



Cathrine Fabricius-Hansen*

Reflections on counterfactuals

<https://doi.org/10.1515/tl-2021-2020>

Received May 9, 2021; accepted May 11, 2021

In his target paper, Wolfgang Klein (WK) takes a fresh look at counterfactual conditionals and simple main-clause counterfactuals, outlining a nonformal compositional analysis that dispenses with notions like presupposition, implicature, and degrees of similarity between possible worlds, i.e. notions that have played a central role in logic-semantic approaches for decades.

WK sums up his analysis as follows:

The present analysis of counterfactuals has two components:

Its linguistic meaning is strictly derived from the meaning of its components; this includes a partition into meaning elements that specify the situation talked about and meaning elements that specify the properties that this situation is said to have.

A counterfactual speaks about a situation in some nonactual world; about that world, we have no knowledge except by what results from the topic-features, including the *if*-clause (and, of course, the context). So, the warrant of our judgment is taken over from reasoning about the actual world and applied analogously to the different topic situations. The likelihood that the counterfactual be true is as good as this common warrant allows us to say.

(WK p. 220)

In other words: The *if*-clause (the protasis) in (1) is adverbial, like e.g. *At five o'clock* in (2), assigning the (past) topic situation *s* of the apodosis to some nonactual world containing a (past) situation with the property [the lights be red],¹ and with (1), the speaker asserts that *s* has the property [it be too late], basing his/her assertion on what the protasis says about *s* and (what she/he knows about) how things usually work in the actual/real world, which tentatively is characterized as the world of the speech act situation.

- (1) If the light had been red, it would have been too late. (= WK (3b))
- (2) At five o'clock, it would have been too late.

¹ Since nothing is said to the contrary, the topic times of the two sub-clauses will be taken to coincide, in accordance with the principle of “topic consistency” (WK, sect. 5.2).

*Corresponding author: Cathrine Fabricius-Hansen, Department of Literature, Area Studies and European Languages, University in Oslo, Oslo, Norway, E-mail: c.f.hansen@ilos.uio.no

The analysis seems quite simple – but, as WK acknowledges (p. 224), it is also rather sketchy. And I must confess that I find it difficult to fully judge its implications and possible advantages over more elaborate theories of counterfactual conditionals (see e.g. Arregui 2021 for a recent nontechnical overview) and simple counterfactuals (see Kasper 1987 for a detailed analysis; c.f. also e.g. Asher and McCreedy 2007; Kasper 1992). In the following sections, I shall briefly mention some of the puzzles I have been struggling with.

1 How do *if* and (subjunctive) mood interact?

It seems difficult to disagree with three fundamentals of the analysis: (i) that the conditional clause syntactic-semantically should be conceived as a specific kind of framing or restricting adverbial in relation to the apodosis (c.f. already Kratzer 1978); (ii) that *if* by itself “marks as UNDECIDED whether the topic situation of that clause has the comment features or not”, i.e. whether the described situation holds or not (p. 217); (iii) that the subjunctive marks that the topic situation “is not in the actual world” (p. 217).

However, it is not clear to me how (ii) and (iii) interact: Does the subjunctive turn the ‘undecidedness’ expressed by *if* into ‘decidedness’ with respect to nonactual worlds? Or does the ‘undecidedness’ at the outset pertain to the actual world alone? In short: How can (ii) be reconciled with what is stated in (48b) below?

- (48) a. Bare *q* indicates that according to the speaker, a situation with the topic-features of *q* has the comment-features of *q*.
 b. When *if p* is added, it contributes to the topic-features of *q*: it adds that the topic situation of *q* is in a world, in which a situation with the topic-features of [*p*] has the comment-features of [*p*].
 c. In each clause, IND adds that the topic situation of that clause is in the actual world; SUB adds that it is not in the actual world.

(WK, p. 217)

2 What is (in) ‘the topic situation’?

The “topic situation” – the situation the sentence (or speaker) “speaks about” – is a central component of WK’s theory. However, is it really needed in addition to the

topic time (and place) as something that must be given by the context if it is not provided by the sentence itself? Isn't a situation – in distinction to a situation type, and with some abstract exceptions – an entity that is defined by its temporal-spatial location in addition to its participants and the relations between them? That is, assuming that the topic time t of the apodosis of (3) is identical to the utterance time, as indicated by *now*, wouldn't (4) suffice as characterization of what (3) expresses according to WK?

- (3) If Ito had eaten the fish, he would now be dead. (WK, p. 214)
- (4) A situation with the property [Ito be dead] holds in some nonactual world in which a situation with the property [Ito have eaten the fish] holds at the time t' , where t' – the topic time of the subordinate clause – is prior to t .

In other words: Does the sentence – or speaker – predicate the comment-property of a given (topic) situation (in some nonactual world), or does it – (s)he – state that there is a situation with that property in some nonactual world at the given (topic) time (and place)? Or should 'the topic situation' be understood in a much broader sense, as a whole "web" of situations or set of circumstances that are given by the context and provide the premises for reasoning about the nonactual world, in addition to general knowledge about the actual world?

3 Against or not against the facts?

WK's proposal nicely accounts for the observation that "a counterfactual protasis need not necessarily indicate the speaker's opinion that the situation is 'against the facts'", and similarly for the apodosis (pp. 7–8). Mostly, however, if nothing is said to the contrary, the listener may safely infer from a counterfactual conditional that in the speaker's opinion, the protasis and the apodosis are contrary to (what the speaker takes to be) the facts; and I am not sure that WK explains why this should be so – or that the explanations quoted below differ essentially from accounts to the effect that counterfactuality is an implicature (see e.g. Ippolito 2003, Arregui 2021 and references therein).²

Under the present analysis, the assertion [of a counterfactual] is confined to some nonactual world. [...] But as a rule, we talk about what is the case in the actual world, and when that is not done, there should be a reason for it; and the most obvious reason is that nothing can be said about the actual world. So, we make the weaker claim – but it may well be justified for the actual world as well.

(WK, p. 222)

² Note that the two statements seem to reflect the speaker's and the listener's view, respectively.

Suppose someone says:

(26) Yesterday, Isa was in Berlin.

That does not preclude that Isa is in Berlin right now; it is just not asserted. [...]

In the same way, something that is true in a nonactual world may also be true in the actual world – but that is not asserted, and the mere fact that it is not asserted may lead us to believe that according to the speaker, it is not true.

(WK, p. 206)

These two passages give rise to further questions: If “something that is true in a nonactual world may also be true in the actual world”, are we then talking about the same situations – or rather different situations with the same descriptive properties (cf. WK, p. 215)? Can one and the same entity – or situation – inhabit different worlds, or must identity across worlds necessarily be understood as a counterpart relation (cf. Lewis 1973)?

How we go about judging the truth of a counterfactual is a central issue in WK’s paper; and the section (6.2) addressing it is a – if I may say so: ‘constructive’ – part of the paper that I find very illuminating and convincing. Of course, it may be asked whether an account to the effect that “the warrant of our judgment is taken over from reasoning about the actual world” (WK, p. 220), differs essentially from the idea that the nonactual world in question must be a “minimal revision” or *ceteris paribus* deviation from what is actually the case – an approach that is dismissed by WK (sect. 2.3.). However, while similarity-based relations in the spirit of Stalnaker (1968) and Lewis (1973) “remain a (more or less) black box” (Arregui 2021: 25), WK illustrates in greater detail the kind of argumentative reasoning underlying our judgments – and why our intuitions about the truth of a counterfactual may remain rather vague. In spirit, then, his proposal belongs to the tradition of premise semantics as developed by Veltman (2005, a.o.) and Kratzer (1981, a.o.) and briefly outlined by Arregui (2021: 25–28).

4 What is (in) a world – and the context?

The paper has fairly little to say about “the context”, except that it contributes to determining the topic situation and other topic entities in cooperation with sentence-internal elements (e.g. p. 210); and the examples used are mostly single sentences without a linguistic context. As a consequence, the role that the discourse context plays in the interpretation does not quite get the attention it deserves. For instance, counterfactuals may be modally subordinated (Asher and McCready 2007, among others); cf. (5).

- (5) That evening, Ito might have been served a poisonous fish. If he had eaten it, he would now be dead.

And while the implicated falseness of a counterfactual protasis, as WK notes, under certain conditions, may be canceled in the apodosis or relativized in following utterances (by the same speaker), the discourse leading up to a counterfactual may strongly restrict its interpretation – or even its applicability (cf. Asher and McCready 2007, among others). Thus, the counterfactual conditional in (3) is unacceptable in (6), where its falseness implicature contradicts the preceding assertion.

- (6) Ito was hungry and immediately ate the fish he was served. #If he had eaten it, he would now be dead.

Notably, in a monologue, indicative assertions set the facts that a counterfactual must be interpreted against, whether the assertions concern the ‘actual’/‘real’ world or are fictional – and consequently relate to a nonactual world according to WK (p. 199).

Interestingly, the paper sometimes, or mostly, talks about what “is (true)” or “holds” in the actual world and sometimes about what according to the speaker “belong(s) to the web of situations that he considers to be the actual world” (WK, p. 206); and footnote 7 mentions that it might be useful “to distinguish between ontologically possible and epistemologically possible worlds” (p. 198).

That indicates that in order to fully understand the semantics of counterfactual constructions, more is needed than on the one hand the distinction between the actual world and nonactual worlds and on the other hand the notion of warrants based on reasoning about the actual world. Asher and McCready, for instance, present an interesting update-semantic account of epistemic modals and counterfactuals that “relativizes the Lewis semantics for counterfactuals to a discourse-dependent set of epistemic possibilities” (Asher and McCready 2007: 125). And at any rate, we need a very rich notion of context, along the lines of Kamp (2020) “Articulated Contexts” and, more generally, his “Mental State Discourse Representational Theory”.

In the communication-theoretic framework described in the last section, language interpretation is a mental process, which the processor, the recipient of linguistic input, must carry out on the basis of the linguistic input he receives and information that is available to him as interpreter, i.e. information that is stored as part of his mental state, and that he is able to retrieve from storage for his current interpretational needs. (Kamp 2020: 79)

5 A final remark

In the sections above, I have mentioned some of the questions and objections that WK’s analysis of counterfactuals may give rise to. Concluding, I should like to

stress that altogether, reading his paper has been a refreshing, thought-provoking, and very useful experience – also because of the puzzles it uncovers but perhaps does not fully solve.

References

- Arregui, Ana. 2021. Counterfactuals. “If kangaroos had no tails ...”. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas Ede Zimmerman (eds.), *The Wiley Blackwell Companion to Semantics*. New York: Wiley Online Library.
- Asher, Nicholas & Eric McCreedy. 2007. *Were, would, might* and a compositional account of counterfactuals. *Journal of Semantics* 24. 93–129.
- Ippolito, Michela. 2003. Presuppositions and implicatures in counterfactuals. *Natural Language Semantics* 11. 145–186.
- Kamp, Hans. 2020. *Trying to see where I am, have been and want to go*. Stuttgart: Ms., University of Stuttgart.
- Kasper, Walter. 1987. *Semantik des Konjunktivs II in Deklarativsätzen des Deutschen*. Tübingen: Niemeyer.
- Kasper, Walter. 1992. Presupposition, composition, and simple subjunctives. *Journal of Semantics* 9. 307–331.
- Kratzer, Angelika. 1978. *Semantik der Rede. Kontexttheorie – Modalwörter – Konditionalsätze*. Königstein: Scriptor.
- Kratzer, Angelika. 1981. Partition and revision: The semantics of counterfactual. *Linguistics and Philosophy* 12. 607–653.
- Lewis, David. 1973. Counterfactuals and comparative possibility. *Journal of Philosophical Logic* 2. 418–446.
- Stalnaker, Robert. 1968. A theory of conditionals. In Nicholas Rescher (ed.), *Studies in Logical Theory*, 98–112. Oxford: Blackwell.
- Veltman, Frank. 2005. Making counterfactual assumptions. *Journal of Semantics* 22. 159–180.