



Lying and self-defeating prophecies

Wolfgang Freitag¹ 

Received: 20 June 2023 / Accepted: 23 December 2023
© The Author(s) 2024

Abstract

The paper disputes the common view which holds that to lie is, essentially, to assert a disbelieved proposition. It shows by reference to self-defeating prophecies that one can lie by asserting a believed proposition, and by reference to self-fulfilling prophecies that one can be truthful by asserting a disbelieved proposition. It concludes that lying is, essentially, asserting a proposition believed to be false conditionally on the occurrence of the assertion.

Keywords Lying · Self-fulfilling prophecy · Self-defeating prophecy · Conditional belief

Funding Open Access funding enabled and organized by Projekt DEAL.

Lying is, essentially, saying something believed to be false. Yet what is it to say something believed to be false? The classical conception construes the disbelief required for lying as directed at the *propositional content* of the assertion:

I am grateful to the members of the Research Colloquium of the Chair of Theoretical Philosophy at the University of Mannheim, in particular Felix Bräuer and Nadja-Mira Yolcu, for some very useful comments. Special thanks to Alexandra Zinke for the many discussions on the topic. Christopher von Bülow did the copyediting.

✉ Wolfgang Freitag
wolfgang.freitag@uni-mannheim.de

¹ Institute of Philosophy, University of Mannheim, 68131 Mannheim, Germany

(DEF) *A* lies if and only if there is a proposition *p* such that

(CON1) *A* asserts *p*, and

(CON2) *A* believes that *p* is false.¹

It is generally acknowledged that (DEF) is but a basic definitional scheme which may call for substantial modification and development: some authors claim that there are nonliteral lies and thus replace (CON 1) with a more complex clause,² and others hold that there need to be further conditions, stating, e.g., that *A* intends to deceive the addressees,³ or that *p* not only is believed to be, but actually is, false.⁴ Yet whatever differences there exist with regard to the necessary refinements of the basic model, it is almost common ground that it is impossible to lie without disbelieving *p* and impossible to be truthful with respect to a disbelieved *p*: to lie is essentially to assert a disbelieved proposition.⁵

This paper argues that the classical conception misconstrues the disbelief required for lying: lying is not to assert a disbelieved proposition, but to assert a proposition that is disbelieved given one asserts it.⁶ The difference and its significance for the concept of lying is subtle and easily missed, since it does not show up for ordinary cases. It is best brought out with the help of examples in which the truth-value of the assertion depends on the occurrence of the assertion. I will therefore discuss *future-directed* assertions—prophecies—and show by reference to *self-defeating* prophecies that it is possible to believe that an assertion will be false if it occurs, to make the assertion, and yet to not disbelieve the asserted proposition.⁷ I argue that this refutes (DEF): there are lies without disbelief of the asserted proposition (in Sect. 1). Cases of *self-fulfilling prophecies*, on the other hand, show that it is possible to make an assertion, believe that

¹ This formulation follows closely that of Viebahn (2021, p. 292). Other proponents of (DEF) and variants thereof are Fallis (2009), Stokke (2018), and Pepp (2019).

² A recent example is Viebahn (2021), who provides what he calls a “commitment-based approach,” articulated in the following complex substitute for (CON 1): “(L1) *A* performs a communicative act *C* with *p* as content; (L2) with *C*, *A* intends to communicate *p* to *B*; (L3) with *C*, *A* commits herself to *p*” (Viebahn, 2021, p. 300).

³ See, e.g., Chisholm & Feehan (1977) and Williams (2002, p. 96).

⁴ See, e.g., Carson (2010).

⁵ In the interest of straightforward presentation, I ignore in the main text occasional voices which require, instead of a belief that *p* is false, only that *A* believes that *p* is *probably false* or that *A* does *not believe* that *p* (cf., e.g., Carson, 2010, p. 39). Note, however, that the examples in Sect. 1 also refute *this* view. I also ignore positions according to which belief does not play a role at all: recently, Krstić (2023) has claimed, referring to cases of “transparent delusion,” that lying is *not* asserting something *believed* to be false, but asserting something *judged* to be false. It may therefore be worth emphasizing that my argument is directed not so much at the kind of mental state or act required, but rather at the *content* of that mental state or act. I think that Krstić’s liar would have to judge, not that *p* is false, but that *p* is false *under the condition that she asserts p*.

⁶ To keep things simple, I ignore in the main text the question *when* the required disbelief would have to occur. But observe that we may understand the examples in the main text in such a way that the agents are not in possession of the required disbelief at any time during the assertion. See also fn. 11 below.

⁷ Pepp (2018, Sect. 8) discusses cases which exhibit an *epistemic* dimension of ‘self-influence’: the assertive act does not influence the truth or falsity of *p*, but rather is *evidence* for her *doxastic state* with regard to *p*. As Pepp’s cases are quite complex and the issues raised largely independent of my present concerns, I will not discuss them in this paper.

the assertion will be true if it occurs, while also disbelieving the asserted proposition. Such cases pose another problem for (DEF): a speaker may be perfectly truthful while disbelieving the asserted proposition (Sect. 2).⁸

I conclude that a lie is not an assertion of a *proposition* that is believed to be false, but an *assertion* that is believed to be false. In Sect. 3, the idea will be cast in terms of conditional disbelief. Lying is not asserting a disbelieved proposition *simpliciter*, but asserting what is disbelieved conditionally on the occurrence of the act of assertion. (DEF) will be revised accordingly.

Note that my argument does not require any finesse with regard to the illocutionary force involved in the potential lie, or some particular view on the relevance of the semantic–pragmatic distinction. The cases to be discussed in this paper involve straightforward assertions of some semantic content *p*.⁹ I will therefore keep things simple and develop my argument with respect to (DEF) alone. The application of the argument to extant modifications of (DEF) will be left to the interested reader.

1 Lying without disbelief

In all cases to be discussed, (CON 1) is fulfilled: the agent asserts some proposition. Whether the agent commits a lie then depends, according to (DEF), on the fulfilment of (CON 2) alone: it depends on whether the agent disbelieves the asserted proposition. I will first show that there are cases of lying in which the agent does not disbelieve the asserted proposition: (CON 2) is not fulfilled and thus (DEF) is too narrow. To prepare the debate, consider the following case:

PROPHECY. If nobody interferes, next Saturday's lottery will yield a random sequence of numbers. But by means of some manipulating device the lottery can be rigged: switch it on, enter a certain sequence of numbers, and the lottery mechanism will invariably spit out exactly this sequence of numbers. Let *P* be the proposition *that the combination XYZ will win in Saturday's lottery* and assume the chance of *P* in an unmanipulated lottery to be enormously small. In full awareness of the situation, Prophet asserts not-*P*.

Is Prophet's assertion a lie? As long as she does not believe that the machine is being rigged, Prophet believes that it is very unlikely that XYZ will win in next Saturday's lottery. Prophet has a very high credence in not-*P*, and we may therefore assume that she does not believe *P*. (CON 2) is not satisfied and Prophet's assertion not a lie according to (DEF). This may well be the right verdict. To make things interesting, consider the following development.

SELF-DEFEATING PROPHECY. Lottie has come in control of the manipulating device. She would never use it, except to take revenge on Prophet. If Prophet

⁸ I stick with the label, though I think it is descriptively inadequate: there is no element of *self*-fulfilment in 'self-fulfilling' prophecies, and no element of *self*-defeat in 'self-defeating' prophecies. A 'prophecy,' a prediction, is fulfilled if and only if it turns out true. What makes a 'self-determining' prediction true or false is, however, not the prediction itself, but some future state of affairs determined by the prediction.

⁹ Throughout this paper, I assume that predictions, even self-fulfilling or self-defeating prophecies, are straightforward assertions.

asserts not- P , Lottie will switch on the manipulating device, type in XYZ and thus guarantee that Prophet's assertion is false. If Prophet asserts P , Lottie will make sure that some other numbers will be drawn. And if Prophet should not make any lottery-related assertion, Lottie will not intervene and let chance decide. Fully aware of Lottie's schemes, Prophet nevertheless asserts not- P .

Is Prophet lying in this case? She knows that any prediction will be self-defeating: that if she actually predicts some lottery outcome, her prediction will be false. In particular, she knows that, if she asserts not- P , the number combination XYZ will be drawn and thus not- P will be false. If, therefore, she sets out to predict not- P , she sets out to lie. And if she manages to bring about the assertion, she succeeds in lying. Making a prophecy known to be self-defeating is to commit a lie. This, however, is not the result of (DEF)—at least not unconditionally.

Whether, according to (DEF), Prophet is lying depends on whether the respective instance of (CON 2) is met: it depends on whether Prophet believes that not- P is false. But Prophet does not disbelieve not- P , unless she also believes that she *asserts* not- P : only when she believes that she asserts not- P does she believe Lottie to interfere and P to turn out true. Thus, according to (DEF), whether Prophet is lying depends on whether she believes in the occurrence of the self-defeating prophecy.

Of course, we would expect Prophet to know about her assertive acts in most situations. And for them, (DEF) may well yield the right results.¹⁰ Yet it is not hard to construct a case in which Prophet ends up asserting that not- P but does not believe that she accomplishes the assertion. Just suppose that she attempts to make the assertion but also knows that she has some severe speech impediment: she knows that she may well not succeed and hence that she may not assert not- P .¹¹ Prophet does not believe that she accomplishes the self-defeating prophecy and hence does not believe that Lottie will intervene. As Prophet then does not disbelieve what she asserts, she is not, according to (DEF), a liar. This is, however, the wrong result. Prophet may not know or believe that she will end up asserting not- P . And therefore she may not know or believe that she commits a lie. But she knows that, *if* she asserts not- P , she will say something false. Thus, if she asserts not- P , she will *be* a liar, even if she does not know that she is because she does not know that she has accomplished the assertive act.¹²

The case can even be dramatized. Assume that Prophet knows that, due to her speech impediment, it is most unlikely that her predictive act will materialize and hence most unlikely that Lottie will intervene. Prophet therefore holds it very probable that XYZ is *not* the winning combination. Prophet's assertion is a lie despite the fact that she has

¹⁰ Even this is not guaranteed: Prophet may know that her assertion of not- P is false if it comes about, know that it comes about, but nevertheless fail to infer that P . She would then be in a position to know that P , but not in fact know or even believe that P . As this would involve some failure of closure, and perhaps some failure of rationality, I will not base my case on the agent's failure of making the right inferences.

¹¹ Just to be on the safe side, assume that at no point during the act of asserting (and possibly even after), Prophet realizes that she in fact accomplishes (or has accomplished) the assertion of not- P : Prophet may be unable to hear the words she speaks, etc.

¹² One might be inclined to object that, when she knows of her speech impediment, Prophet is less guilty than when she believes that she will assert not- P , because she then knows that there is a chance that mischief will still be prevented. Yet while this observation is correct as far as it goes, it is also irrelevant. Our question is whether the successful attempt constitutes a lie, not whether, or to which extent, Prophet is guilty in attempting to make the prophecy that would be a lie.

a high credence in not- P . Perhaps we may even assume that Prophet has the outright belief that P is false. Prophet then is lying while actually believing what she asserts.

It would be of little use to reject, in response to this dramatized case, the assumption that one can believe not- P while thinking that there is some small chance of P coming true: the appeal of a definition of lying would be greatly diminished if it required some contentious views on the relation between believed probability and outright belief. And of course, there are other examples. King Charles III might believe that his jewels will be safe in the Welsh town Eglwysrw. He knows, of course, that once he asserts this, he has revealed the location of the jewels and thieves will try to steal them. For some political reason, however, he must state, or at least attempt to state, that the jewels will be safe there, but he also thinks that he won't be able to pronounce the name of the town—he has failed often enough in the past—and hence won't be able to complete the assertion. That's why Charles believes that the jewels will remain safe in Eglwysrw. Quite unexpectedly, and unbeknownst to himself, Charles succeeds in pronouncing the difficult name and hence in completing the self-defeating prophecy. Charles does not know that he is lying, because he does not know that he is actually making the assertion. But that does not alter the fact that he *is* lying. Charles ends up lying by asserting something he in fact believes.

If this is correct, (DEF) is too narrow: (CON 2) does not allow for lying in the absence of disbelief, let alone lying in the presence of belief in the asserted proposition. It does not guarantee that known-to-be-self-defeating prophecies are lies. I will now show that (DEF) is also too wide.

2 Truthfulness in the presence of disbelief

Consider an alternative development of PROPHECY:

SELF-FULFILLING PROPHECY. Winnie was able to outsmart Lottie and gain control over the manipulating device. In contrast to Lottie, Winnie is indebted to Prophet and therefore determined to make Prophet's lottery-related assertion *true*. If Prophet asserts P , XYZ will be the winning combination. If Prophet asserts some other outcome, she will be right too. And if Prophet does not make any assertion on this matter, Winnie does not interfere and lets chance decide. Fully aware of the new situation, Prophet attempts and manages to assert P .

Is Prophet a liar? She knows and hence believes that, if she predicts the lottery outcome, this prediction will be true. In particular, she knows that if she asserts P , then P is true and the number combination XYZ will be drawn. If, therefore, she sets out to predict P , she intends to say something true. And she knows that, if she manages to bring about the assertion, she succeeds in saying something true. Making a prophecy which is known to be self-fulfilling cannot constitute a lie. But again, the classical conception does not allow for this unconditional result.

According to (DEF), whether Prophet is truthful or a liar depends on whether she believes that P . And this, in turn, depends on whether she believes that the self-fulfilling prophecy occurs. Thus (DEF) also makes the truthfulness of Prophet's assertion dependent on her believing in the occurrence of the assertion. If Prophet believes that she

asserts P , (DEF) provides the right verdict: she then believes that Winnie interferes and that P is true. Yet suppose again that Prophet knows that, due to her speech impediment, the attempt to assert P will most likely be—or even has most likely been—unsuccessful and the chances of her actually asserting or having asserted P very small. Prophet also knows that, if she does not assert P , Winnie will not intervene and P will most likely be false. We may even assume that, knowing that it is unlikely that she accomplishes or has accomplished the assertion, Prophet believes that P is false: Prophet then disbelieves what she asserts and hence is a liar according to (DEF).

Yet again I submit that this is the wrong result: Although Prophet believes that, given her speech impediment, she won't bring about the self-fulfilling prophecy, that Winnie won't interfere, and hence that P will be false, she also knows that *if* she manages to accomplish the assertion, the assertion will be true and XYZ the winning combination. More generally, it is as impossible to lie with a prophecy known to be self-fulfilling as it is to be truthful with a prophecy known to be self-defeating. If this is correct, there are not only lies in which the asserted content is not disbelieved (or even actually believed), but also perfectly truthful assertions while the agent *disbelieves* the asserted proposition. In order to save (DEF) from this further kind of counterexample, one might reject the assumption that it is possible to disbelieve P while thinking there is some small chance of P coming true. But whereas this maneuver might indeed avoid that (CON 2) is fulfilled and that Prophet technically comes out a liar, this would not be of much help: Prophet *almost* disbelieves P , which would render her assertion of P *almost* a lie, which is wrong enough.¹³

It is easy to construct other cases. Having, to his own surprise, previously revealed the location of the jewels to possible thieves, Charles now believes that his valuables will no longer be safe in Eglwysrw. Of course, he knows that, if only he manages to again publicly assert that the jewels will be safe there, all the police will gather and protect his treasure: the police will do everything to establish public trust in their new king. Yet he thinks that he won't be able to pronounce the name of the town a second time and hence won't be able to make the assertion that the valuables will be safe in Eglwysrw. (For some odd reason, he has no other way of communicating.) As it happens, he succeeds in pronouncing the name again and thus in making the self-fulfilling assertion. But he does not realize that he has succeeded in making this assertion, and so he actually disbelieves what he asserts. He believes that the jewels will not be safe in Eglwysrw. Charles ends up truthfully asserting something he in fact disbelieves.

I don't want to exclude that there are ways for proponents of (DEF) to get around examples such as these, but I will leave matters here. It is obvious that also self-fulfilling prophecies mean trouble for (DEF). And since we need to revise the classical conception of lying in reaction to self-defeating prophecies anyway, it is commendable to search for an alternative that provides a general solution to the problem created by the fact that the agent's belief in the asserted proposition may depend on her belief in the occurrence of the assertoric act.

¹³ Perhaps we should conceive of lies as coming in degrees, depending on the degree of disbelief. (For the basic idea of degrees of lying, see already Simpson, 1992, p. 631.) Given such a refinement, Prophet's assertion that P would come out a lie to a high degree which, given his knowledge that the accomplished prophecy will definitely be true, seems mistaken.

3 The essence of lying

An assertion is true if and only if the asserted proposition is true. Yet as our examples demonstrate, believing an *assertion* to be true or false is not the same as believing the asserted *proposition* to be true or false. In the case of a prophecy known to be self-defeating a person may believe that the assertion is false if it occurs, without believing the asserted proposition to be false. And in a case of a prophecy known to be self-fulfilling a person may believe the assertion to be true if it occurs, while believing the asserted proposition to be false.

Given that there is a difference between believing the asserted *proposition* to be false and believing the *assertion* to be false, there are two ways of construing lying based on the idea that lying is, essentially, to say something believed to be false. The classical account of lying assumes the relevant belief to concern the asserted proposition, thus arriving at (CON 2). A proponent of this view must assume that there are cases of prophecies known to be self-defeating that are *not* lies and cases of prophecies known to be self-fulfilling that *are* lies—and that the difference between lying and not lying can depend on whether the agent believes that the intended assertive act really occurs. I have claimed that this is a most implausible position. In the Lottie situation, Prophet knows that her assertion of not-*P* will be false. Her attempt to assert that not-*P* is an attempt to lie. And if she succeeds in making the assertion, she succeeds in lying. Prophet is lying although she might not know that she is lying. Conversely for the Winnie situation: even though Prophet may not believe that she accomplishes the assertion and hence may believe that not-*P*, she is not lying when she in fact asserts *P*; she knows, and hence believes, that the assertion is true if it comes about. I take these cases to speak strongly against (CON 2).

The alternative holds that the disbelief relevant for lying must be directed, not at the asserted proposition, but at the assertion itself. A liar believes that the assertion is false—if, of course, the assertion comes about. Whether or not she also believes that the assertion does come about, and hence that the asserted proposition is indeed false, is immaterial. We can give a precise formulation of this alternative condition once we observe that believing an assertion to be false is believing the asserted proposition to be false *conditional on the occurrence of the assertion*: if the subject revises her belief state by the information *that she asserts p*, she believes that *p* is false. I therefore suggest that we require, in a definition of lying, instead of an unconditional disbelief, a disbelief conditional on the occurrence of the assertion.¹⁴ We thus arrive at the following modification of (CON 2):

(CON 2*) A believes (not-*p* | A asserts *p*).¹⁵

Condition (CON 2*) is a ‘conservative’ modification of (CON 2) in that it will not make a difference for ‘ordinary’ cases in which *A* disbelieves *p* conditional on her *p*-assertion iff *A* disbelieves *p* unconditionally: *A* believes that it rained last night just in case *A* believes that it rained conditional on her asserting it.

¹⁴ Note that I am requiring conditional disbelief, not disbelief in a conditional.

¹⁵ I ignore here possible complications deriving from the plausible idea that the assertion condition might have to be phrased in *de se* terms.

That the defect of (CON 2) does not show up for ordinary cases explains the common acceptance of something like (DEF). (CON 2*) makes a difference for prophecies which are known or believed to be self-defeating or self-fulfilling, however. It renders Prophet's self-defeating prophecy a lie and avoids the consequence that we can lie with prophecies believed to be self-fulfilling. The essence of lying is not asserting what is disbelieved *simpliciter*, but asserting what is disbelieved conditional on the occurrence of the assertion.

This is surely not the final word about lying. Very likely (CON 1) needs significant modification and refinement; perhaps in (CON 2*) disbelief should be replaced by some other, weaker doxastic condition; and it may very well be that further clauses are necessary to get the definition right. Pointing out that the required doxastic state needs to be directed at the assertion and hence at the propositional content conditional on it being asserted can only be the beginning. I have tried to show that any attempt at a definition of lying must start from there.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest I have no conflicts of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Carson, T. L. (2010). *Lying and deception: Theory and practice*. Oxford University Press.
- Chisholm, R. M., & Feehan, T. D. (1977). The intent to deceive. *Journal of Philosophy*, 74, 143–159.
- Fallis, D. (2009). What is lying? *Journal of Philosophy*, 106, 29–56.
- Krstić, V. (2023). Lying to others, lying to yourself, and literal self-deception. *Inquiry*. <https://doi.org/10.1080/0020174X.2023.2206850>
- Pepp, J. (2018). Truth serum, liar serum, and some problems about saying what you think is false. In E. Michaelson & A. Stokke (Eds.), *Lying: Language, knowledge, ethics, and politics* (pp. 43–64). Oxford University Press.
- Pepp, J. (2019). Assertion, lying, and untruthfully implicating. In S. Goldberg (Ed.), *The Oxford handbook of assertion* (pp. 829–850). Oxford University Press.
- Simpson, D. (1992). Lying, liars and language. *Philosophy and Phenomenological Research*, 52, 623–639.
- Stokke, A. (2018). *Lying and insincerity*. Oxford University Press.
- Viebahn, E. (2021). The lying–misleading distinction: A commitment-based approach. *Journal of Philosophy*, 118, 289–319.
- Williams, B. (2002). *Truth and truthfulness*. Princeton University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.