case may be). If the output effects of instituting a new language convention were isolated to an observation point outside of the totality of facts, then *any* condition about the state of that totality could be used as truth condition. That is, if our linguistic practice took place in a view-from-nowhere location separate from the world, we could “view” the world to determine whether the condition was satisfied and “record” that in the truth or non-truth of the sentence in question, without affecting the world.²² For instance, we could use the criterion that the liar is not appropriately assertible in the world and, if that criterion were satisfied, record that fact in the truth of “The liar is not true” in the view-from-nowhere location; and doing so would have no effect on the world and hence not undermine itself. But in fact, the output must have an effect in the world that supplies the input, and that makes it impossible for certain conditions to be our actual truth conditions.

The objectivisation impulse is built into the way we as a language community create conventions for naïve truth conditions for a sentence. It is (implicitly) assumed that the community can choose *any* subclass of the class of all possible worlds to be the worlds wherein the sentence is true. But in any possible world where a sentence is true, it has to be true. That is, in any possible world where a sentence is true because its truth conditions are satisfied, so that it has to be true to live up to its job of describing the objective facts, it also has to be true in the sense of it being a fact among the totality of objective facts of the world that the sentence is true. Thus, the subclass cannot contain any possible worlds in which the sentence itself is not true.

If the language community nevertheless attempts to assign truth conditions in a way that presupposes that they are standing outside the world that they are creating a language to describe, then they may fail. The truth conditions they have attempted to assign may not become the actual truth conditions. (This can come as a surprise to them if their attempt at assigning truth

concepts of *closure* and *transcendence*, with which he makes points connected to Nagel’s theme of totalities of some given kind of entities versus entities of that kind outside of that totality. However, whereas Nagel recognizes the conflict as genuine, Priest insists that it is no worse than that it can be accommodated by a *dialetheia*.

²²A more familiar way to describe this is to say that if, *per impossible*, we were in possession of a language that were a meta-language relative to all languages, including itself, then any condition could be used as the truth condition of any sentence in that language. But whereas this description is in terms that are particular to the semantic paradoxes, the description in terms of the view-from-nowhere metaphor unifies the psychological explanation for our confusion concerning the liar with Nagel’s explanations for a wide range of phenomena.

conditions is via compositional rules, so they do not think about the truth conditions of each individual sentence.) What they attempt to assign as truth conditions is what I have called “naïve truth conditions”. Inspired by Nagel, we can instead refer to them using the more descriptive name “view-from-nowhere truth conditions”. The view-from-nowhere truth conditions of a sentence are the conditions under which that sentence would be true if all facts (including facts about truth values) were given independently of the facts that are the effects of the language conventions that govern the sentence. Let me unpack that a bit. Let a possible world W have a language community that functions as in the possible worlds we have described above, in the respect that truth conditions are given via view-from-nowhere truth conditions. This language community has made an attempt to assign truth conditions to a sentence ϕ . The facts on which the truth value of ϕ depends are facts in W , and the truth value of ϕ is also a fact in W . But imagine that we instead had two copies of W , W_1 and W_2 , and that the facts on which the truth value of ϕ depends are facts in W_1 , while the resulting truth value of ϕ creates a fact in W_2 (W_1 is the “viewed world” and W_2 is the “point from nowhere”). Then, the truth conditions that the language community attempted to assign could be applied even if they are self-referential. (For instance, if the liar is not true in W_1 , the truth conditions that the English-speaking language community has attempted to assign to the liar could be applied to that fact to make the liar true in W_2 .) The view-from-nowhere truth conditions are the truth conditions that would be used in that fictive scenario.

However, as this is a fictive scenario, view-from-nowhere truth conditions can be used directly as truth conditions (the conditions under which the sentence is actually conventionally appropriately assertible) only in those cases where it does not matter that the viewed world and the world from which it is viewed are identical. For instance, the view-from-nowhere truth conditions of “Grass is green” can be used as the truth conditions of that sentence because the fact of the truth of the sentence does not affect the greenness of grass. In the remaining cases, the truth conditions have to be given in some other way (although they can be defined *in terms of* the view-from-nowhere truth conditions, as they are in Revision World, Dialetheism World, and Kripke World).

So, to sum up how the concept of view-from-nowhere truth conditions can be used to understand the case of the liar: There is an objective fact about whether the liar is conventionally appropriately assertible or not. If it is, the view-from-nowhere truth conditions of the liar are not satisfied, and if it is not, they are satisfied. In the latter case, the liar itself would naïvely seem to be the right tool for someone who knows that the liar is not conventionally appropriately assertible to communicate that fact to someone who doesn’t know it. But implicit in that naïve thought is the presupposition that they have a “nowhere place” in which to communicate, because (per the assumption that the liar is not conventionally appropriately assertible) the liar is not available to them to serve that purpose, according to the conventions of their world.

So the answer to the puzzle of why it seems like the liar ought to be true iff

the liar is not true is that, because of our tendency to naïvely assume that we command a view-from-nowhere perspective on the world, we think of language as a detached medium with which we can describe an objective world that exists independently of it—as if the liar in the detached medium could be true to reflect the non-truth of the liar in the unaffected objective world (or the other way around). But everything must be, by definition, within the world; and thus the liar marks one of the limits of the transcendent impulse. This way of thinking has formed our way of instituting language conventions. The way our conventional truth conditions are primarily given is based on a presupposition of ideal objectivity (similar to the presupposition of ideal objectivity that was shown to be mistaken when the observer effect was discovered in physics), and we have either not formed language conventions that take into account that this is an over-idealisation, or we are not fully aware of the conventions we have formed that do so. Since this presupposition is implicit and not a conscious part of human beings’ use of language,²³ it has seemed to us that there were, or ought to be, identity between truth conditions and view-from-nowhere truth conditions; and that is why the liar paradox has seemed to be an *insoluble*. The liar has clear view-from-nowhere truth conditions, so it has been assumed to have clear truth conditions. However, every attempt to specify those truth conditions and the resulting truth value(s) of the liar has in some way or other conflicted with our intuitions about how truth conditions work. But those intuitions arise from the false expectation that view-from-nowhere truth conditions can be identical to truth conditions in all cases, and that the same world can serve both as the viewed world and as the viewpoint from nowhere when they do. Realizing this ought to dispel the psychological *pull* that the paradox can exert, even after one has been made aware of the more elementary elements of the solution presented above.

As the reader has probably guessed, the content of this section has not simply been an elaboration on my stipulations about the Simple, Revision, Kripke, and Dialetheism worlds; it has been an attempt to describe the actual world. In addition, I want to put forward the claim that everything I have stipulated to be the case for *all* of those four possible worlds is the case for the actual world, and that the actual world is located somewhere in their neighborhood. However, I will not go any further. I do not know what, if any, conventions are in fact in force regarding the circumstances in which indicative sentences are appropriately assertible, in those cases where their view-from-nowhere truth conditions are not sufficient to determine that by themselves.²⁴ In particular, I do not have a solution to the *descriptive liar problem*: determining what the actual truth conditions of the liar are, according to existing conventions.²⁵ But that was not my goal. My goal

²³Or, at least, it rarely is. Philosophers working on the liar are sometimes exceptions. That seems, for instance, to be the case in this quote from Soames (1999, 6, my emphasis), who, commenting on Tarski’s theory, writes that its most serious problem is “the irresistible urge to violate the hierarchy’s restrictions on intelligibility in the very process of setting it up and describing it. We tend to forget this because *we imagine ourselves taking a position outside* the hierarchy from which it can be described.”

²⁴It is common for participants in a convention not to be able to account exactly for what the convention is; see Lewis (1969, 63–64).

²⁵In addition to the liar paradox, the liar expressibility problem, and the descriptive

was to convince the reader that our language conventions could easily have developed in such a way that we ended up in one of the four possible worlds described, and that the solution to the liar paradox is not sensitive to the contingent differences between them. The solution (in a narrow sense) to the liar paradox is simply that the truth value of the liar is governed by our language conventions, and that no language convention can make it the case that the liar is both appropriately assertible and not appropriately assertible.

Let me conclude on the relationship between truth conditions and naïve truth conditions. We had to drive a wedge between these two concepts, because if the former had the same extension as the latter, we would have a contradiction. Giving a fully satisfying solution to the semantic paradoxes requires understanding each concept well enough to realize *why* they are not co-extensional. We have done so by (extensionally) identifying each of them with some other concept. Following Lewis, we have identified truth conditions with the conditions under which the given sentence is conventionally appropriately assertible, and drawing on Nagel, we have identified naïve truth conditions with view-from-nowhere truth conditions.²⁶ And even though we have not determined the conditions under which a self-referential sentence is actually conventionally appropriately assertible (only that it could possibly be as it is in Simple World, Revision World, Dialetheism World, or Kripke World), we have convinced ourselves that they must be different from the view-from-nowhere truth conditions.

Among the virtues of this solution to the liar paradox is theoretical coherence: the theory builds on work by Lewis and Nagel that has much wider areas of application than semantic paradoxes. And their books (Lewis 1969; Nagel 1986) were written, it seems, without the authors ever thinking about those paradoxes, which means that ad hoc-ness is avoided. Also, the theory is simple: a great deal can be explained merely by making the distinction between truth conditions and view-from-nowhere truth conditions.

liar problem, one can also distinguish the *liar inference problem*: which inference rule used in the typical deduction of a contradiction using the liar should be abandoned? As some take this to be the primary version of the liar paradox, let me briefly address it. We need to split the notion of valid inference, as defined by necessary truth preservation, in two. There are valid inferences in the sense that conventionally appropriate assertibility is guaranteed to be preserved, and there are valid inferences in the sense of the preservation of satisfaction of the view-from-nowhere truth conditions. In the latter case, it is the T-in and the T-out rules that are not (generally) valid: the view-from-nowhere truth condition of a sentence $T\phi$ saying that another sentence ϕ is true is that ϕ is conventionally appropriately assertible, and not that the view-from-nowhere truth conditions of ϕ are satisfied: so $T\phi$ may have its view-from-nowhere truth condition satisfied, while ϕ does not, or vice versa. In the former case, the question of which inference rule has to go depends entirely on the conventions that are in force.

²⁶That the naïve truth conditions are the view-from-nowhere truth conditions is a contingent fact, which is due to a certain psychological propensity. It is conceivable that humans could have been naïve about self-reference in some other way, or not naïve about it at all.

6 Comparison with Chihara’s position

This paper has a lot in common with Chihara (1979).²⁷ The part of the solution presented in section 1 is essentially identical to Chihara’s: both solve the paradox by arguing that truth must be like the *kratosk* predicate. But there is also an important difference between truth and *kratosk*, which has to be accounted for to make the solution convincing, and which Chihara failed to account for. Concerning truth, there is an insistent inner voice that says “There is a definite way the world is, and therefore a clear separation of the indicative sentences into those that describe that world (or a part or aspect of it) and those that don’t; so there is a definite set of sentences that are true; so there *should* be a legitimate predicate with *all* of the properties we tend to attribute to the truth predicate”. Something similar is not the case for the *kratosk* predicate. There is no intuition that the world of physical things is (clearly and exhaustively) separated into those that are *kratosk* and those that are not that survives after it has been pointed out that the definition of “*kratosk*” is inconsistent. To account for the difference, we need two things: first, the concept of view-from-nowhere truth conditions, which sentences predicating truth of some entity have, while sentences predicating *kratosk*-ness of some entity do not (in all cases); and second, the observation that instituting language conventions in order to be able to describe the world changes the world. Then, we can answer the insistent inner voice: yes (setting vagueness and the conceivable correctness of certain kinds of anti-realism aside), there is a definite set of the indicative sentences that describe the world, in the sense of having their view-from-nowhere truth conditions satisfied. But we cannot make exactly those sentences true (i.e., conventionally appropriately assertible), on pain of changing the set of sentences that describe the world.

Chihara refers to his own position as “the inconsistency view of truth”, but rather than a logical inconsistency, there is something more akin to a performative contradiction involved in the paradoxical uses of the truth predicate. It is not that something can be both true and not true according to what is intended with the truth predicate (as given by how the predicate contributes to view-from-nowhere truth conditions), in the way that something can be both *kratosk* and not *kratosk* according to what is intended with the *kratosk* predicate (as given by its two-part definition). Rather, it is that the “propositional content” (as given by the view-from-nowhere truth conditions) of a sentence predicating truth or non-truth of something can conflict with a presupposition of appropriately asserting that sentence: namely, that the language community in question has made it true in the circumstances where it is asserted, similarly to how the propositional content of a sentence saying that I am asleep conflicts with a presupposition of me asserting it, namely that I am awake.

Now that we have seen that there is an important difference between the truth predicate and the *kratosk* predicate, it will be useful to develop a system for classifying the types of “defectiveness” of predicates. Let us

²⁷The same holds of Eklund (2002) and the descriptive parts of Scharp (2013).

call a predicate *T-determined* if its application to a referring singular term of the language always results in a sentence that is, according to existing conventions, appropriately assertible or, according to existing conventions, not appropriately assertible. Its opposite is a *T-underdetermined* predicate: there is a referring singular term of the language such that the predicate's application to it results in a sentence whose appropriate assertibility is not covered by existing conventions. It should be obvious that the predicate "sratosk" will be T-underdetermined in the Simple World if it is introduced by the explicit stipulation "it is appropriate to assert of an object that it is *sratosk* if it is more than 2 metres high, and it is inappropriate to assert of an object that it is *sratosk* if it is less than 1 metre high". But the kratosk predicate is also T-underdetermined, at least in the Simple World. There, the language users have attempted to make kratosk T-*overdetermined*, but they cannot possibly succeed: when they place inconsistent obligations on themselves in the case of sentences such as "*x* is kratosk" where *x* is 1.5 metres high, they simply create confusion and fail to establish a convention that covers the case.

The truth predicate is also T-underdetermined in the Simple World, so we need another distinction to differentiate between truth and kratosk: a predicate is *view-from-nowhere-truth-conditions determined* or *VFNTC-determined* if its application to a referring singular term of the language always results in a sentence that has view-from-nowhere truth conditions according to existing conventions; and it is *VFNTC-underdetermined* if not. If we again set vagueness aside, the truth predicate is VFNTC-determined, because a sentence of the form " ϕ is true" has its view-from-nowhere truth conditions satisfied if ϕ is an appropriately assertible sentence according to existing conventions, while its view-from-nowhere truth conditions are unsatisfied according to existing conventions if not. The kratosk predicate, on the other hand, is VFNTC-underdetermined.

Let us apply these conceptual distinctions to the predicate "the view-from-nowhere truth conditions of . . . are satisfied". This predicate needs to be commented on, because on the one hand, I may have given the impression that it comes closer to doing what the truth predicate is supposed to do than the truth predicate itself, while, on the other hand, anyone who has heard of the revenge phenomenon will know that there must therefore also be a problem with this predicate hidden somewhere. The problem is, in fact, not hidden very well. It is located in the first place you would look, namely in the sentence "The view-from-nowhere truth conditions of this sentence are not satisfied". When you try to figure out whether it is true, you end up in a circle. You consider whether an imagined subject enjoying a view-from-nowhere perspective on the world would judge the sentence to be true. But that subject would have to consider the very same question. So in a version of Simple World where the predicate has been introduced in the same way as I have introduced it in this paper, there is no convention that settles the appropriateness of asserting the sentence. That is, the predicate is T-underdetermined in Simple World (because in Simple World a sentence lacks truth conditions iff its negation does). But not only that: not even view-from-nowhere truth conditions have been assigned to the sentence, so

the predicate is also VFNTC-underdetermined. The conclusion is that in the present way of categorizing types of defectiveness, the predicate “the view-from-nowhere truth conditions of . . . are satisfied” is different from the truth predicate, but similar to the kratosk and sratosk predicates, with respect to VFNTC-determinedness. So any naïve hopes one might have had of this predicate solving the liar expressibility problem are dashed.

Am I nevertheless so fortunate that every sentence I have asserted in this paper is true according to actual conventions iff its view-from-nowhere truth conditions are satisfied, so that my attempt at communicating with you, the reader, has not been obstructed? I do not know, because I do not have a detailed account of actual conventions. However, referring back to the end of section 2, where I mentioned that the solution would not be undermined if I got certain contingent facts wrong, I want to add that it could also survive my failure to communicate some of those facts. For this solution, revenge is at most a communication restrainer, not a paradox reviver.

7 Dialectical position vis-à-vis the dialetheist

The purposes of this paper have been, firstly, to understand the liar phenomenon independently of features of our actual language conventions that could easily have been different; and, secondly, to argue against strong dialetheism, according to which one and the same sentence both has the property of being true and fails to have that property. The primary point in favour of dialetheism is the alleged failure of every consistent theory about the semantic paradoxes. The job of pointing out these purported failures individually, for a large number of the proposed consistent theories, has been undertaken with admirable diligence by Priest. However, Priest’s recurring line of attack against most of those theories consists in (explicitly or implicitly) identifying a sentence ϕ such that it follows from the theory that the view-from-nowhere truth conditions of ϕ are satisfied, while ϕ is not true according to the theory; and concluding on that basis that the theory is wrong.²⁸ Having separated satisfaction of view-from-nowhere truth conditions from truth, we are now in a position to recognise this type of argument as fallacious.

It should be clear that the conventionalism regarding truth, which I have advanced as the solution to the liar paradox, has nothing to do with a general relativism. To avoid that, we only need to distinguish properly between sentences and states of affairs: there are sentences (written, spoken, thought, etc.) with conventional truth conditions, and in addition, there are facts (obtaining states of affairs) that, in conjunction with the truth conditions, determine truth values. The solution is consistent with a strong realism, meaning that facts are objective and most of them are independent of what humans think and do. The separation of truths and facts is essential to avoiding the liar paradox: it is what makes it possible to claim that

²⁸See Priest (1984; 1987; 1993; 1995; 2002; 2005; 2006b; 2007; 2010; 2012), Priest and Routley (1989), and Beall and Priest (2007).

the truths can fail to provide a perfect representation of the facts in accordance with view-from-nowhere truth conditions. Then again, it *is* possible to maintain that there is a perfect match between truths and facts, if one is ready to reject the “fact-bivalence” (that states of affairs either obtain or not, and not both) that I have assumed. It is possible if one is a glut theorist who claims that facts are exactly as glutty as the sentences that are about them.²⁹ I have not *proved* that position wrong. I am simply claiming that it is less plausible than my combination of fact-bivalence with an account of the liar phenomenon based on the fallibility of the human endeavour to create a system with which to represent the facts.

References

- Beall, J. (2009). *Spandrels of Truth*. Oxford University Press.
- Beall, J. and G. Priest (2007). Not so deep inconsistency: a reply to Eklund. *Australasian Journal of Logic* 5, 74–84.
- Belnap, N. and A. Gupta (1993). *The Revision Theory of Truth*. The MIT Press.
- Burge, T. (1975). On knowledge and convention. *Philosophical Review* 84, 249–255.
- Chihara, C. (1979). The semantic paradoxes. *Philosophical Review* 88, 590–618.
- Eklund, M. (2002). Inconsistent languages. *Philosophy and Phenomenological Research* 64, 251–275.
- Frege, G. (1918). Der Gedanke. In *Beträge zur Philosophie des deutschen Idealismus*, Volume 1. Verlag der Kenerschen Buchhandlung. Translated as “Thoughts” in M. Black, V. Dudman, P. Geach, H. Kaal, E.-H. W. Kluge, B. McGuinness and R. H. Stoothoff (Transl.) (1984), *Collected Papers on Mathematics, Logic and Philosophy*. Basil Blackwell.
- Gilbert, M. (1989). *On Social Facts*. Routledge.
- Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic* 11, 1–60.
- Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy* 72, 690–716.
- Lewis, D. (1969). *Conventions*. Harvard University Press.
- Maudlin, T. (2004). *Truth and Paradox: Solving the Riddles*. Oxford University Press.
- Miller, S. (2001). *Social Action: A Teleological Account*. Cambridge University Press.
- Nagel, T. (1986). *The View from Nowhere*. Oxford University Press.
- Priest, G. (1984). Semantic closure. *Studia Logica* 43, 117–129.

²⁹This is suggested explicitly in Priest (2006a, 51–54).

- Priest, G. (1987). Unstable solutions to the liar paradox. In S. J. Bartlett and P. Suber (Eds.), *Self-Reference: Reflections on Reflexivity*. Martinus Nijhoff.
- Priest, G. (1993). Another disguise of the same fundamental problems: Barwise and Etchemendy on the liar. *Australasian Journal of Philosophy* 71, 60–69.
- Priest, G. (1995). Gaps and gluts: Reply to Parsons. *Canadian Journal of Philosophy* 25, 57–66.
- Priest, G. (2002). *Beyond the Limits of Thought* (second ed.). Oxford University Press.
- Priest, G. (2005). Review of “Truth and Paradox” by T. Maudlin. *Journal of Philosophy* 102, 483–486.
- Priest, G. (2006a). *Doubt Truth to be a Liar*. Oxford University Press.
- Priest, G. (2006b). *In Contradiction* (second ed.). Oxford University Press.
- Priest, G. (2007). Revenge, Field, and ZF. In J. Beall (Ed.), *Revenge of the Liar: New Essays on the Paradox*. Oxford University Press.
- Priest, G. (2010). Hopes fade for saving truth. *Philosophy* 85, 109–140.
- Priest, G. (2012). Read on Bradwardine on the the liar. In C. D. Novaes and O. T. Hjortland (Eds.), *Insolubiles and Consequences: Essays in Honour of Stephen Read*. College Publications.
- Priest, G. and R. Routley (1989). The philosophical significance and inevitability of paraconsistency. In G. Priest, R. Routley, and J. Norman (Eds.), *Paraconsistent Logic: Essays on the Inconsistent*. Philosophia Verlag.
- Putnam, H. (1975). The meaning of ‘meaning’. In *Philosophical Papers, volume 2: Mind, Language and Reality*, pp. 215–271. Cambridge University Press.
- Scharp, K. (2013). *Replacing Truth*. Oxford University Press.
- Skyrms, B. (1998). Salience and symmetry-breaking in the evolution of convention. *Law and Philosophy* 17, 411–418.
- Soames, S. (1999). *Understanding Truth*. Princeton University Press.
- Tarski, A. (1944). The semantic conception of truth. *Philosophy and Phenomenological Research* 4, 341–376.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Suhrkamp. Translated as “Philosophical Investigations” (Third ed.) (2001). Blackwell Publishing.