



Why a Gricean-style defense of the vacuous truth of counterpossibles won't work, but a defense based on heuristics just might

Tomasz Puczyłowski¹ 

Received: 14 May 2023 / Accepted: 19 November 2023 / Published online: 5 January 2024
© The Author(s) 2024

Abstract

Counterpossibles are counterfactuals with an impossible antecedent. According to the orthodox view of counterfactuals, all counterpossibles are vacuously true. This is puzzling because some counterpossible statements seem to be false. The paper analyzes two approaches to explaining why certain counterpossibles, though perhaps true, may appear to be false. The first, which appeals to the Gricean mechanism of conversational implicatures, asserts that some counterpossibles appear to be false because their assertion carries with it a false conversational implicature. However, I argue that, under a closer scrutiny, this approach collapses. I therefore turn to a second approach, proposed by Timothy Williamson. It appeals to a heuristic according to which speakers may regard a counterpossible to be false if they have previously accepted its opposite. Since the applicability of Williamson's solution is limited, I suggest a more general account. Its underlining idea is that a counterpossible is rejected if the speaker cannot find what they regard as a universally true conditional function derivable from the counterpossible by substitutions and syntactic transformations.

Keywords Counterpossibles · Counterfactuals · Conversational implicatures · Warranted assertability maneuver · Heuristics

1 Introduction

Counterpossibles are counterfactuals with an impossible antecedent. As a species of counterfactuals, they can naturally be expected to have truth conditions that follow from the truth conditions of counterfactuals tout court. But it's debatable what the truth conditions of counterfactuals are.

✉ Tomasz Puczyłowski
tomasz.puczylowski@uw.edu.pl

¹ Faculty of Philosophy, University of Warsaw, Krakowskie Przedmieście 3, 00-927 Warsaw, Poland

The question of the truth conditions of counterfactuals has received extensive attention in analytic philosophy. Given that only declarative sentences can have truth values, researchers explored ways of translating counterfactuals into declaratives. Most theorists agree that, whatever their truth conditions might be, they must differ from those of material implication. Alas, the agreement stops here. Some scholars, such as Henry Hiz (1951) and Roderick Chisholm (1946), have argued that counterfactuals are essentially metalinguistic expressions, while others, such as Henry Finch (1958), attempted to establish their truth conditions based on conditional probability. However, since the late 1960s, thanks to the contributions of Robert Stalnaker (1968) and Lewis (1973), one approach has been so widely adopted that Timothy Williamson (2018) calls it the orthodox view of the semantics of counterfactuals. According to it,

a counterfactual of the form ‘ $A > B$ ’ is true at a possible world w iff either (i) there is no possible world accessible from w where A is true, or (ii) there is a possible world w^* accessible from w where A and B are true, and w^* is more similar to w than any world accessible from w where A is true while B is not true.

Since all counterpossibles are statements with an impossible antecedent, condition (i) straightforwardly applies. Therefore, all counterpossibles, regardless of the truth value of their consequents, must be true.

But the rub is that many authors find this consequence to be unpalatable (e.g., Yagisawa, 1988; Nolan, 1997; Brogaard & Salerno, 2013; Berto et al., 2018). Surely, in any world, the following sentences cannot all be true:

- (1) If the sum of 2 and 3 were 6, then pigs would not fly.
- (2) If the sum of 2 and 3 were 6, then the sum of 2 and 3 would not be 6.

So why accept condition (i), leading to the conclusion that all counterpossibles are vacuously true? There are at least two arguments for adopting (i) (see Kocurek, 2021). The first begins with the assumption that the intension of a counterfactual is a function of the intensions of its antecedent and consequent, or to put it formally:

$$|A > B| = f(|A|, |B|),$$

where $|\alpha|$ is the set of all possible worlds at which α is true. Because the truth value of the consequent is irrelevant to the truth value of a counterfactual with a false antecedent, we obtain:

$$|A > B| = f(|A|, |A \cap B|).$$

But if A is impossible—i.e., not true at any possible world—then, for any B ,

$$|A > B| = f(\emptyset, \emptyset \cap |B|) = f(\emptyset, \emptyset).$$

Therefore, the intension, and consequently the truth value, must be the same for all counterpossibles. Because all counterpossibles of the form ‘ $A > A$ ’ seem to be true, it is only natural to assert that all counterpossibles must be true, rather than false.

A second argument in favor of condition (i) proceeds by assuming that if A strictly implies B, then it is also the case that 'A > B' (symbolically: $\Box(A \rightarrow B) \rightarrow (A > B)$). But, if that holds, then, by the laws of modal logics, the following is also true for an arbitrary B

$$\Box\neg A \rightarrow \Box(A \rightarrow B).$$

And from this, by propositional logic, we get:

$$\Box\neg A \rightarrow (A > B),$$

which can be read as 'the impossibility of A implies that if A were the case, then B would be the case'.¹

The upshot of all this is that we are confronted with a dilemma. We must either adopt the orthodox view and accept the consequence that all counterpossibles are vacuously true, or find a different semantics for counterfactuals.² For someone who chooses the first horn of the dilemma, the orthodox view can be seen as facing an explanatory problem: if all counterpossibles are true, then why are some counterpossibles judged to be false? The easiest answer would be that such judgments are only made by speakers ignorant of the semantic and philosophical subtleties involved in interpreting modal expressions. Yet, as we saw, there are knowledgeable and reflective scholars who also regard some counterpossibles to be false. How can an advocate of the orthodox view account for this?

In order to answer this question, I focus on two possible explanations of why certain counterpossibles may appear to be false. The first, proposed by Emery and Hill (2017), appeals to the Gricean mechanism of conversational implicatures. Its basic idea is that, although all counterpossibles are vacuously true, some of them seem to be false because their assertion in certain contexts triggers false conversational implicatures, which the speaker may confuse with the semantic content of the counterpossible in question. Upon closer examination, this pragmatic defense of orthodoxy proves unsatisfactory, however. My argument to this effect relies on a novel test for conversational implicatures that, when applied to at least some propositions that are allegedly conversationally implied by assertions of counterpossibles, reveals that these propositions are not conversational implicatures. Therefore, I proceed to consider a second approach, based on fallible heuristics. Here, I discuss a heuristic proposed by Williamson (2018), according to which speakers may regard a counterpossible to be false if they have previously accepted its opposite. However, since the applicability of Williamson's solution is relatively limited, I suggest a more general explanation. In rough outline, its underlining idea is that a counterpossible is rejected (evaluated as false or having no truth value) if the speaker cannot find what they regard as a universally true conditional function derivable from the counterpossible by substitutions and syntactic transformations. Although I confess that I cannot marshal any empirical data in support of the psychological reality of this heuristic, I hope that the proposed explanation proves sufficiently plausible to merit further investigation.

¹ For a comprehensive exposition of these and other arguments for the orthodox view, see (Kocurek, 2021) and (Williamson, 2007, p. 174; 2018, pp. 360–362).

² Authors who choose the second horn of the dilemma include Nolan (1997), Kment (2006), Priest (2009), Brogaard & Salerno (2013) and Berto et al. (2018).

The paper unfolds in the following manner. In Sect. 2, I discuss Emery and Hill's (2017) proposal to defend the orthodox view by appealing to conversational implicatures. In Sect. 3, I critically examine Maciej Sendłak's (2019, 2021) argument against Emery and Hill's solution. Although I agree with Sendłak's conclusion that, ultimately, a pragmatic defense of the sort proposed by Emery and Hill does not succeed, I take issue with one of his premises. In Sect. 4, I present my own argument against any Gricean-style defense of the orthodox view of counterpossibles. Finally, in Sect. 5, I propose a heuristics-based mechanism responsible for evaluating counterpossibles as true or false. Section 6 offers a brief summary of the conclusions.

2 A Gricean-style defense of the orthodox view

Emery and Hill (2017) explicitly invoke concepts developed by H. P. Grice to explain why there may be some discrepancy between the content of a sentence and what is communicated by its assertion in a given context. They also aim to explain why some counterpossibles seem to ordinary interpreters to be not only non-vacuously true but also informative and interesting (Emery & Hill, 2017, p. 5). As Grice famously noted, by assuming that participants of a conversation are cooperative in the process of communication, it is possible to convey information not included in the semantic content of uttered words or sentences. It is not always easy to distinguish between the literal content of a sentence and information indirectly communicated by asserting the sentence in a given context. Therefore, when some conveyed information is true and some is false, and possibly only one is part of the semantic content of the assertion, an erroneous judgment of the assertion is possible, especially if one takes a piece of pragmatic information for semantic content. For example, sometimes the conjunction 'Mark fell asleep and yelled at Julian' might be mistakenly deemed false, if it is true that Mark did both actions, albeit not in the implied order. The error occurs because one may incorrectly assume that pragmatically communicated information—namely, the order of events—is part of the semantic content of the conjunction.

It seems that Emery and Hill address a common type of argument levelled against the orthodox interpretation of counterpossibles:

P1. According to the orthodox view of counterfactuals, any sentence of the form 'A > B' where A states an impossible state of affairs is true.

P2. However, on some occasions some such sentences seem to be false.

P3. Such sentences seem to be false because their assertion is not warranted.

P4. If such a sentence is not warrantably assertible in a context, then it must be false in that context.

P5. There is no other plausible explanation why 'A > B' with an impossible antecedent might seem to be false.

Therefore,

C1. Some counterpossibles are false.

And, by extension,

C2. The orthodox view of counterfactuals is false.

Emery and Hill challenge P5 by presenting reasons why some true counterpossibles may appear to be false. In particular, they argue that true counterpossibles can generate false conversational implicatures, leading to confusion about their truth value. This observation, combined with the belief that one can conflate the semantic content of a sentence with its conversational implicatures, sheds light on how such confusion can occur.

How can one explain that assertions of some counterpossibles may give rise to conversational implicatures? Emery and Hill (2017, p. 6) suggest that conversational implicatures stemming from these assertions would likely result from the flouting of the Maxim of Quantity:

As Grice observed in his original discussion of implicature, conversations seem to be governed by a ‘Maxim of Quantity’ which advises speakers to ‘be as informative as required’ (Grice, 1989). Insofar as participants in a conversation are aware of this maxim, and believe each other to intend to conform to it, they can exploit it to communicate propositions other than the ones that they explicitly assert. For if a speaker asserts a proposition that is trivially true, and therefore uninformative, the audience will assume that the speaker intends to communicate a more substantial proposition that is related to the asserted proposition in subject matter and will look around for salient propositions that have these properties. Now we can suppose that something like this is what happens when a speaker asserts a counterfactual with an impossible antecedent. Recognizing that the proposition is a triviality, and believing that the speaker would not wish to violate the Maxim of Quantity, a philosophically sophisticated audience will look for a more interesting proposition that has more or less the same subject matter and that the speaker is likely to have in mind.

If the orthodox account is accurate and a counterpossible conveys only its literal meaning, all counterpossibles should appear informatively and cognitively trivial and uninteresting.³ Therefore, by uttering them, one can only pragmatically convey some nontrivial information. If a speaker chooses to be cooperative in a conversation and aims to convey nontrivial information, they should refrain from asserting a vacuously true counterfactual when their sole intention is to communicate the literal meaning of their utterance. Otherwise, they would violate the Maxim of Quantity. Nevertheless, if they opt for asserting a counterpossible, their cooperation hinges on the intention to communicate nontrivial information, with the expectation that a discerning hearer can decode the implied meaning.

But it might be the case that other maxims besides the Maxim of Quantity come into play when calculating the implicatures potentially conveyed by counterpossibles. For, as Williamson (2018, p. 363) pointed out, instead of asserting a counterpossible, one could express equally relevant (although also equally trivial) information more concisely by a sentence stating the impossibility of the antecedent of the respective counterpossible. The observation that any assertion of a counterpossible merely expresses trivially true information in an unnecessarily complicated form would activate the process of working out what is actually communicated by the speaker

³ This is what Brogaard and Salerno (2013, p. 642) explicitly claim.

(provided she was cooperative, etc.). This unnecessarily complicated form encoding trivial information must be observed in the process of working out what is meant by uttering a counterpossible. Therefore, in contrast to what Emery and Hill suggest, not only the Maxim of Quantity but also the Maxim of Manner would be necessary in explaining what may be conversationally implied by asserting a counterpossible, provided the defense based on conversational implicatures is sound.

In addition, it is worth pointing out that two logically equivalent sentences can differ in meaning because of having different intensional structures (Carnap, 1947, §13–15). Therefore, although all counterpossibles may be logically equivalent on account of being necessarily true, they could still differ in meaning. Hence, the question of the type of conversational implicatures that are presupposed in this kind of defense remains open: they may involve implicatures of Quantity alone or implicatures of Quantity and Manner.

3 Sendłak's argument against Emery and Hill's defense of the orthodox view

According to Sendłak (2019, 2021), Emery and Hill's account fails because the same kind of mechanism could be invoked to argue for the vacuity of all counterfactual propositions. In other words, one could just as well argue that although every counterfactual proposition is vacuously true, some counterfactual statements seem to be false because their assertion in certain contexts gives rise to false conversational implicatures. That, however, would go against the chief motivation for interpreting counterfactuals in terms of possible worlds, and thereby undermine, rather than support, the orthodox view.

I believe that Sendłak's criticism misses its mark. Contrary to what he contends, one cannot adopt the same kind of pragmatic defense as that proposed by Emery and Hill to argue that every counterfactual is vacuously true. The reason is that Emery and Hill's solution is an instance of a more general argumentative strategy called the *warranted assertability maneuver*, or *WAM* (DeRose, 1998, 2002), which is successful just in case it satisfies certain conditions.⁴ The trouble is that Sendłak's putative argument concerning the truth value of all counterfactuals does not meet them.

The three conditions on any successful WAM are as follows (DeRose, 1998, §10; 2002, pp. 172–176):

1. it needs to explain away only the appearance of falsity—in particular, in the same context not only a given sentence must appear to be false, but also its negation must seem to be false;
2. it can provide the explanation by appealing only to the generation of a false implicature (or false implicatures);
3. it appeals to general rules of conversation in explaining why the apparently false assertion is unwarranted.

⁴ This kind of rebuttal is not rare in philosophical debates concerning semantics. For example, in a debate concerning fiction, some argue that fictional sentences appear to have a truth value only because their assertions conversationally implicate some propositions (Adams et al., 1997; Adams & Fuller, 2007; cf. Puczyłowski 2021).

One cannot overemphasize the importance of condition 1. As Brown rightly observes (2006, p. 411), ‘DeRose argues that a good candidate for a WAM should start from a set of intuitions which, plausibly, cannot all be correct so that one or more needs explaining away.’ In the case of counterpossibles, condition 1 seems to be satisfied, at least when a counterpossible appears to be false. For example, both ‘If Napoleon had been Socrates, Marco Polo would have been a sailor’ and ‘It might be the case that Napoleon had been Socrates but Marco Polo had not been a sailor’ appear to be false.

But consider the case of counterfactuals with merely possible antecedents. According to a WAM-style explanation of the apparent falsity of ‘If Columbus had reached the place he was planning to reach in 1492, he would not have arrived in India,’ an assertion of that sentence conveys a false proposition—namely, that Columbus was not planning to reach India. However, before offering such an explanation, one must ensure, as the criteria for a successful WAM require, that the negation of that counterfactual also appears to be false in the same context of evaluation. This is not the case. Sentences such as ‘It might be the case that Columbus had reached the place he was planning to reach in 1492 and he would have arrived in India’ and ‘If Columbus had reached the place he was planning to reach in 1492, he would have arrived in India’ appear to be true.

To sum up my reservation toward Sendlak’s argument: to propose a similar and successful pragmatic account of counterfactuals with merely possible antecedents, one would have to demonstrate that, whenever a counterfactual ‘ $A > B$ ’ seems to be false, there exists another one—that is, its negation—‘It might be the case that A and it is not the case that B ’ or ‘If it were the case that A , it might be that not B ’ or ‘ $A > \text{not } B$ ’ (or whatever the negation of a counterfactual is, see Espino et al. 2022) that appears to be equally false in the given context of evaluation. My contention is that this is not always the case, so WAMs aiming to demonstrate that some apparently false counterfactuals with merely possible antecedents are vacuously true and their apparent falsehood is produced by some pragmatics mechanisms cannot get off the ground.

4 An argument against any defense of the orthodox view that invokes conversational implicatures

Although I find Sendlak’s argument wanting, I agree with its conclusion about the inadequacy of the pragmatic approach to counterpossibles. Naturally, this does not necessarily mean that advocates of the orthodox view are wrong about the semantics of counterpossibles. They may be correct in claiming that all counterpossibles are true, but the mechanism responsible for assessing some of them as false may be different from that suggested by Emery and Hill. In this section, I develop my own argument for rejecting any defense of the orthodox view of counterpossibles that is based on conversational implicatures. However, before I can present the argument itself, I have to do a bit of stage-setting—this is the job of Sect. 4.1.

4.1 A novel test for conversational implicatures

If false conversational implicatures are responsible for some judgments that a counterpossible is false, they should, naturally, have all the distinguishing features of conversational implicatures in general. Among the properties that distinguish conversational implicatures from other types of inferences, some are canonical—i.e., introduced by Grice himself. They include: cancellability, non-detachability, calculability, and non-conventionality. These features can be provisionally defined as follows⁵:

If an utterance of sentence S in context C conveys proposition p , then p is *non-detachable* from the content of the utterance of S in C if and only if p is conveyed in C by the utterance of any sentence synonymous with S .

If an assertion of sentence S in context C conveys proposition p , then p is a *non-conventional* part of the overall content of the act of uttering S if and only if there is a context C^* (different from C) in which the assertion of S does not convey p .

Proposition p is a *calculable* conversational implicature of an utterance made by person A if and only if, in justifying the fact that p is implied by A 's utterance in a given context, it is necessary to refer to the Cooperative Principle and conversational maxims, the content of A 's utterance, its context, and background knowledge.

If an utterance of sentence S in context C conveys proposition p then p is a *cancellable* part of this act iff there is a sentence S^* and a context C^* such that (i) S is a proper part of S^* , (ii) S^* entails S , and (iii) p is not conveyed by S^* in C^* , but assertion of S^* is admissible in C^* .

Now note that, for each potential implicature triggered by the use of a sentence, there must be a context where that implicature does not arise. This is what distinguishes conversational implicatures from logical consequences, conventional implicatures, and presuppositions. Non-conventionality of conversational implicatures suggests a further test for implicatures and a corresponding distinguishing feature, which I call *non-preservation in the context of a pressing yes-or-no question*:

If an utterance of sentence S in context C conveys proposition p , then p is *not preserved in a direct answer to a pressing yes-or-no question* (part of communication content of S) if and only if an assertion of S (in a different context C^*) as a direct affirmative answer to the pressing yes-or-no question such as 'Is it the case that S ? Is S true? Yes or no?' or 'I am not interested if p or not- p or something else—only in this: is it true that S ? Yes or no?' does not convey p in C^* .⁶

The test is based on a hypothesis, which—I admit—may require further investigation, that, when one says (or makes as if to say, by answering 'Yes') that P in response

⁵ In addition, conversational implicatures are expected to be reinforceable (Sadock, 1978, p. 294), and not preserved under negation and question (Kroeger, 2018, p. 152). These are non-canonical features of conversational implicatures.

⁶ For the sake of concision, I have opted for a simplified version of the question in some cases.

to a pressing yes-or-no question ‘All that I am interested in is the following: Is it the case that P? Yes or no?’ then the assertion that P does not convey any conversational implicature. Let us examine some typical examples of conversational implicature and test whether they are preserved in direct answers to pressing yes-or-no questions. The examples are structured as follows: first, I show a context where an assertion of a sentence generates a conversational implicature. Then, I take a similar but different context (the context of a pressing yes-or-no question) where an assertion of the same sentence doesn’t give rise to the implicature.

It is commonly acknowledged by philosophers of language that the following example illustrates an application of the Maxim of Relevance: Suppose someone were asked to write a letter of recommendation for Mr. X, and all they wrote in response was:

- (3) *Mr. X’s command of English is excellent, and his attendance at tutorials has been regular.*

By such an utterance, they would conversationally imply that, in their opinion, Mr. X was not highly competent in the relevant field. However, the same conclusion seems not to follow in the circumstances where academic qualities of Mr. X are not at issues, and what is relevant is determined by the following question:

- (4) *What do you think—is Mr. X’s command of English excellent? Yes or no? Has his attendance at tutorials been regular? Yes or no?*

In such circumstances, the assertive answer (including a simple ‘yes’) conveys the expected quantity of relevant information, so no need arises for conversational maxim exploitation. Therefore, in such circumstances, (3) does not convey what was conversationally implicated in some other scenario. Note, however, that all logical consequences and presuppositions are preserved in the context of a pressing yes-or-no question. For example, a positive answer in both contexts conveys that Mr. X’s command of English is excellent and that Mr. X speaks English.

For the next illustration, suppose you are asked.

- (5) *Is Tom seeing anyone?*

to which you respond assertively, but ostensibly in an irrelevant manner as if flouting the Maxim of Relevance:

- (6) *He has been paying a lot of visits to New York lately.*

Clearly, by saying so, you imply that Tom is seeing someone in New York (provided that you are being cooperative) and this is why he has been travelling there. However, if your assertion were a direct answer to the questions:

- (7) *Has Tom been paying a lot of visits to New York lately? Yes or no?*
 (8) *I am not interested in whether or not Tom is seeing anyone, nor do I want to know why he has been going to New York. I only want to know this: is it true that Tom has been paying a lot of visits to New York lately: yes or no?*

you would not blatantly violate any conversational maxim; therefore, your answer would not give rise to conversational implicature.

Even if an implicature is triggered in an uncontentional way, that is, by the assumption that all maxims are observed, it is not preserved in the direct answer to a pressing yes-or-no question. In some context, asserting.

(9) *Alfred finished his manuscript and went on his holidays.*

conveys that—by the assumption that a speaker observes the sub-maxim ‘be orderly’—events are recounted in the order in which they occurred (Levinson, 1983, p. 108). But responding with (9), which counts as a direct affirmative answer to the following pressing yes-or-no question, doesn’t convey this conclusion:

(10) *I am not interested in what happened first—I only want to know this: is it true that Alfred finished his manuscript and went on his holidays: yes or no?*

For a final illustration, suppose that we are in a context where, due to John’s negligence, some documents are missing. We want to establish the extent of the loss. Then if someone said.

(11) *John has lost many of the documents.*

the rest of us would be right to adopt the belief that he has not lost all of the documents. Let us now change the context to:

(12) *Has John lost many of the documents? Yes or no?*

(13) *I am not interested in whether or not John has lost all the documents—I only want to know this: is true that he has lost many of the documents: yes or no?*

Under such circumstances, the conclusion—that the speaker implicates that John has not lost all of the documents—would not be rational if (11) were used as a reply.

If such elaborate answers strike you as extravagant and clumsy, and you worry that their length suggests that whoever is giving them is conveying something more, just assume that the answers given were a simple ‘Yes’. It should be clear that the answer ‘Yes’ counts in the circumstances as saying, in an elliptical manner, that, respectively, Mr. X’s command of English is excellent, and his attendance at tutorials has been regular, John has been paying a lot of visits to New York, Alfred finished his manuscript and went on his holydays, and John has lost most of the documents. Nevertheless, nothing seems to be implied by such answers, since they convey exactly the right amount of relevant information and do not flout any of the conversational maxims in contexts where they serve as direct answers to pressing yes-or-no questions.

4.2 The argument

If I am correct about the proposed feature of conversational implicature, it is unlikely that counterpossibles could appear to be false solely because their assertions generate false conversational implicatures. For instance, suppose someone says.

(14) *All I want to know is whether the following is true: if it were the case that $2 + 2 = 5$, it would be the case that $2 + 3 = 127$. Is it: yes or no?*

I wager that, if someone rejects.

(15) *If $2 + 2$ were equal to 5, $2 + 3$ would be 127.*

in some contexts of evaluation, their assessment would not change in the context of a pressing yes-or-no question—they would still respond in the negative to (14). A

Gricean-style defense of the orthodox view amounts to the idea that, by saying ‘yes,’ one would endorse a false proposition implicated conversationally by an assertion of (15). But if conversational implicatures are not preserved in direct answers to pressing yes-or-no questions, one would have no reason not to reply in the affirmative. Therefore, what is responsible for the impression that some counterpossibles are false are not false conversational implicatures but something else.

If this argument is not fully convincing, let us look at propositions allegedly conversationally implied by assertions of counterpossible and see if they have the canonical properties of conversational implicatures. It will turn out that they lack at least one such feature. Regarding calculability, however, one can argue that the propositions conversationally implied by assertions of counterpossibles can be calculated (as shown in Sect. 2). If I am right about them not being preserved in direct answers to pressing yes-or-no questions, they are non-conventional. However, are they cancellable? Is the following sentence that cancels the alleged implicature admissible?

- (16) *If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared, but I don't want to imply that the first would affect the latter, I am just trying to say something undeniably true.*

I must admit that my intuition fails me here. Nevertheless, to my ear, such an assertion sounds odd and uninformative.

What about non-detachability? This is a delicate matter, but if one can agree that sentences ‘ $a = b$ ’ and ‘ $b = a$ ’ are synonymous (where a and b stand for proper names), then it seems that whatever assertions of a counterpossible convey, it is detachable, and hence not a conversational implicature. To see this, consider two counterpossibles that seem to differ in terms of their truth value:

- (17) *If Bonaparte had been Julius Caesar, he would have spoken Latin.*
 (18) *If Julius Caesar had been Bonaparte, he would have spoken Latin.*

But being synonymous, they should conversationally generate the same implicatures. So, if false implicatures are to blame for the impression that (17) is false, then (18) should be implicating the same false implicature, and hence the first counterpossible should appear to be false as well. But it does not.

5 A heuristics-based defense of the orthodox view

In this section, I discuss possible reasons for why people tend to perceive such sentences as false, without necessarily delving into their semantic status. My approach is intended to be sufficiently general to avoid commitment to any claims about whether or not such sentences express propositions. Even if they do not, the explanation I offer will nevertheless account for why they are perceived as meaningful and truth-evaluable.

My proposal builds on the idea that people use fallible heuristics to evaluate or explain the apparent truth values of some sentences, including counterpossibles. The specific heuristics may vary depending on the type of sentence being assessed. For instance, one heuristic may explain errors involving universal quantification (Brooks & Sekerina, 2005), while another may account for why conditionals are sometimes

interpreted as if they were biconditionals (Wyer & Srull, 1989, p. 281). The idea of employing heuristics to explain the apparent falsehood of counterpossibles originates from the work of Timothy Williamson (2018). He argued that all counterpossibles are true, but some are erroneously assessed as false because a seemingly opposite counterpossible had been accepted previously.⁷ The fallible heuristic he proposed can be summarized as follows:

If you accept counterfactual ‘ $A > B$ ’ and A states some impossible state of affairs, then reject ‘ $A > \text{not-}B$ ’

or

If you accept one of ‘ $A > B$ ’ and ‘ $A > \text{not-}B$ ’, reject the other.

I believe that this heuristic provides a plausible explanation for why certain counterpossibles can appear false. This outcome is not surprising, especially if we have previously accepted a counterpossible with a contradictory consequent. However, it has a limitation: for a counterpossible to be considered false, there must exist a corresponding counterpossible that has been accepted as true.⁸ This means that the heuristic cannot explain why some counterpossibles ‘ $A > B$ ’ and ‘ $A > \text{not-}B$ ’ seem equally false. As an example, consider:

(19) *If $2 + 2 = 5$, Plato would have been born in Paris.*

(20) *If $2 + 2 = 5$, Plato would not have been born in Paris.*

It is likely that we wouldn’t deem either (19) or (20) as acceptable.⁹

My proposed heuristic for establishing the truth value of counterpossibles is intended to avoid the limitation of Williamson’s solution. It consists of two main parts. First, it suggests that counterpossibles can only be assessed as plausibly acceptable if they can be viewed as substitutions of indicative conditional propositional functions that seem to be universally satisfied. To make this basic premise clear, let us consider counterpossibles of the following two kinds (V stands for a verb):

(I) *If A had been B , A would (not) have V -ed...*

(e.g., *If Bonaparte had been Caesar, he would have lived in Rome*)

(II) *If A had V ed, (A) would have V^* -ed...*

(e.g., *If Bonaparte had squared the circle, he would have been a famous mathematician*).

The heuristic, at least to a first approximation, stipulates that these counterpossibles can be deemed acceptable or assertable only if they can be obtained by simple grammatical

⁷ McLoone et al. (2023) suggest that Williamson is not always consistent, since on one occasion (Williamson, 2017, p. 161) he “acknowledges that we can reason non-vacuously about impossible scenarios”.

⁸ Berto et al. (2018) raise a number of objections to Williamson’s heuristic; for responses, see (Williamson, 2020, pp. 257–261).

⁹ Interestingly, Williamson (2018, p. 263) also offers a Gricean explanation of why using a counterpossible seems wrong when the speaker is fully aware of the impossibility of the state of affairs stated by its antecedent. However, he invokes Gricean mechanisms to explain why some counterpossibles may seem informative; he does not explicitly and unambiguously attribute their sense of falsity to implicatures.

transformations from seemingly universally satisfied indicative conditional propositional functions (sometimes augmented by some contextually relevant assumptions and addenda; hereinafter, given in parenthesis):

(i) *If x is B , x V ...*

(If x is B (and B lives in C), then x lives in C)

(ii) *If x V s, x V^* s...*

(If x solves P (and P is a famous problem that belongs to the filed F), x becomes a famous representative of F)).

Second, if a given counterpossible cannot be derived from a universally true conditional function by substitutions and syntactic transformations, the heuristic recommends rejecting it. For the sake of brevity, my outline of the proposed heuristic is restricted to only a few kinds of counterfactuals with impossible antecedents, but, hopefully, these steps can be applied to other types of counterpossibles by following a similar pattern and can be summarized in the following list:

1. Discover a seemingly universally true indicative conditional propositional function, with all underlying assumptions embedded in the antecedent of the counterpossible, and all omitted arguments and additions explicitly expressed (that are implicitly assumed in assessing the counterpossible).¹⁰
2. Uniformly substitute all variables, whether they are individual, propositional, or predicate variables.
3. Convert the indicative conditional structure to the second or third conditional, and if necessary, remove any additional assumptions in the antecedent and arguments by utilizing ellipsis and elision, to obtain the counterpossible.
4. Accept the counterpossible if, and only if, you can perform steps 1–3.

Some proponents of the pragmatic defense of the orthodox view may consider such an indicative propositional function (with all relevant assumptions embedded in its antecedent and augmented by omitted arguments) as the proposition pragmatically implied by an assertion of the counterpossible. For example, Emery and Hill (2017, p. 3) think that an assertion of.

(21) *If Obama had had different parents, he would have had different DNA.*

implies that Obama's DNA depends causally on properties of his parents. I would rather say that it is viewed as plausibly asserted or acceptable because the following seems to be universally true (which could be substantiated by referring to the causal connection Emery and Hill have in mind):

(22) *If x has different parents than y , x has different DNA than y .*

However, even if (22) may play a role in taking the counterpossible as true, this proposition is not conversationally implied by assertions of (21). If I were to speculate on

¹⁰ Here and in subsequent instances, the adjective 'seemingly' used for 'true sentence/function/proposition' implies the agent's belief that the sentence/function/proposition is true. In other words, the sought-after indicative propositional form need not be objectively true; it only needs to appear true to the agent. Alternatively, the condition can open with '1. Discover an indicative conditional propositional function that you take to be true.'

the matter, I would say that the sense of the informativeness of counterpossibles stems from the fact that, in the process of evaluating such sentences in terms of their truth value, the interpreter looks for such—usually non-trivial—universally true, possibly non-vacuously true, indicative conditional propositional functions and often finds them. Also, disagreement among interpreters about the truth value or acceptability of a given counterpossible may be the result of assigning to a given counterpossible different indicative conditional propositional functions as step 1 of the heuristic postulates—some interpreters find true ones, others don't or find only those that are false.

Another possible source of disagreement may result from adopting different views on what constitutes an adequate semantics for indicative conditionals. For instance, let's assume that one agent adopts the orthodox analysis, according to which such a sentence is false only if its antecedent is true while its consequent is false. Others may disagree and oppose the idea of treating indicative conditionals as material implication. However, let's assume that both parties view a counterpossible as a substitution of the same indicative conditional propositional function: 'If x is a man and has 10 heads, x is French.' In this case, they can differ in their assessments of the function. The first may insist that it is universally true, even if it is vacuously true, since no man can satisfy its antecedent. The other may disagree and argue that 'if, then' is not necessarily a truth-functional connective and for that reason the function is not universally satisfied and true. Thus, the difference in assessing the truth-value of an indicative conditional may lead to different assessments of the relevant counterpossible. Consequently, even if the agents agree that the only indicative conditional closely resembling the form and content of a counterpossible has an impossible antecedent, disagreement may persist if agents differ in what semantics they adopt for indicative conditionals. For example, assume both agents find an indicative conditional.

If x is P and x is not P , then q .

as the one that most closely resembles, with respect to form and content, the counterpossible in question. One can assess it as true (and hence accept the counterpossible) while the other can reject it and, consequently, reject the counterpossible, if they differ in what it takes for an indicative conditional to be true.

Let us consider a few examples of counterpossibles to see how the heuristic is supposed to work. The counterpossible.

(23) *If $2 + 3 = 6$, then $2 + 4 = 7$.*

can seem to be acceptable to an agent who takes the following to be true:

(24) *If $x + y = z$, then $x + (y + 1) = (z + 1)$.*

or because the agent believes that:

(25) *If $x + 3 = z$, $x + 4 = (z + 1)$.*

or that.

(26) *If $2 + y = z$, $2 + (y + 1) = (z + 1)$.*

By contrast, the counterpossible.

(27) *If $2 + 3 = 6$, then $2 + 4 = 17$.*

seems to be unacceptable because it is hard to find a general conditional it instantiates.

The counterpossible.

(28) *If Cesare had been Hitler, he would have exterminated the Gauls*

may seem to be true or acceptable to someone who accepts (roughly speaking) one of the following:

(29) *If x is y (and y exterminates y 's foes of nationality N), x exterminates those of nationality N .*

(30) *If x is (acting like) y (and y exterminates y 's foes of nationality N), x exterminates those of nationality N .*

Probably, agents who reject (29) or (30) as universally true, would reject the counterpossible as well.

Sometimes it may be difficult to find a proper general statement that an agent regards to be true. Suppose that I am asked 'How much is $5 + 7$?' and answer '11'. Under the circumstances, the assertion of

(31) *If $5 + 7$ were 13, I would have got that sum right.*

seems unacceptable, whereas the assertion of.

(32) *If $5 + 7$ were 13, I would have got that sum wrong.*

appears to be acceptable. I suppose one can reconstruct the rationale for such judgements by appealing to the following complex line of reasoning:

(33) *Propositions p and q are contrary. I accept that p . So if it is the case that q , then I am wrong (accepting that p). So it is not the case that if q , then I am right accepting that p .*

Or, to put it even more generally, a possible reason why (32) can be judged acceptable is that it can be seen as a special case of the apparently true conditional (clauses in parenthesis explicitly state what is implicitly assumed during an evaluation of a given counterpossible):

(34) *(If I claim that p which is contrary to q , then) if q , I am wrong (claiming that p)*

The above conditional can be substituted accordingly and transformed into the following third conditional:

(35) *(If I had claimed that $5 + 7 = 11$ which contradicts $5 + 7 = 13$, then) if it were the case that $5 + 7 = 13$, then I would be wrong (claiming $5 + 7 = 11$).*

Therefore, according to the proposed heuristics, if one takes the conditional as being universally true, one accepts the counterpossible as plausibly assertable.

How can one discover an apparently universally true indicative conditional function that applies to a counterpossible? How can one determine that in order to assess (28) as true, (30) should be identified? I do not believe that a universally applicable rule can be provided. There may be other conditional propositional forms that an agent can find universally true and accept the counterpossible on that basis. It must be emphasized is

that there isn't a single correct indicative conditional that must be found. If one finds one that fits the criteria, it provides a basis for accepting the relevant counterpossible; if not, and if the individual is not an advocate of the orthodoxy, they may reject the counterpossible. However, in some cases, more than one indicative conditional may be applicable to a given counterpossible—if they differ in truth value, it can be particularly challenging for an agent to evaluate the counterpossible. Nevertheless, in these circumstances, if I may speculate on this psychological matter, one tries to assess which of the two better captures the relevant counterpossible in terms of its structure and subject matter. There may not be a final, decisive or correct answer here.

What motivates the heuristic? Assessing the truth value of a counterpossible does not always seem straightforward, especially when the impossibility of its antecedent is acknowledged. Since the antecedent and often the consequent do not pertain to facts but rather to states of affairs that have not occurred or may even be impossible, evaluating the truth value of the sentence is not as straightforward as with simple indicative sentences. However, even the assessment of indicative conditional sentences can be perplexing, as any logic teacher can probably attest. Indicative conditionals appear to be less problematic in this respect than counterpossibles. It seems that typical users may lack the cognitive or theoretical tools to determine precisely what is being communicated and how to assess the truth value of such complex and compound sentences. Sometimes, however, they face the issue of whether a given counterpossible should be accepted and they are expected to decide whether they take a counterpossible as true or false. In such cases, when they are pressed to judge a counterpossible in these terms, users may, I hypothesize, resort to a crude method for arriving at an answer.

6 Conclusion

I have argued that a Gricean defense of the orthodox view fails. If assertion of some counterpossibles gave rise to false conversational implicatures, which the speakers might mistake for semantic content, then the alleged implicatures would be suspended in the context of a pressing yes-or-no question. Sometimes, however, assertions of such sentences appear to be false even when they are direct answers to a pressing yes-or-no question. Therefore, a defense of the orthodox view that appeals to false conversational implicatures cannot be effective.

As an alternative defense, I have proposed a fallible but reliable heuristic that seems to be operative when people are confronted with the problem of assessing the truth values of counterpossibles. Failure to find an appropriate universally true conditional propositional function may be responsible for judging some counterpossibles as false.

I believe that empirical studies designed to assess the proposed test for conversational implicature could help investigate its reliability. If shown to be reliable, the test could serve as a valuable tool for evaluating philosophical arguments that employ the mechanism of conversational implicature to support specific semantic theories. Additionally, conducting an empirical study to determine whether users employ the outlined heuristics would be valuable. If people indeed accept or reject counterpossibles in the way described in this paper, this would provide indirect support for the orthodox view.

Acknowledgements I would like to thank Mariusz Popieluch and Witold M. Hensel for comments and criticisms on earlier drafts of this paper. I am grateful to three anonymous reviewers who provided me with excellent comments on the penultimate version of the paper.

Declarations

Conflict of interest The author has no competing interests to declare that are relevant to the content of this article.

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

- Adams, F., & Fuller, G. (2007). Empty names and pragmatic implicatures. *Canadian Journal of Philosophy*, 37, 449–462. <https://doi.org/10.1353/cjp.2007.0024>
- Adams, F., Fuller, G., & Stecker, R. (1997). The semantics of fictional names. *Pacific Philosophical Quarterly*, 78, 128–148. <https://doi.org/10.1111/1468-0114.00032>
- Berto, F., French, R., Priest, G., & Ripley, D. (2018). Williamson on counterpossibles. *The Journal of Philosophical Logic*, 47, 693–713. <https://doi.org/10.1007/s10992-017-9446-x>
- Brogaard, B., & Salerno, J. (2013). Remarks on counterpossibles. *Synthese*, 190(4), 639–660. <https://doi.org/10.1007/s11229-012-0196-6>
- Brooks, P. J., & Sekerina, I. (2005). Shortcuts to quantifier interpretation in children and adults. *Language Acquisition*, 13(3), 177–206. https://doi.org/10.1207/s15327817la1303_2
- Brown, J. (2006). Contextualism and warranted assertibility manoeuvres. *Philosophical Studies*, 130(3), 407–435. <https://doi.org/10.1007/s11098-004-5747-3>
- Carnap, R. (1947). *Meaning and necessity. A study in semantics and modal logic*. The University of Chicago Press.
- Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind*, 55(220), 289–307.
- DeRose, K. (1998). Contextualism: An explanation and defence. In J. Greco & E. Sosa (Eds.), *Blackwell guide to epistemology* (pp. 196–203). Blackwell.
- DeRose, K. (2002). Assertion, knowledge and context. *Philosophical Review*, 111(2), 167–203. <https://doi.org/10.1215/00318108-111-2-167>
- Emery, N., & Hill, C. H. (2017). Impossible worlds and metaphysical explanation: Comments on Kment's modality and explanatory reasoning. *Analysis*, 77(1), 134–148. <https://doi.org/10.1093/analys/anw068>
- Espino, O., Orenes, I., & Moreno-Ríos, S. (2022). Inferences from the negation of counterfactual and semifactual conditionals. *Memory and Cognition*, 50, 1090–1102. <https://doi.org/10.3758/s13421-021-01252-4>
- Finch, H. A. (1958). An explication of counterfactuals by probability theory. *Philosophy and Phenomenological Research*, 18(3), 368–378.
- Grice, P. (1989). *Studies in the ways of words*. Harvard University Press.
- Hiž, H. (1951). On the inferential sense of contrary-to-fact conditionals. *The Journal of Philosophy*, 48, 586–587.
- Kment, B. (2006). Counterfactuals and explanation. *Mind*, 115(458), 261–310. <https://doi.org/10.1093/mind/fzl261>
- Kocurek, A. W. (2021). Counterpossibles. *Philosophy Compass*, 16(11), e12787. <https://doi.org/10.1111/phc3.12787>

- Kroeger, P. R. (2018). Analyzing meaning: An introduction to semantics and pragmatics (textbooks in language sciences 5). Language Science Press. <https://doi.org/10.5281/zenodo.2538330>
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511813313>
- Lewis, D. (1973). Counterfactuals. *Harvard University Press*. <https://doi.org/10.2307/2273738>
- McLoone, B., Grützner, C., & Stuart, M. T. (2023). Counterpossibles in science: An experimental study. *Synthese*, 201, 27. <https://doi.org/10.1007/s11229-022-04014-0>
- Nolan, D. (1997). Impossible worlds: Modest approach. *Notre Dame Journal of Formal Logic*, 38(4), 535–572. <https://doi.org/10.1305/ndjfl/1039540769>
- Priest, G. (2009). Conditionals: A debate with Jackson. In I. Ravenscroft (Ed.), *Minds, worlds and conditionals: Themes from the philosophy of Frank Jackson* (pp. 311–335). Oxford University Press.
- Puczyłowski, T. (2019). Fictional sentences and the pragmatic defence of direct reference theories. *Studia Semiotyczne*, 33(2), 259–276. <https://doi.org/10.26333/sts.xxxiii2.09>
- Sadock, J. (1978). On testing for conversational implicature. In P. Cole (Ed.), *Pragmatics (syntax and semantics 9)* (pp. 281–297). Academic Press.
- Sendlak, M. (2019). On the pragmatic approach to counterpossibles. *Philosophia*, 47(2), 523–532. <https://doi.org/10.1007/s11406-018-9979-4>
- Sendlak, M. (2021). Counterpossibles, story prefix and trivialism. *Synthese*, 199, 7283–7301. <https://doi.org/10.1007/s11229-021-03114-7>
- Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory (American Philosophical Quarterly Monographs 2)* (pp. 98–112). Blackwell.
- Williamson, T. (2007). *The philosophy of philosophy*. Wiley.
- Williamson, T. (2017). Model-building in philosophy. In R. Blackford & D. Broderick (Eds.), *Philosophy's future: The problem of philosophical progress* (pp. 159–172). Wiley.
- Williamson, T. (2018). Counterpossibles. *Topoi*, 37(3), 357–368. <https://doi.org/10.1007/s11245-016-9424-x>
- Williamson, T. (2020). *Suppose and tell. The semantics and heuristics of conditionals*. Oxford University Press.
- Wyer, R. S., Jr., & Srull, T. K. (1989). Memory and cognition in its social context. *Psychology Press*. <https://doi.org/10.4324/9781315807591>
- Yagisawa, T. (1988). Beyond possible worlds. *Philosophical Studies*, 53, 175–204.

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