

Counterfactuals, Accessibility, and Comparative Similarity

abstract

Berit Brogaard and Joe Salerno (2008) have defended the validity of counterfactual hypothetical syllogism (CHS) within the Stalnaker-Lewis account. Whenever the premisses of an instance of CHS are non-vacuously true, a shift in context has occurred. Hence the standard counterexamples to CHS suffer from context failure. Charles Cross (2011) rejects this argument as irreconcilable with the Stalnaker-Lewis account. I argue against Cross that the basic Stalnaker-Lewis truth condition may be supplemented in a way that makes (CHS) valid. Yet pace Brogaard and Salerno, there are alternative ways of spelling out the basic truth condition which are standard in most debates; and given these ways, the counterexamples to CHS are successful.

Charles Cross (2011) has taken issue with Berit Brogaard and Joe Salerno's argument for the validity of counterfactual hypothetical syllogism (CHS).¹ The classical example:

(Hoover)

[H1] If J. Edgar Hoover had been a communist, he would have been a traitor. [H2] If he had been born a Russian, he would have been a communist. [H3] Therefore, if he had been born a Russian, he would have been a traitor. (Brogaard and Salerno 2008: 39)

Brogaard and Salerno argue for the following claim:

¹ I am very grateful to Professor Cross for intense discussion.

The ability of the possible worlds account to explain the failure of these inferences is thought to be one of its great strengths. However, as we will argue, the results yielded rest on a contextual fallacy. (Brogaard and Salerno 2008: 40)

Cross, in contrast, has the following aim:

I will argue that no contextual fallacy is committed in the standard Stalnaker/Lewis counterexamples to the rules in question. (Cross 2011: 91)

What is the Stalnaker-Lewis semantics as presupposed by Brogaard and Salerno? Cross notes that it is neither textbook Stalnaker nor textbook Lewis (Cross 2011: 92 ann. 1). Nor can it be identified with the maximal consensus of both positions. In Brogaard and Salerno's view, the Stalnaker-Lewis semantics boils down to the following truth condition:

(SL) $p > q$ is true at possible world w iff q is true at all of the closest possible worlds to w at which p is true. (Cross 2011: 92)

Of course, in order to understand SL, we have to add some implications like a similarity ordering that discerns the set of closest possible worlds. Brogaard and Salerno add:

...a consequence of the standard account is that counterfactuals with impossible antecedents (i.e. counterpossibles) are vacuously true; if there are no closest A-worlds, then vacuously all the closest A-worlds are B-worlds. (Brogaard and Salerno 2008: 40)

So we should keep in mind that the Stalnaker-Lewis semantics Brogaard and Salerno commit themselves to just consists in SL and its implications plus the condition for vacuous truth.

As we will see, Cross argues to the stronger conclusion that Brogaard and Salerno's account is irreconcilable with the Stalnaker-Lewis semantics. I will argue, however, that neither Brogaard and Salerno nor Cross are completely right. Pace Cross, Brogaard and Salerno are right as far as there is a neat way of spelling out SL according to which the Stalnaker-Lewis counterexamples do rest on a contextual fallacy. In contrast, Cross is right in claiming that the *standard* Lewisian way of spelling out SL does *not* lead to the contextual fallacy.

Here is Brogaard and Salerno's defence of CHS. An inference must be valid only provided context is held fixed throughout this inference. Regarding CHS, there are two possibilities. Either context is held fixed; then CHS comes out valid. Or it is not; then the failure of CHS does not impair the claim that CHS is valid in principle. The two alternatives arise as follows. Closest antecedent worlds are determined by holding fixed a context of background facts. In order for H1 to be non-vacuously true, the closest worlds where Hoover is a communist must be worlds where Hoover is a traitor. To achieve this, we hold fixed the contextual background fact that Hoover is the American director of the FBI. Holding this fixed, the antecedent of H2 and H3 is inaccessible. H2 and H3 are vacuously true. The inference is valid. In contrast, in order for H2 to be non-vacuously true, we must give up the background fact that Hoover is the American director of the FBI. If we hold the resulting context fixed, the premiss H1 turns out false and (Hoover) fails for this reason. What remains is that the context of background facts changes such as to make first H1 and then H2 true. Then the inference fails due to context failure. All counterexamples to CHS must be treated in the same way.

Cross denies that for both premisses H1 and H2 to be non-vacuously true, a change in context *must* occur. To be sure, whenever both premisses are true, a change in *what would be true if the antecedent were true* occurs. But this change does not have to amount to a shift in context:

The issue for Brogaard and Salerno is not whether, in real-world examples, *what would be true if p were true* varies with *p*; the issue is whether this variation entails a change of

context. Brogaard and Salerno's argument assumes that a change of context *must* occur in that case, but this assumption is simply wrong as claim about the semantics of the Stalnaker-Lewis conditional.(Cross 2011: 94)

Why is the assumption that a change of context *must* occur wrong? The variation in background facts that occurs whenever H1 and H2 are true may be owed to the variation in the antecedent given a fixed similarity ordering and not to a shift in context. Brogaard and Salerno have not shown otherwise.

I want to argue, however, that the most charitable interpretation of Brogaard and Salerno's way of spelling out SL indeed substantiates their key assumption. Their view is anticipated by Wright's discussion of CHS:

...when a number of counterfactuals are at play in a single context, some single range of relevant worlds –perhaps that, determined à la Lewis, for the most remote of them- governs them all.(Wright 1983: 138)

I introduce a specific notion of background facts held fixed, facts_B , which is subject to the following conditions: Context fixes the similarity ordering by holding fixed facts_B . Different facts_B , different context. A world is accessible in a context iff facts_B obtain at that world. Any difference in accessible worlds must be a difference in facts_B . Hence any difference in accessible worlds must amount to a difference in context.

Cross rejects this notion of a background fact:

Where exactly did Brogaard and Salerno go wrong? The problem is their claim that 'if context must remain fixed when evaluating an argument for validity, the set of contextually determined *background facts* must remain fixed when evaluating an argument involving

subjunctives for validity.’ On a Stalnaker/Lewis account of counterfactuals, context alone does not determine a set of *background facts*. What context determines is a measure of comparative world similarity, and it takes both a comparative world similarity measure and a counterfactual antecedent to determine a set of *background facts*. In short, the set of contextually determined *background facts* is always *antecedent-relative*... (Cross 2011: 94, emphasis on ‘background facts’ mine)

But I do not see any reason within SL to eschew background facts_B. The distinction between worlds being accessible and being inaccessible looms large in the Stalnaker-Lewis semantics. For instance, we may want to constrain accessible worlds by their being nomically possible. This is done by holding fixed the natural laws as background facts_B. And it is plausible to say that *these* background facts are not antecedent-relative. What is special about Brogaard and Salerno’s use of facts_B is that accessibility is both very flexible (it may vary with context from H1 to H2) and very restrictive (mundane possibilities like Hoover being Russian may become inaccessible due to a context shift). Cross introduces a very different notion of background facts, let us call them facts_C, which are partly determined by the contextually shaped similarity ordering together with the antecedent of a counterfactual. On Cross’s epistemic account of facts_C, the background facts held fixed by p are the propositions q such that a given agent would use $p > q$ to settle her beliefs about the truth values of other counterfactuals. This is perfectly legitimate, but so far it is perfectly reconcilable with the conception of Brogaard and Salerno. Just add the subscripts ‘_B’ to the first occurrence of ‘background facts’ and ‘_C’ to the other three occurrences in the passage quoted from Cross. Then it becomes obvious that Brogaard and Salerno can happily agree with what Cross says about the Stalnaker-Lewis account. Facts_B must remain fixed when evaluating counterfactual syllogisms, facts_C vary with the antecedent. Both kinds of background facts serve very different tasks. Facts_B are the

contextual features that shape the similarity ordering, facts_C are determined by the similarity ordering together with the antecedent in order to settle the truth value of other counterfactuals.

Now in my charitable reading of Brogaard and Salerno, a crucial further premiss of the semantics Brogaard and Salerno build around SL enters the stage:

(Accessibility) Antecedent worlds are either closest or inaccessible.

This *follows* in no way from SL; but it is *reconcilable* with SL. SL does not have to be given up if supplemented by this additional premiss. For H1 to be (non-vacuously) true, the set of accessible worlds from which precisely the antecedent worlds are selected must be confined to worlds where Hoover is the American director of the FBI. This is the decisive background fact_B held fixed by the context. Yet for H2 to be (non-vacuously) true, some accessible world must be a world where Hoover is not American. In order to reconcile both, a change in context is required.

Why can Brogaard and Salerno be confident that the same must go for all other examples where CHS fails? The general scheme of CHS is:

$$p > q, q > r; p > r$$

If we consider *a fixed set of worlds* from which *precisely* the antecedent worlds are selected, we can be confident that this scheme is valid: All p -worlds in the set are q -worlds. All q -worlds in the set are r -worlds. Thus all p -worlds in the set are r -worlds. As a consequence, the set of worlds considered must vary throughout the inference for the scheme to fail. In Brogaard and Salerno's account as I have presented it, there is such a fixed set: the set of accessible worlds. If context varies whenever the set of accessible worlds varies, CHS can only fail due to a change in context.

In sum, I doubt that Cross successfully defeats Brogaard and Salerno. Their account is a *legitimate* way of spelling out SL. What justifies (accessibility)? It provides an explanation why CHS fails that rivals with the standard explanation. Its advantage is that it preserves the principled validity of a syllogistic pattern we feel familiar with. Thus it allows to explain the pull (Hoover) unfolds *as an inference*. Consider

(Hoover) rearranged

(H2) If J. Edgar Hoover had been born a Russian, he would have been a communist. (H1) If he had been a communist, he would have been a traitor. (H3) Therefore, if he had been born a Russian, he would have been a traitor.

Here it seems to me that we feel compelled by the inference but do not as readily accept the second premiss (H1) as in other situations. In my view, this is explained by the inferential pattern ($p > q, q > r; p > r$) becoming *manifest* upon rearranging (Hoover). On behalf of Brogaard and Salerno, I offer the following additional explanation: We settle for the first premiss we encounter (H2) to be non-vacuously true. Yet since CHS *is valid in principle*, we also feel compelled to expect a commitment to its validity whenever the pattern is manifest. Thus we hold the context fixed.² Fortunately, this is irreconcilable with the truth of H1. So we do not have to accept the untenable conclusion. If this view is correct, it shows our commitment to the principled validity of CHS, contrary to the standard explanation.

Yet since Brogaard and Salerno do not provide sufficient evidence of the sort just outlined, I join Cross in denying that they have convicted the usual counterexamples to CHS of a

² Wright's proposal that we tend to hold context fixed when counterfactuals are uttered 'in one breath' does not explain why (Hoover) and (Hoover) rearranged behave differently (Wright 1983: 138; cf. Lowe 1984, 1990, 1995; Wright 1984).

contextual fallacy. My reasons are different from Cross's. There is an alternative way of spelling out SL in which no contextual fallacy is committed. In that reading, CHS comes out invalid. And this alternative is standard in most philosophical debates. The contestable premiss is (accessibility): *Antecedent worlds are either closest or inaccessible*. I focus on Lewis' (1973) treatment. In Lewis' picture, a similarity ordering does not simply distinguish between worlds being closest and worlds being inaccessible. Some worlds are not as close as others but still accessible.³ This changes our picture of (Hoover). Remember: In order for H1 to be non-vacuously true, the closest worlds where Hoover is a communist must be worlds where he is a traitor. To achieve this, the closest worlds where Hoover is a communist must be worlds where Hoover is still the American director of the FBI. Yet in accepting a similarity ordering that makes these worlds the closest worlds where Hoover is a communist, one does not commit oneself to the antecedent of H2 and H3 being inaccessible. To be sure, the worlds where Hoover is Russian and a communist cannot be *as* close as the closest worlds where he is a communist; but they may still be accessible. Nor does the truth of H1 preclude H2 from being true. For contrary to Brogaard and Salerno's view, holding fixed background facts is not inevitably an all-or-nothing matter ('either worlds are such as to preserve that Hoover is American, or they are inaccessible'). For reasons of clarity, I introduce a further notion of background facts, $facts_L$, that replace $facts_B$ in determining the similarity ordering. Nothing seems to preclude a more flexible unified similarity ordering that has the following features: *As long* as this is reconcilable with the antecedent to be assessed, it holds onto the contextually salient features of the actual world, both the $fact_L$ that Hoover is American and the $fact_L$ that Russians are (used to be) communists. This ordering *interacts with the antecedent* of a counterfactual as follows: If one of the $facts_L$ proves irreconcilable with the

³ Or, if we follow Lewis in dropping the limit assumption, there tend to be only such worlds but none that are closest tout court.

antecedent, it is given up. The result is not that the antecedent worlds are inaccessible; rather they are, other things being equal, less close than an antecedent world that allows to hold onto the fact_L in question. The fact_L that Hoover is the American director of the FBI must be given up in order to maintain the antecedent that he is Russian. As a consequence, the closest worlds where he is Russian are less close than the closest worlds where he is a communist. But they are not *ipso facto* inaccessible. If they are not, no prerequisite of H1 being non-vacuously true precludes H2 from being non-vacuously true as well. This is not to give up the idea that certain other background facts, facts_B, play a role in determining the range of accessible worlds. But the accessibility relation is not relevant to assessing (Hoover).

We have seen that, if there is only one set of worlds from which all antecedent worlds are selected to evaluate the scheme $(p > q, q > r; p > r)$, Brogaard and Salerno's results seem compelling. But now there are several sets of antecedent worlds for consideration as selected from the set of accessible worlds. Firstly, there is the set of *accessible* antecedent worlds. And then there is the set of *closest* antecedent worlds. Both will usually come apart. Then CHS may well fail without a shift in context: The closest antecedent worlds form a subset of the accessible antecedent worlds. As a consequence, not all accessible *p*-worlds must be *q*-worlds for the first premiss to come true. And not all accessible *q*-worlds must be *r*-worlds for the second premiss to come true. Thus the closest *p*-worlds may well be *q*-worlds without being *r*-worlds provided the closest *q*-worlds are both *r*-worlds and not-*p*-worlds.

As a consequence, Brogaard and Salerno are right in that there is a legitimate way of spelling out SL according to which the usual counterexamples to CHS involve a contextual fallacy. Yet Cross is right that they do not attain their more ambitious aim: they fail to establish that within the Stalnaker-Lewis approach, CHS comes out valid. For the standard way of cashing out SL has CHS come out invalid. In order to further substantiate their claim, Brogaard and

Salerno would have to show that it is superior in dealing with the data, for instance (Hoover) rearranged.⁴

References

Brogaard, B., Salerno, H. 2008. Counterfactuals and context. *Analysis* 68: 39-46.

Cross, C. 2011. Comparative world similarity and what is held fixed in counterfactuals. *Analysis* 71: 91-96.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.

Lowe, E.J., 1984. Wright versus Lewis on the Transitivity of Counterfactuals. *Analysis* 44: 180-183.

Lowe, E. J. 1990. Conditionals, Context and Transitivity. *Analysis* 50: 80–87.

Lowe, E. J. 1995. The Truth about Counterfactuals. *The Philosophical Quarterly* 45: 41–59.

Wright, C. 1983. Keeping Track of Nozick. *Analysis* 43: 134-140.

Wright, C. 1984. Comment on Lowe. *Analysis* 44: 183-185.

⁴ Cf. Lowe's counterexample to the Lewisian account (Lowe 1984: 182-183). Yet Lowe argues against context-relativity as espoused by Wright and later by Brogaard and Salerno (Wright 1984: 184; Lowe 1984: 181).