

Counterfactuals, irrelevant semifactuals and the \$1.000.000 bet

Lars Bo Gundersen & Jesper Kallestrup

To cite this article: Lars Bo Gundersen & Jesper Kallestrup (2023): Counterfactuals, irrelevant semifactuals and the \$1.000.000 bet, *Inquiry*, DOI: [10.1080/0020174X.2023.2249510](https://doi.org/10.1080/0020174X.2023.2249510)

To link to this article: <https://doi.org/10.1080/0020174X.2023.2249510>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 24 Aug 2023.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Counterfactuals, irrelevant semifactuals and the \$1.000.000 bet

Lars Bo Gundersen^a and Jesper Kallestrup^b

^aDepartment of Philosophy, School of Culture and Society, Aarhus University, Aarhus, Denmark; ^bDepartment of Philosophy, University of Aberdeen, Aberdeen, UK



ABSTRACT

You've just read the first sentence of this paper. Would you have read it if some butterfly in Brazil had had some extra nectar for breakfast? You probably think so. But this trivial observation apparently has very dramatic consequences. For instance, it seems to imply that you would have read that very sentence even if someone had offered you \$1.000.000 not to do so. This paper is about what thus looks like a paradox in that a counterintuitive conclusion can seemingly be derived from plausible premises and assumptions. The key is to recognise that 'you would have read the sentence if the butterfly had feasted' admits of distinct readings: one on which it is false, which is the traditional counterfactual implying causal relevance, and another on which it is true, which we call an irrelevant semifactual. While a fully satisfactory solution would need to develop and defend a semantic analysis of such conditional sentences, our modest aim is merely to sketch how the paradox might be solved. The underpinning theoretical work is for a different paper.

ARTICLE HISTORY Received 25 March 2022; Accepted 13 August 2023

KEYWORDS Counterfactuals; irrelevant semifactuals; semantics

You've just read the first sentence of this paper. Would you have read it if some butterfly in Brazil had had some extra nectar for breakfast? You – and most other sensible people – probably think so. True, small deviations from the actual course of events may bring about more widespread and dramatic changes.¹ Many things could be affected by such an extra nectar intake. But not *everything*. Many other things would remain unaffected, including the fact that you just read the first sentence of this paper.²

CONTACT Jesper Kallestrup  jesper.kallestrup@abdn.ac.uk  Department of Philosophy, University of Aberdeen, 50-52 College Bounds, Aberdeen AB24 3DS, UK

¹As witnessed by the so-called *butterfly effect*. See Lorenz (1963).

²If you are sceptic, feel free to tighten up the time index for the consequent until you feel confident, e.g., the butterfly had the extra nectar only a few minutes or seconds prior to your reading.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

However, this trivial observation apparently has very dramatic consequences. For instance, it seems to imply that you would have read that very sentence *no matter what* – even if, say, someone had offered you \$1.000.000 not to do so.

This paper is about what thus looks like a *paradox* in that a counterintuitive conclusion can seemingly be derived from plausible premises and assumptions. The plan is as follows. In (§1) it is explained why you *would* have read the sentence even if our butterfly had had an extra breakfast, and then in (§2) why it follows that you would have read it *no matter what*. The key is in (§3) to recognise that our counterfactual, and others like it, admit of distinct readings: one on which it is false, and another on which it is true, that you would have read the sentence if the butterfly had feasted. While a fully satisfactory solution would need to develop and defend a *semantic analysis* of such conditional sentences, our modest aim is merely to sketch how the paradox might be solved. The underpinning theoretical work is for a different paper.

1. Why you would

Why, then, is it that you would have read the first sentence even if some Brazilian butterfly had drunk a bit more nectar? As hinted at above, the intuitive answer is that the behaviour of the butterfly is completely *irrelevant* when considering what might have prevented you from doing so. Your reading – and a whole lot of other things, for that matter – would still have taken place regardless of whether the butterfly had or had not feasted. When it comes to your reading, the eating habits of a remote butterfly is simply neither here nor there.

In slightly more technical parlance, let an ‘irrelevant semifactual’ be a counterfactual with a true consequent C (hence *semifactuals*) where the antecedent A is irrelevant to the truth of C (hence *irrelevant semifactuals*). A defining feature of irrelevant semifactuals is the following equivalence:

(E) $A > C$ is an irrelevant semifactual if and only if $\neg A > C$ is an irrelevant semifactual.

For we evaluate the truth of $A > C$ and $\neg A > C$ by considering the antecedent while holding fixed as much as possible about the rest of the world (staying as close as possible to the actual world).³ And since,

³See e.g. Kvat (1986, 44). These intuitions are nicely mirrored in the standard Lewis/Stalnaker semantics for counterfactuals according to which a counterfactual $A > C$ is true at a world iff all the closest A -

by hypothesis, C is unrelated to A , the assumption of A , or $\neg A$, will change nothing with respect to C . Hence, if C is true in the actual world, C is also true both in the closest A -worlds and in the closest $\neg A$ -worlds. Thus, A 's irrelevance to some true C ensures the truth of both $A > C$ and $\neg A > C$.⁴

In support of this line of thinking, we may cite the example of Morgenbesser's coin.⁵ Someone tosses an indeterministic coin and, while the coin is in mid-air, offers you good odds that it will land heads. You refuse, and the coin lands heads. It is now intuitively true that:

(1) If you had bet heads, you would have won.

And (1) owes its truth to the following irrelevant semifactual:

(2) If you had bet heads, the coin would still have landed heads.

The idea is, again, that the antecedent (your betting) is irrelevant to the de facto true consequent (the coin landing heads), and so, the irrelevant semifactual (2) should be evaluated as true. So should the semifactual:

(3) If you had *not* bet heads, the coin would still have landed heads.

After all, your betting behaviour, whatever it may be, does not influence the trajectory of the coin. In short, irrelevant semifactuals are always true:

(4) $(C \wedge (A \text{ is irrelevant to } C)) \rightarrow (A > C)$.⁶

If C is true, and A is irrelevant to the truth of C , C would still have been the case if A had been the case. As a special instance of (4) we get: if you did in fact read the first sentence of this paper, and the eating habits of some Brazilian butterfly is irrelevant to your reading, then you *would* have read the first sentence if that Brazilian butterfly had had an extra breakfast.

worlds are C -worlds (with closeness accounted for in terms of overall similarity with respect to laws of nature and particular facts).

⁴Cf. standard possible world semantics for counterfactuals. The same line of reasoning is easily adapted to a semantics modelled on branching time framework, as in Wawer and Wroński (2014). We shall later make the notion of an irrelevant semifactual more precise, and also introduce a notation $A \triangleright C$ different from $A > C$.

⁵As cited in Slote (1978, 27).

⁶The first conjunct in the antecedent in (4) ensures that $A > C$ is a semifactual, i.e., a counterfactual with true consequent, and the second conjunct states that $A > C$ is an irrelevant semifactual. The consequent in (4) then states that such an irrelevant semifactual ($A > C$) is true.

2. Why you would not

The key intuition why you would have read the first sentence despite the butterfly's feasting is thus captured by (4), i.e. that irrelevant semifactuals are always true. But now the question is what 'A is irrelevant to C' in (4) means. One very natural way of characterising what it is for A to be *irrelevant* to C is to say that whether C occurred is *causally* independent of whether A occurred; indeed, the two examples in (§1) of the butterfly and of coin-flipping involve precisely a causal conception of irrelevance. Schaffer (2004), Edgington (2004) and Bennett (2003, 234-7) all take 'irrelevance' to be understood as causal independence in this sense.⁷

How then should 'causal independence' be understood? Well, causal dependence is typically spelled out as *counterfactual* dependence, following Lewis' (1973; 1986)⁸ proposal that:

(5) C is causally dependent on A if and only if $(A > C) \wedge (\neg A > \neg C)$.

If 'relevance' is interpreted as causal dependence in this Lewisian sense, *irrelevance* should be expressed as the negation of the conjunction in (5):

(6) $\neg((A > C) \wedge (\neg A > \neg C))$.

At this point one may worry that (5) fails to adequately characterise causal *relevance* (and hence that (6) fails to capture causal irrelevance). The reason would be that $(A > C) \wedge (\neg A > \neg C)$ is sufficient, but not necessary, for causal relevance. Suppose soldier S shoots and kills a convict, but had S not fired his rifle, S* would have fired hers and killed the convict. While both take aim, S* desists when seeing that S has already pulled the trigger. The problem is that the second conjunct of $(A > C) \wedge (\neg A > \neg C)$ is false when A stands for 'S shoots' and C for 'victim is killed'. The reply is that the intuition elicited by such *pre-emption* cases pertains to *causation proper* rather than the different notion of relevance as causal dependence. And there's no reason why $(A > C) \wedge (\neg A > \neg C)$ cannot characterise the latter.⁹ Yablo (1992; 1997; 2003) is a case in point.

⁷There are of course other notions of (ir)relevance in the literature, such as metaphysical, epistemic, or semantic (ir)relevance, but these shall not detain us here, as our paradox clearly pertains to a causal notion.

⁸Lewis (op. cit.) required that A and C be numerically distinct and actually occurring events, and that the counterfactuals not be backtracking.

⁹Indeed, as Lewis (1973) himself observed, causal dependence is sufficient, but not necessary, for causation. For while causation is transitive, causal dependence is not. To overcome this problem Lewis (op. cit.) defined causation in terms of the ancestral of causal dependence: C causes distinct event E if and

Without endorsing his specific account, the reasoning is illustrative of the distinctions we wish to draw. He imposes a *proportionality constraint* on causation, which says that a cause must be specific enough but not too specific for its effect. Now consider Shoemaker's (2007, 14) pigeon Alice, who is conditioned to peck at scarlet to the exclusion of other shades of red. Suppose she is presented with a scarlet chip. Scarlet-ness is the best candidate for a cause, screening off redness for not being specific enough. But, by being causally necessary, redness is still relevant to the pecking. Importantly, both counterfactuals in (5) are true: (redness > pecking) is automatically true given that redness is an actually instantiated property and the pecking is an actually occurring event, and (\neg redness > \neg pecking) is also true in that redness is necessitated by scarlet-ness, which causes the pecking. The example thus serves to illustrate that the distinct notion of relevance can do important explanatory work when understood counterfactually as in (5).¹⁰

With the foregoing in mind, we can now proceed to rephrase (4) as:

$$(4^*) (C \wedge \neg((A > C) \wedge (\neg A > \neg C))) \rightarrow (A > C)$$

However, (4*) entails what we shall call *Strengthened Conjunction Conditionalization* (SCC)¹¹:

$$(7) C \rightarrow (A > C)$$

To see this, assume the *first* conjunct in the antecedent of (4*):

$$(8) C$$

only if there is a chain of causal dependences between *C* and *E*. As is familiar, he later (2000) refined this account in terms of a pattern of counterfactual dependence of alterations of *E* upon alterations of *C*. We shall continue to restrict causal dependence, and hence relevance, to whether-whether counterfactuals, but everything we say would apply *mutatis mutandis*, were we to include the when-and and how-how counterfactuals that also are part of Lewis' notion of influence.

¹⁰In fact, (5) can also handle Yablo's (1992, 257) case of pigeon Sophie who is trained to peck at red to the exclusion of other colours. She also pecks upon being shown a scarlet chip, but in this case, redness is the cause, screening off scarlet-ness for being too specific. But scarlet-ness is still relevant to the pecking, in the sense of being causally sufficient (without being a sufficient cause). Scarlet-ness necessitates redness which causes the pecking. Again, both counterfactuals in (5) are true: (scarlet-ness > pecking) is automatically true for the same reason as before, and (\neg scarlet-ness > \neg pecking) is also true, at least if we follow Lewis' recommendation (1973; 2000), cf. Bennett (2003), that such counterfactuals should be given a deletion, rather than a replacement, reading. That is to say, the closest world in which the antecedent is true is not a world in which the chip is some other specific shade of red, but rather one in which it is no particular colour, and in such a world there is no pecking.

¹¹For more on *Conjunction Conditionalization* $(A \wedge C) \rightarrow (A > C)$, see Walters and Williams (2013).

Assume then, for *reductio*, that the consequent in (4*) is false:

$$(9) \neg (A > C)$$

Then, by *modus tollens*, the antecedent in (4*) must be false too. But according to (8), the first conjunct in the antecedent is *true*. Hence, the *second* conjunct must be false:

$$(10) \neg \neg ((A > C) \wedge (\neg A > \neg C))$$

By double negation elimination, this leads to:

$$(11) (A > C) \wedge (\neg A > \neg C)$$

However, the first conjunct in (11), and thus (11) itself, contradicts (9). Hence, the assumption in (9) must be false:

$$(12) \neg \neg (A > C)$$

So, by double negation elimination, we get:

$$(13) A > C$$

This proves that (13) is true if assumption (8) is. That is, by conditional proof, we can infer (SCC):

$$(14) C \rightarrow (A > C)^{12}$$

It follows that if 'irrelevance' is interpreted as causal independence, endorsement of (4) commits one to (SCC). But the problem is now that (SCC) is a *highly implausible* principle;¹³ adopting it nearly collapses the entire semantics for counterfactuals to the truth-functional semantics characterising the material conditional. Thus, counterfactuals with true consequents, and counterfactuals with false consequents and true

¹²A more succinct, but also less easily comprehensible, proof of (SCC) from (4*) goes as follows: Suppose that *C* is true, but *A > C* is false. Then it holds that $C \wedge \neg(A > C)$, which immediately implies that $C \wedge \neg((A > C) \wedge (\neg A > \neg C))$. But then (4*) implies that *A > C*, which contradicts the *reductio* hypothesis. Thanks to an anonymous referee.

¹³And besides, (SCC) threatens – like *Conjunction Conditionalization* ($A \wedge C \rightarrow (A > C)$) does – to render counterfactuals infelicitous for their various theoretical tasks in conditional theories of knowledge, disposition, mind, aesthetics, etc.

antecedents, will receive the same truth value as their material cousins do. Only counterfactuals with false antecedents and false consequents will potentially differ in truth value from the corresponding material conditional.

(SCC) claims that when C as a matter of fact is the case, C would have been the case *no matter what*. For any A – regardless of whether A is relevant or irrelevant to C – whenever C is actually true, C would have been the case if A had been the case. For instance, given that you did read the first sentence of this paper, (SCC) dictates that you would have read it come what may. Apart from a somewhat deterministic flavour, it faces you with a glitch in practical reasoning in that you would have read the first sentence even if someone had offered you \$1.000.000 not to do so. Even worse: if you had *not* read the first sentence, then you had indeed read that sentence.

At this juncture, two distinct, but related, worries spring to mind. The first is that if, as we propose, the causal dependence of C on A is best understood as the conjunction $(A > C) \wedge (\neg A > \neg C)$, then causal *independence* should be understood, not as in (6), i.e. the negation of the conjunction, but as the negation of both conjuncts: $\neg(A > C) \wedge \neg(\neg A > \neg C)$. In other words, if C is causally independent of A , then A is neither causally sufficient nor causally necessary for C .¹⁴ Suppose that (6) is true but only because one of the conjuncts is false. Then, even if C may not be causally *dependent* on A , at least by the definition in (5), it's far from clear that C is causally *independent* of A . This may generally be the case for *background conditions*, which are necessary, but insufficient, to bring about an effect, e.g. the presence of oxygen (A) is not causally sufficient for the lighting of a match (C), and so the first conjunct in (6) $A > C$ is false, but if oxygen is absent, the match would not light, and so the second conjunct in (6) $\neg A > \neg C$ is true. In this case we would say that the presence of oxygen is causally relevant to, but not causally sufficient for, the lighting of the match. If that is correct, (6) does not capture our intuitive notion of causal irrelevance. (6) should rather be expressed as the negation of both conjuncts:

$$(6^*) \neg(A > C) \wedge \neg(\neg A > \neg C)$$

And, accordingly, (4) should be rephrased, not as (4*), but as:

¹⁴Thanks to an anonymous referee for raising this point.

$$(4^{**}) (C \wedge (\neg(A > C) \wedge \neg(\neg A > \neg C))) \rightarrow (A > C)$$

We do not share this worry. In our opinion, (6) is the most appropriate expression of *irrelevance*. We believe that (5) offers a plausible account of relevance, and we take two events to be irrelevant to each other just in case they are not relevant to each other, where these are all causal notions. However, rather than arguing for this claim, it suffices to show – for our purposes – that nothing much hangs on it. If you happen to favour (6*) as the most appropriate formulation of irrelevance, you will still, by standard Lewisian logic, find yourself stuck with the unpalatable (SCC). The reason is simply that (SCC) *also* follows from (4**). Here's how. Assume the *first* conjunct in the antecedent of (4**):

$$(15) C$$

Assume then, for *reductio*, that the consequent in (4**) is false:

$$(16) \neg(A > C)$$

Then, by *modus tollens*, the antecedent in (4**) must be false too. But according to (15), the first conjunct in that antecedent is *true*. Hence, its *second* conjunct must be false:

$$(17) \neg(\neg(A > C) \wedge \neg(\neg A > \neg C))$$

Note that (17) is also a negation of a conjunction, which means that at least one of its embedded conjuncts must be false. (16) suggests that it is the second conjunct. By double negation elimination, this leads to:

$$(18) \neg A > \neg C$$

Assume then, for *reductio*, that *A* is false:

$$(19) \neg A$$

By counterfactual *modus ponens*, this leads to:

$$(20) \neg C$$

However, (20) contradicts (15). Hence, the assumption in (19) must be false, and so by double negation elimination, this gives us:

(21) A

Now, by applying conjunction conditionalisation (CC) on (21) and (15) we get:

(22) $A > C$

This proves that the assumption in (16) is false and hence, by double negation, that (22) is true if the assumption (15) is true. That is, by conditional proof we can infer (SCC):

(23) $C \rightarrow (A > C)$

The second, related worry arises from the fact that we are interested in semifactuals with irrelevant antecedents, which are evaluated by considering whether the occurrence of A would, or would not, have *causally prevented* C from occurring. Hence, the appropriate notion of (ir)relevance should be *negative*: A is negatively relevant to C just in case the occurrence of A would causally prevent C from occurring; otherwise, A is irrelevant to C . But we interpret relevance *positively*, as causal dependence of C on A , cashed out counterfactually as in (5). Moreover, if such notion of negative (ir)relevance is adopted, one may suspect that our paradox is a non-starter, because the absurd (SCC) may no longer be derived from a corresponding version of (4), i.e. $(C \wedge (A \text{ is negatively irrelevant to } C)) \rightarrow (A > C)$.¹⁵

The point is well taken, but the first question is how to understand the envisaged notion of negative relevance to do with A causally preventing C . Here is a proposal that naturally springs to mind: A causally prevents C if and only if A caused $\neg C$. That is, for A to prevent C is for A to *cause the absence* of C , rather than for A *not to cause the presence* of C . On Lewis' view, such absences, or omissions in particular, as non-occurrences of events, are not themselves events, but they can still feature as *causal relata*.¹⁶ It is thus still the case on his view that causation involving absences is a matter of a chain of causal dependences. That means we

¹⁵We are grateful to an anonymous referee for alerting us to this objection.

¹⁶Indeed, on Lewis' view, their causal efficacy is grounded in true counterfactuals. For instance, the gardener failing to water the plants caused them to die, because they would not have died, had the gardener watered them.

should continue to use counterfactuals to flesh out prevention, and hence the notion of negative relevance; the only difference is that C is a non-occurrence of an event: A causally prevents C , i.e. C is negatively relevant to A , if and only if $A > \neg C \wedge \neg A > C$.¹⁷ The next question is now whether we can deduce the unpalatable (SCC) $C \rightarrow (A > C)$ from $(C \wedge (A \text{ is negatively irrelevant to } C)) \rightarrow (A > C)$, where negative irrelevance is the negation of negative relevance: $\neg(A > \neg C \wedge \neg A > C)$.¹⁸ Basically, to say that A is not negatively relevant to C is to say that it's not the case that the occurrence of A would prevent C from occurring. Consider again (4), but where irrelevance is understood in this negative sense:

$$(4^{***}) (C \wedge \neg((A > \neg C) \wedge (\neg A > C))) \rightarrow (A > C)$$

Now recall from (§1) the equivalence:

(E) $A > C$ is an irrelevant semifactual if and only if $\neg A > C$ is an irrelevant semifactual

(E) means that A 's irrelevance to some true C ensures the truth of both $A > C$ and $\neg A > C$. Given these defining features of irrelevant semifactuals, we can safely expand on the consequent in (4^{***}):

¹⁷Maybe our objector has a different notion of negative relevance in mind according to which A can be negatively relevant to C despite lack of causal dependence between A and not- C and not- A and C , i.e., even if $\neg((A > \neg C) \wedge (\neg A > C))$. Suppose for instance that our Brazilian butterfly would have flapped its wings in order to get some extra nectar and so, via some butterfly effect, caused a storm outside your window. Suppose furthermore that you are slightly more likely not to read philosophy papers when it is windy. In that case the feasting would not have *prevented* you from reading the first sentence of this paper. There would still be no *causal dependency* between the feasting and your not reading. Nevertheless, the feasting would have slightly decreased the *objective probability* that you did so and in that sense been negatively relevant to the reading. The worry can be rephrased thus: A may be either *strongly* and *weakly negatively* relevant to C , where $\neg((A > \neg C) \wedge (\neg A > C))$ only rules out strong negative relevance. The reply is that a more precise definition of irrelevant semifactuals likewise should be offered in terms of strong (ir-)relevance; that is, as counterfactuals with true consequents where the antecedent is not strongly relevant for the consequent (in the negative sense specified). Such counterfactuals are always true. Obviously, much more needs to be said about the exact borderline between strong and weak relevance. How much, exactly, must A decrease the objective probability of C in order to constitute counterfactual, and hence causal, dependence between A and not- C to qualify as strongly relevant? For present purposes, it will suffice to constrain the domain of A and C to instances where A is strongly (ir-)relevant to C . Henceforth, '(ir-)relevance' should thus be understood as shorthand for 'strong (ir-)relevance'.

¹⁸Again, one may worry that negative irrelevance may be better captured as negative causal *independence*: $\neg(A > \neg C) \wedge \neg(\neg A > C)$; rather than as lack of negative causal *dependence*: $\neg((A > \neg C) \wedge (\neg A > C))$. We do not share this worry. But, again, it suffices to stress that nothing essentially hangs on this: (SCC) also follows from (4) when rephrased in terms of negative causal *independence*. To see this, work through step (24) to (28) below but with the amendments added above in steps (17) to (23). Thanks to an anonymous referee for raising this point.

(4****) $(C \wedge \neg((A > \neg C) \wedge (\neg A > C))) \rightarrow ((A > C) \wedge (\neg A > C))$ ¹⁹

With (4****) in mind, assume now for *reductio*:

(24) $C \wedge \neg(\neg A > C)$

From (24) we can infer the negation of the consequent of (4****):

(25) $\neg((A > C) \wedge (\neg A > C))$

And given the first conjunct C from (24), the negation of the second conjunct in the antecedent of (4****) follows via double negation elimination:

(26) $(A > \neg C) \wedge (\neg A > C)$

From (26) we can infer:

(27) $(\neg A > C)$

But reflect that (27) contradicts our assumption in (24), which is therefore false:

(28) $C \rightarrow (\neg A > C)$

In sum, (SCC) $C \rightarrow (A > C)$ *does follow* from $(C \wedge (A \text{ is negatively irrelevant to } C)) \rightarrow (A > C)$.²⁰ Because the notion of negative (ir)relevance therefore makes no difference to our proof, we shall revert to our original – simpler and more familiar – conception of relevance as Lewis-style causal dependence, namely:

¹⁹One may worry that (E) is inconsistent with (4****) since they together lead to (4****). The reason is that (4****) apparently commits us to:

(E*) $C \wedge (\neg((A > \neg C) \wedge (\neg A > C)))$ if and only if $C \wedge \neg(\neg A > \neg C) \wedge (A > C)$,

which simply cannot be true. Consider A and C such if A were the case, C would be the case, and if A were not the case, C would not be the case. In other words, $A > C$ and $\neg A > \neg C$. Suppose furthermore that C is true. Surely, such A and C must be possible. But then it is easy to see that $\neg A$ is *not* negatively irrelevant to C (since it is true that $A > C$ and that $\neg A > \neg C$), but that A is negatively irrelevant to C (it is neither true that $A > \neg C$ nor that $\neg A > C$). This means that (E*) is false in this case.

This worry, which we owe to an anonymous referee, regarding (E*) is real. However, the good news is that (E*) does not follow from (4****). In order to derive (E*) from (4****), an additional assumption is needed:

(#4****) $C \wedge \neg((A > \neg C) \wedge (\neg A > C)) \leftrightarrow A > C$

And, to be sure, (#4****) is certainly not an assumption that we share. Furthermore, to argue for something like (#4****) would seem to be futile, unless one from the outset adopts the idea that semi-factuals are governed by a semantics distinct from that governing traditional counterfactuals.

²⁰To be clear, (28) is a substitution instance of (SCC).

(5) C is causally dependent on A if and only if $(A > C) \wedge (\neg A > \neg C)$.

These considerations put (4*) or any of (4**), (4***) or (4****) – and hence (4) – under a fair amount of pressure. But (4) is, recall, a way of formalising what essentially drives the intuition from (S1) that you *would* have read the sentence if the butterfly had feasted, namely that such a conditional holds whenever the antecedent is irrelevant to a true consequent. Giving up (4) seems to completely undermine the rationale for thinking that you would have read that sentence regardless. So maybe you would not – after all – have read that sentence if the butterfly had had an extra nectar.²¹

3. Why you both would and wouldn't

Our proposed solution to this unpleasant dilemma is to claim that you both would and wouldn't have read the first sentence if the butterfly had feasted. Or, more precisely, that the conditional in question is *ambiguous* (in a manner to be specified below). In one sense it is true; in another false. There is, in fact, a clear sense in which it is false that you would have read the sentence if the butterfly had feasted. To see this, consider:

(29) If I had told you a joke, I would have received a large fine.

(29) is arguably false, as there is no causally relevant dependence between joke telling and being fined. The implication that (29) carries any such connection is simply false. However, we could also argue that (29) is true. Assume that I have indeed received a large fine (due to parking illegally) and then ask yourself: would I have been fined, had I told you a joke? Given the actual truth of the consequent, and since telling you a joke would have done nothing by way of changing that fact, had I told you a joke, I would still have been fined. A natural way to explicate the two readings – on which it is respectively false and true – is to qualify the consequent:

(29*) If I had told you a joke, I would *consequently* have received a large fine.

(29**) If I had told you a joke, I would *still* have received a large fine.

²¹Strictly speaking, to deny (4) where C is true and irrelevant to A is to claim that $\neg(A > C)$. But, as Williams (2010) observes, $\neg(A > C)$ and $(A > \neg C)$ are naturally treated as equivalent when the antecedent is possibly true.

With that in mind, consider now:

(30) If some Brazilian butterfly had had some extra nectar for breakfast, you would have read the first sentence of this paper,

which equally admits of similarly distinct readings – one that implies a causal dependence between antecedent and consequent, and one that does not:

(30*) If some Brazilian butterfly had had some extra nectar for breakfast, you would *consequently* have read the first sentence of this paper.

(30**) If some Brazilian butterfly had had some extra nectar for breakfast, you would *still* have read the first sentence of this paper.

To see this, suppose you skipped the first sentence of this paper. Thus, you jump directly to the question posed in the second sentence: ‘*Would* you have read the first sentence, if some butterfly in Brazil had had some extra nectar for breakfast?’ In that case, you would probably find the continuation most disagreeable: ‘You – and most other sensible people – probably think so’. And the reason for this is that you now consider (30*). That is, you take the antecedent and the consequent to be false, and then consider whether the truth of the antecedent would have brought about the truth of the consequent. Since there is no causally relevant connection between antecedent and consequent, this is not the case and, accordingly, you (and most other sensible people) take (30*) to be *false*.

As things turned out, you *did* read the first sentence. Thus, when confronted with the question in the second sentence, you ponder whether (30**) is true. That is, you take the consequent to be *true* and consider whether it would still have been so if the antecedent had obtained, i.e. whether the obtaining of the antecedent would have *prevented* the consequence from obtaining. And since there is no causally relevant connection between antecedent and consequent, that is not the case and, accordingly, you take the irrelevant semifactual (30**) to be *true*.

The final question is how this semantic distinction actually solves the paradox.²² In order to spell this out in detail, it will prove useful to introduce some new notation. Let *small* arrow ($A \succ C$) stand for semifactuals such as in (29**) and (30**), expressing that *C* would *still* happen, if *A*

²²Thanks to an anonymous referee for requesting (and sketching) an answer to this question.

occurred; and let *big* arrow ($A > C$) stand for traditional counterfactuals such as in (29*) and (30*), expressing that C *would* happen if, and as a *consequence* of, A . So far, we have used ($A > C$) to formalise both semifactuals and counterfactuals, but henceforth we shall reserve it to denote the latter. With that in mind, reconsider (4*). On the one hand, it should be obvious that (4*) is meant to be a principle for the introduction of semifactuals (more precisely, for the introduction of irrelevant semifactuals). Indeed, we argued in (S2) that (4) captures the intuition that irrelevant semifactuals are always true, and (4*) is a natural way of using counterfactual dependence to render the pertinent notion of relevance more precise. The appropriate formalisation of the consequent of (4*) should accordingly be ($A \succ C$). The Lewis-style account of causal dependence, on the other hand, assumes the conventional counterfactual ($A > C$). Thus, if irrelevance is understood as absence of causal dependence, as we propose, it should be spelled out using the traditional counterfactual ($A > C$). This means that (4*) should be reformulated as follows:

$$(\#4^*) (C \wedge \neg((A > C) \wedge (\neg A > \neg C))) \rightarrow (A \succ C)$$

As for (SCC), we ought likewise to distinguish between its two versions, depending on how the consequent is interpreted:

$$(\text{SCC for semifactuals}) C \rightarrow (A \succ C)$$

$$(\text{SCC for counterfactuals}) C \rightarrow (A > C)$$

Now, obviously, neither of these versions of (SCC) can be derived from (#4*) in the manner proposed in (S2). The proof to that effect, understood as a proof of (SCC for semifactuals) is blocked from line (2) to (3) since ($A > C$) does *not* entail ($A \succ C$). Nor is there any possibility of deriving (SCC for counterfactuals) from (#4*) by the crisper version of that argument in (15) – (19). The *reductio* step would presuppose that ($A \succ C$) entails ($A > C$). But ($A \succ C$) does *not* entail ($A > C$). Either way, the paradox is avoided. And by extension, the same considerations apply to (4**), (4***), and (4****) *mutatis mutandis* when these are properly formalised with ($A \succ C$) in the consequent. So, no matter how (4) is interpreted, no untoward consequences follow, if only the relevant conditionals are disambiguated in this way.

4. Conclusion

In (S1) we argued that (30) is indeed true. However, in (S2) we then proceeded to question the driving principle, (4), behind that reasoning and,

eventually, to suggest that (30) is *false*. We are now in a position so see that *both* of these lines of reasoning may be perfectly valid. (30) (and (4) for that matter) *is* true – when (30) is interpreted as:

(#30**) $(A \succ C)$,

with *A* standing for some Brazilian butterfly having some extra nectar for breakfast, and *C* for reading the first sentence of this paper. At the same time (21) is also *false* – when interpreted as:

(#30*) $(A > C)$

In other words, maybe the correct lesson to draw from these cases is simply to grant the truth of semifactuals such as (29**) and (30**), but then to point out that these irrelevant semifactuals differ in content from (29*) and (30*). In that case, the truth of (29*) and (30*) does not automatically follow from the truth of (29**) and (30**), just like the falsehood of (29*) and (30*) does not automatically entail the falsehood of (29**) and (30**). Rather (29) and (30) are – at least in certain contexts – *equivocal* between, on the one hand, (29*) and (30*), and, on the other hand, (29**) and (30**). The two sets of counterfactuals may instead represent two asymmetrical modes of thinking; insisting that they are equivalent, apt to be treated as if they were semantically on a par, is deeply problematic; or so we argued. Counterfactuals such as (29*) and (30*) affirm a *causal relevance*, as spelled out in terms of a counterfactual dependence, between antecedent and consequent, whereas the corresponding semifactuals (29**) and (30**) cast doubt on the counterfactual dependence between the antecedent and the negated consequent. They suggest the antecedent would *not have prevented* the consequent.

A similar proposal has been aired by other philosophers of modality such as Goodman (1947) with the explicit proposal that:

(31) $(A \succ C) \leftrightarrow \neg(A > \neg C)$ ²³

In the same spirit, Pollock (1976) proposed that counterfactuals have two distinct sets of truth-conditions; one where there is a counterfactual *dependence* between antecedent and consequent, and one where there is a *lack* of such dependence. Pollock (op. cit.) did, though, subsume both

²³One difficulty with this particular proposal is that $A \succ C$ is then given the same semantic treatment as *might* counterfactuals, which seems implausible under the assumption of interdefinability between *might* – and *would*–counterfactuals.

sets of conditions under *one* unifying (disjunctive) semantics. However, it would be much more natural, as Goodman (op. cit.) proposed, to work with *two* distinct semantic accounts, one based on each set of truth-conditions, such that counterfactuals and semifactuals are each assigned their own distinctive semantics.²⁴

Never mind the details of Goodman or Pollock's proposals, the important point for our purposes is that there be two distinct sets of truth-conditions, such that one cannot univocally infer the truth of a corresponding counterfactual ($A > C$) from the truth of an irrelevant semifactual ($A \triangleright C$), or the falsity of the latter from the falsity of the former. That is to say, the inference from (4**) to (SCC) is blocked by our proposed solution to what seems like a paradox. Reverting to our opening question, the answer is therefore that you both would and would not have read the initial sentence of this paper if some butterfly in Brazil had feasted. You would in the sense of (30**), but you would not in the sense of (30*).

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.
- Bennett, K. 2003. "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It." *Noûs* 37: 471–497.
- Douven, I. 2008. "The Evidential Support Theory of Conditionals." *Synthese* 164: 19–44. doi:10.1007/s11229-007-9214-5.
- Douven, I. 2016. *The Epistemology of Indicative Conditionals*. Cambridge: Cambridge University Press.
- Edgington, D. 2004. "Counterfactuals and the Benefit of Hindsight." In *Cause and Chance: Causation in an Indeterministic World*, edited by P. Dove, and P. Noordhof, 12–27. London: Routledge.
- Goodman, N. 1947. "The Problem of Counterfactual Conditionals." *The Journal of Philosophy* 44: 113–128. doi:10.2307/2019988.
- Kvart, I. 1986. *A Theory of Counterfactuals*. Indianapolis: Hackett Publishing Company.
- Lewis, D. 1973. "Causation." *The Journal of Philosophy* 70: 556–567. doi:10.2307/2025310.
- Lewis, D. 1986. "'Causation' and Postscripts to 'Causation'." In *Philosophical Papers: Volume II*, edited by D. Lewis, 172–213. Oxford: Oxford University Press.
- Lewis, D. 2000. "Causation as Influence." *The Journal of Philosophy* 97: 182–197. doi:10.2307/2678389.

²⁴Such a division also seems to be obligatory for any probability-based semantics, because only the former, counterfactual *dependence*, can be modelled probabilistically. For more recent discussion see Douven (2008; 2016), Shanks (1995) and van Rooij & Schulz (2019; 2022).

- Lorenz, E. N. 1963. "Deterministic Nonperiodic Flow." *Journal of the Atmospheric Sciences* 20 (2): 130–141. doi:[10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- Pollock, J. 1976. *Subjunctive Reasoning*. Philosophical Studies Series in Philosophy 8. D. Reidel Publishing Company, Dordrecht and Boston.
- Schaffer, J. 2004. "Counterfactuals, Causal Independence and Conceptual Circularity." *Analysis* 64: 299–309. doi:[10.1093/analys/64.4.299](https://doi.org/10.1093/analys/64.4.299).
- Shanks, D. R. 1995. *The Psychology of Associative Learning*. Cambridge: Cambridge University Press.
- Shoemaker, S. 2007. *Physical Realization*. Oxford: Oxford University Press.
- Slote, M. 1978. "Time in Counterfactuals." *The Philosophical Review* 87: 3–27. doi:[10.2307/2184345](https://doi.org/10.2307/2184345).
- van Rooij, R. A. M., and K. Schulz. 2019. "Conditionals, Causality and Conditional Probability." *Journal of Logic, Language and Information* 28: 55–71. doi:[10.1007/s10849-018-9275-5](https://doi.org/10.1007/s10849-018-9275-5).
- van Rooij, R., and K. Schulz. 2022. "Causal Relevance of Conditionals: Semantics or Pragmatics?" *Linguistics Vanguard* 8 (4): 363–370. doi:[10.1515/lingvan-2021-0030](https://doi.org/10.1515/lingvan-2021-0030).
- Walters, L., and J. R. G. Williams. 2013. "An Argument for Conjunction Conditionalization." *The Review of Symbolic Logic* 6: 573–588. doi:[10.1017/S1755020313000191](https://doi.org/10.1017/S1755020313000191).
- Wawer, J., and L. Wroński. 2014. "Towards a new Theory of Historical Counterfactuals." In *Logica Yearbook 2014*, edited by P. Arazim, and M. Dancak, 293–310. Hejnice: College Publications.
- Williams, J. R. G. 2010. "Defending Conditional Excluded Middle." *Noûs* 44: 650–668.
- Yablo, S. 1992. "Mental Causation." *The Philosophical Review* 101: 245–280. doi:[10.2307/2185535](https://doi.org/10.2307/2185535).
- Yablo, S. 1997. 'Wide Causation', *Philosophical Perspectives* 11, *Mind, Causation and World*, 251-281.
- Yablo, S. 2003. 'Causal Relevance', *Philosophical Issues* 13, *Philosophy of Mind*, 316-328.