

# The obvious solution to the liar paradox given mainstream background assumptions

Casper Storm Hansen

Draft of August 2, 2019

ABSTRACT: This paper argues that, given certain mainstream assumptions, the solution to the liar paradox is straightforward, once a few simple distinctions have been made, namely between states of affairs and truths, between the truth property and the truth predicate, and between truth conditions and so-called naïve truth conditions. The solution is really what Chihara would call a “diagnosis”, for no claim is made about whether the liar sentence has the truth property or not. That is because the question of whether gap theory, revision theory, or dialetheism (to name just a few options) is correct is an empirical question inessential to the philosophical solution to the paradox.

The liar paradox arises from the sentence “the liar is not true”, when that sentence is called “the liar”. If it is true, it seems to follow that it is not true, and if it is not true, it seems to follow that it is true. Since it must be either true or not true, it seems to follow that it is both true and not true.

One possible reaction to this argument is to accept the conclusion that the liar is both true and not true. That is, in effect, to accept that the liar paradox constitutes a philosophical challenge of such an extreme character that it can be used to justify what would otherwise be considered an outrageous view. The principle that is discarded to accommodate the liar is the principle that any state of affairs either obtains or fails to obtain, and not both: with the state of affairs of the liar being true considered to be a counterexample. Rejection of this principle constitutes a revision of standard views that could hardly be more fundamental.

An argument could be made that any coherent alternative account of the liar should be preferred, if it required less extensive revisions to a mainstream web of beliefs. But as it happens, I do not think it necessary to accept such a low standard for alternative accounts. For, as I will argue, there is an obviously correct alternative account—not obvious in the sense that anyone could easily figure it out for themselves, but obvious in the sense that when you see it, you should not be in doubt of its validity. I realize that this seems

like an absurd and arrogant claim, given how many previous attempts to solve the paradox there have been. However, most such solution attempts have been formulated under the influence of a certain dogma about what a solution is supposed to accomplish. I will return to what that dogma is; but if we free ourselves from it, a simple solution in line with standard background views is readily available.

Another possible reaction to the liar argument is to deny the last premise: that the liar is either true or not true. However, that would imply rejection of the same fundamental principle as before, just with a *gap* instead of a *glut*, which is no less outrageous. By a “state of affairs” I mean a way the world can be, independently of descriptions of that way in language. So the principle that any state of affairs either obtains or not—never both, never neither—is absurd to deny.<sup>1</sup> Let us take it for granted from here, and see if we can solve the paradox without calling it into question.

The liar being true is a state of affairs. So it either obtains, or it does not. This is the fact that seems to lead to the contradiction; but by distinguishing between the truth property and the truth predicate, we can avoid that contradiction. Let me first discuss the truth property. It is obvious what this property consists in: it is a property of sentences, and a given sentence has it if and only if (1) truth conditions have been assigned to the sentence by convention and (2) those conditions are satisfied. While it is possible to disagree with this analysis, you would have to be pretty far out of the mainstream to do so. Indeed, it is so minimalistic that almost anyone should be able to accept it. I suppose you are more likely to object that the analysis is superficial and that there is much more to be said about (1) and (2). But that does not matter, for the analysis is sufficient to solve the paradox.

If (1) and (2) are the case for the liar, then it has the truth property or, in other words, the state of affairs of the liar being true obtains. If at least one of them fails, then not.

While truth conditions are assigned to sentences by convention, they are typically not assigned individually, but rather collectively through rules of compositionality. That is, semantic conventions are typically about how words contribute to sentences’ truth conditions, and thus only indirectly about the truth conditions themselves. And the semantic convention about how a given word contributes to truth conditions is typically uniform, in

---

<sup>1</sup>One might think that whether a state of affairs obtains may be vague, so that there are more than the two options I mention. I happen to believe otherwise, but I do not need that assumption. For the sake of argument, I can allow that a state of affairs obtains to degree 0.5 and fails to obtain to degree 0.5, so to speak. If you believe that this is possible, just replace the liar with the definite liar in the rest of this paper, where “the definite liar” refers to “the definite liar is not definitely true”. What I meant to rule out is that a state of affairs obtains to degree 1 and fails to obtain to degree 1, or that the sum in some other way is different from 1.

the sense that there is a (relatively) simple rule for how that word contributes to truth conditions across sentences. But obviously, that is not so by necessity: we *could* decide, collectively, that from this day forward, the truth condition of “the meat loaf is not in the oven” is that the Moon is full. Exceptions to the simple, compositional rules are—although generally unwanted—possible. Again, if you deny this, you have some very deviant views.

Next, let us look at the truth predicate. It is obviously a predicate that has been added to language with the intention of facilitating description of (and enquiry about, and speculation about, etc.) states of affairs consisting of a sentence having the truth property. That is, the idea was that a sentence of the form “ $\phi$  is true” should have the truth property if and only if  $\phi$  has the truth property. That was the simple, compositional rule intended by the language community. While it would probably require a considerable amount of work to make that claim entirely precise, it should be obvious that there is a reasonable sense in which it is correct.

In the same sense, the language community intended for the negation to contribute to truth conditions by the simple compositional rule that a sentence of the form  $\neg\phi$  has the truth property if and only if  $\phi$  does not have the truth property.

Thus, it was indirectly intended that “the liar is not true” should have the truth property if and only if “the liar is not true” does not have the truth property. By this being the intention *indirectly*, I mean that it would have to be satisfied for the two rules that were *directly* intended to hold without exception to both do so.

However, it cannot be the case that “the liar is not true” has the truth property if and only if “the liar is not true” does not have the truth property, because the liar being true is a state of affairs that either obtains or not. The flip-side of the fact that it is within our powers to make exceptions to general compositional rules is that it is not within our powers to make general compositional rules exception-free, if doing so would conflict with the basic metaphysical law that any state of affairs either obtains or not. If the language community had attempted—say before the general truth predicate and the general negation had been introduced—to make the sentence “aghafyi”, with no meaningful proper parts, true if and only if “aghafyi” is not true, then they would have failed. And the language community also failed when it, indirectly and without realizing it, intended for the liar, which *does* have meaningful proper parts, to be true if and only if the liar is not true.

This is quite obvious. Of course, admitting it implies admitting that we humans are imperfect when it comes to designing a linguistic system: our intentions are not always realized. But it is hardly news that we are im-

perfect in general, so it should not be so difficult to accept that we are also imperfect in this specific regard.

Acknowledging this specific type of imperfection means that we need to make a distinction between the actual truth conditions of sentences and their “intended” truth conditions: that is, the truth conditions they would have if the general compositional rules were exception-free. I will call the latter “naïve truth conditions”. The (actual) truth conditions of a sentence are the conditions such that necessarily, given the actual language conventions, the sentence has the truth property if and only if those conditions are satisfied.

Typically, of course, truth conditions and naïve truth conditions coincide. It is equally obvious that they do not and cannot, in the case of the liar sentence. For if they did coincide for the liar, then the liar would be true iff the liar were not true.

That the liar’s truth conditions and its naïve truth conditions are different is, I submit, the solution to the liar paradox. With that realization, we see that the liar being true is not, despite appearances to the contrary, a state of affairs that both obtains and fails to obtain. That is all that is required to solve a paradox.

But I hear an objection: “You have not told us whether the liar has the truth property or not. Any proposed solution to the paradox must give an answer to that question.” I beg to differ. That is the dogma I alluded to earlier: that the hunt for a solution to the liar paradox should be conceived of as the hunt for a truth value for the liar sentence. Countless, increasingly complex semantics for formal languages that contain a liar have been proposed with the aim of providing that truth value. But figuring out the truth value is orthogonal to solving the paradox, which consists merely of recognizing that there is a difference between actual truth values and naïve truth values. If you do the former without doing the latter, there are two options (assuming that you respect the constraint that states of affairs either obtain or not). The first is that you conclude that the liar has the truth property, and then you will, in effect, be claiming that the actual truth condition of the liar is satisfied (because otherwise it would not have the truth property) and that the naïve truth condition is not (because the naïve truth condition is that the liar does not have the truth property). However, you will not have made that distinction, so it will seem to your critics that you are contradicting yourself. The second is that you conclude that the liar does not have the truth property: in effect claiming that the actual truth condition of the liar is not satisfied but its naïve truth condition is. And that has the same effect. In fact, this has happened so many times that one could get the impression that semantic theories adhering to the principle that any state of affairs either obtains or not are hopeless. Indeed, it is for this reason that Priest (1987) calls the search for a successful consistent theory a degenerating research

program.

Understanding the conceptual difference between naïve truth conditions and truth conditions, and realizing that they must diverge in the case of the liar, are the philosophical insights needed to solve the philosophical puzzle that is the liar paradox. *How* they diverge is a contingent matter that depends on the fine details of our language conventions. Whatever the truth value of the liar is—and whatever exact way we have collectively failed in our doomed attempt at making our compositional rules exception-free—it could have been the opposite and we could have failed in a different way. Maybe the truth values *are* opposite in, say, English and Chinese.

To illustrate and exemplify what I have said so far, I will nevertheless comment on three of the most popular theories about the precise semantics of the liar—Kripke’s (1975) gap theory, revision theory (Gupta 1982), and dialetheism (Priest 2006)—not with the aim of declaring which are correct and which are not, but of determining which are *possibly* correct. It turns out that the answer for all three is the same: some versions are possibly correct, and some could not possibly be (given mainstream assumptions).

A moderate version of Kripke’s theory merely holds that neither “the liar is not true” nor “the liar is true” has the truth property. That is certainly a possibility, for “the liar is not true” having the truth property and “the liar is true” having the truth property are two different states of affairs, and thus might both fail to obtain. The explanation would be, in effect, that the actual truth condition of the liar is as follows: the liar does not have the truth property *and* the liar is grounded. That is one way for the truth condition of the liar to differ from the naïve truth condition in a manner that makes for a possible convention.

There is also an extreme version of Kripke’s theory that goes like this. Semantic facts, such as which truth values sentences have, are metaphysically special. They are (by necessity?) determined by a recursive “process” that begins with an incomplete world in which all states of affairs concerning semantics neither obtain nor fail to obtain, and ends with an incomplete world where *some* states of affairs concerning semantics neither obtain nor fail to obtain. This is a way to avoid my conclusion that there are exceptions to compositional rules and that truth conditions and naïve truth conditions diverge. For according to this theory, the relevant instances of the biconditionals that express the compositional rules are not false, for there is no fact of the matter about whether they are true. And the truth condition of the liar really is that the liar is not true; there is just no fact of the matter about whether that condition is satisfied. While this is coherent, it is also preposterous: there can be no justification for bringing obvious metaphysical principles into doubt just to avoid the banal admission that human systems

of communication are imperfect.<sup>2</sup>

One version of revision theory would be that the liar is true *now*, false *now*, then true again *now*, and so on. This, too, is a *possibly* correct theory (metaphysically at least, if not epistemically). It implies that the actual truth condition of the liar is that the liar was not true two seconds ago, so the principle that any state of affair either obtains or not (at any given time) is upheld. Also, because the liar sentence is in the present tense, its truth value would never match its naïve truth condition.

As far as I know, no one actually believes the above version of revision theory. Instead, a revision theorist might say<sup>3</sup> that the liar sentence, rather than having the semantic value *true* or *not true*, has the entire revision sequence as its semantic value, or something along those lines. I do not consider this to be a possibly correct theory, because I do not know what the social praxis of communication would have to be like for that theory to be correct. (And that is in spite of me being quite liberal about what I would count as a fact of the social praxis of communication. For instance, I would be willing to consider it a fact of the social praxis of communication that we, in some idealized sense, aim at asserting  $\phi$  and not asserting  $\neg\phi$  if  $\phi$  is an unknowable truth.)

Dialetheism, like Kripke's theory, has a moderate version: in this case holding merely that "the liar is not true" and "the liar is true" both have the truth property. This is possible (and, like all possibilities, a violation of our intentions with the truth predicate and the negation). Likewise, there is an extreme version according to which "the liar is not true" both has and fails to have the truth property, which like the extreme version of Kripke's theory avoids the consequence that human language is imperfect, but at an outrageous and disproportionate price.

I also want to mention the possibility that we simply haven't attached truth conditions to the liar: that we have not instituted the relevant kind of convention. Then, by (1) above, the liar would not have the truth property, and hence its naïve truth condition would be satisfied. But that would not make it true.

Keeping our system of naïve truth conditions fixed, every possibility for what the corresponding actual truth conditions are implies that there is some mismatch between the truth conditions and naïve truth conditions of some sentences. Because they do not distinguish between these two, many attempted solutions to the paradox are instead presented as claims about the correct *logic* being different from classical logic. That would be a revision to mainstream views almost as startling as the idea that there are states of

---

<sup>2</sup>If I interpret them correctly, Kripke himself represents the moderate version and Field (2008) the extreme one.

<sup>3</sup>Based on conversation and personal correspondence.

affairs that neither obtain nor fail to obtain. But the liar provides as little reason for the former revision as for the latter, for it does not constitute a challenge to this much more reasonable view: classical logic is the correct logic of *actual* truth conditions. For instance, if the actual truth condition of one sentence is the negation of the actual truth condition of another sentence, then exactly one of them is true. The revision of logic that would be called for if, say, the moderate version of Kripke's theory happens to be correct with respect to English is superficial, as it would only pertain to naïve truth conditions.

Whatever the exact nature of the mismatch between truth conditions and naïve truth conditions is, it leads to some communication problems. That is bad, for sure, but it is a pragmatic problem for people who want to communicate, and should not be confused with the philosophical problem of the liar *paradox*. I claim that the latter is solved, and that part of the solution is to understand the nature of the former problem.

I maintain this claim even though the pragmatic problem of communication may affect our ability to communicate about the subject of the liar itself. One may try to come up with revenge problems for my proposed solution, for instance by formulating versions of the liar that employ the term "naïve truth conditions" or the term "truth property". However, any such attempt would at most show that pragmatic communication problems also affect some uses of those terms. There is no way to produce a revenge liar that can be used in a sound argument in favor of the conclusion that there is a state of affairs that both obtains and fails to obtain.

In the previous literature on the liar paradox, the theory that comes closest to my position is Chihara's (1979) so-called inconsistency theory of truth. A comparison is therefore in order. According to Chihara, the assumptions that lead to paradox are *prima facie* convincing because they seem to have been made true by fiat, and derive further plausibility from the fact that they hold in the vast majority of cases. Nevertheless, one of them is in fact false. I essentially agree that far. (My inessential dissent is that the assumptions *may* all be *true*; but one of them does not have its naïve truth conditions satisfied.)

Yet, Chihara's diagnosis is only partially correct. This is revealed by, among other things, his claim of essential similarities between the liar paradox and the "paradox" that arises from the predicate "glub", when one attempts to define it as follows: "glub" applies to  $x$  iff  $x$  is not a mouse and "glub" fails to apply to  $x$  iff  $x$  is neither a mouse nor different from  $x$ . For any given thing, you can infer that that thing both is and is not a glub using only this "definition".

I disagree with Chihara's claim. There are at least two important differences between the truth predicate and the liar, on the one hand, and "glub" and

sentences of the form “ $x$  is glub”, on the other. First, the naïve truth conditions of the liar are either satisfied or not, while “ $x$  is glub” has not even been assigned any naïve truth conditions. Second, while the truth predicate is intended (in the sense used above) to correspond to what is a perfectly fine property, there is no property at which the attempted definition of “glub” takes aim. Because of these differences, Chihara is mistaken when he claims (p. 603) that self-reference is not essential to the family of paradoxes that has the liar as a member. Unless there is self-reference (or more generally, non-well-foundedness) to prevent it, there will not be a situation in which naïve truth conditions are given in terms of a predicate that is intended to correspond to a bona fide property and yet cannot be the actual truth conditions.

Chihara also concludes that the T-schema is incorrect. That conclusion is itself at most halfway correct, for the T-schema as he states it (p. 605) is ambiguous: “A sentence is true if, and only if, what is said to be the case by the sentence is in fact the case”. What is “said to be the case” by a sentence? We can distinguish between two versions of the T-schema, namely one in which we answer that question in terms of truth conditions, and another in which we answer it in terms of naïve truth conditions. Given the former answer the T-schema is not only correct, but necessarily so (at least if we set aside the fact that instances of the T-schema are themselves *sentences*, so they may suffer from a mismatch between truth conditions and naïve truth conditions). What Chihara says about the T-schema is only correct under the latter interpretation.

#### REFERENCES

- Chihara, C. (1979). The semantic paradoxes. *Philosophical Review* 88, 590–618.
- Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press.
- 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.
- 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. (2006). *In Contradiction* (second ed.). Oxford University Press.