Wolfgang Klein*
Another analysis of counterfactuality: replies

1 Introduction

In the year of 1979, John Robert ("Haj") Ross talked with us, a small discussion group, about the cognitive metaphor which underlies scientific discourse. It is "war", he said, and in his usual style, he bombed us with numerous convincing examples. No one was happy with this idea, and so, I asked him what he thinks the underlying metaphor should be. This was when he was in his Buddhist period, and so, he said: "It should be a joint work of love". Everybody loved that idea, but nobody was convinced. The academic world is not like that, and it will never be. But what may be achieved in scientific discourse is that it does not matter so much *who* is right but *what* is right. When I read the comments on my analysis of counterfactuality, it occurred to me that they might not fully fulfill this ideal but come quite close to it. That is encouraging, and I will try to maintain this spirit in my replies.

I will begin with two general points. The first one is not directly addressed in any of the comments; but in several, it lurks in the background, and when I submitted the first version of the paper, it played an important role in the referees' comments: why don't you use the well-elaborated tools of formal semantics? The reply is simple: with those instruments, I cannot express what I want to say. This is not a mere lack of competence on my part, as became clear to me when in 1992, Arnim von Stechow and I decided to write a paper together on tense and aspect, issues on which both of us were working at that time. I had just finished a first draft of what was to become the book "Time in Language" (Klein 1994), Arnim read the first chapters, and after two or three days, he said: "I understand perfectly well what you are saying, and I think it is correct; but I cannot express it." So, my conclusion was that the problem must somehow be inherent to the instrument that he knew to play so well. Formal languages, sophisticated as they have become over

*Corresponding author: Wolfgang Klein, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, E-mail: Wolfgang.Klein@mpi.nl

the last fifty years, hopelessly lag behind natural languages in their versatility and their expressive potential. If you don't believe that, then express what you just read, understood, but perhaps not believe, in a formal language. At the very best, formal languages capture a few selected aspects of natural language semantics. They hardly say anything about the lexical meaning of expressions, which is the very basis of what a natural language can express. In a word-by-word translation of a Japanese newspaper article, even someone who does not know a single word of Japanese except *sayonara* can understand large parts of the meaning. In a formal semantic representation of this meaning, he or she understands virtually nothing, if the lexemes are retained in Japanese.

Now, the aim of formal semantics is perhaps not to account for the entire meaning of natural languages; it rather addresses some selected features, such as definiteness, tense, counterfactuality, and a few others. What is the added value of a formal analysis over analysis in normal language, enriched by some additional terms, abbreviations, or figures, where appropriate? Take, for example, the Stalnaker-Lewis account of counterfactuals. Lewis (1973b: 424–425) sums it up as follows (here quoted after the target article, in which for ease of typing the symbol >> is used instead of Lewis' original symbol):

(7WK)[1]   A counterfactual conditional p >> q is true at the actual world iff (a) there
            are no (accessible) worlds at which p is true, or (b) if some (accessible)
            world at which p and q are true is closer to the actual world than any
            world at which p is true and q is not.

The definition is in plain English prose, enriched by the symbol >> and the word "accessible" in an unusual meaning. In Arregui (2008: 18), the Stalnaker-Lewis analysis is (roughly) summed up as follows:

(1AA)   α *would* β is true in (a world) w_0
         iff the α-worlds most similar to w_0 are also β-worlds
         iff $\{w: S(w_0)(\alpha)(w)\} \subseteq \{w: \beta(w)\}$"

The first two lines give a version in natural language, enriched by symbols, which, of course, have to be explained in natural language; otherwise, the definition would not be understandable. It may be arguable whether this version indeed captures the Stalnaker-Lewis idea. But that is not my point here; it is rather what is gained by the formal definition in the third line over the prose version in the second line. This depends on how the additional symbol S is interpreted. Arregui adds the following explanation: "[…], S stands for a contextually supplied

---

similarity relation." Strictly speaking, this goes beyond bare "most similar", insofar as it adds context-dependence to similarity. But as far as I can tell, the S is in no way more informative or more precise than a prose definition like "contextually supplied similarity relation". **We understand the formal definition only to the extent, to which we understand its informal explanation**. And that seems to be characteristic to me for any attempt to analyze the semantics of natural languages in terms of formal languages. So, what is the added value of the formal definition? I don't see any, and until it is clearly demonstrated that this impression is false, I rather stick to usual natural language, enriched by some additional words, abbreviations, or figures, as a means of meaning analysis. In fact, I am not aware of other disciplines except mathematics or logic which proceed otherwise. They often use mathematical formulae when it comes to quantitative givens. But they do not try to express their insights and results in a formal language. Never.

The preceding remarks should not be read as an attack on formal semantics whose intellectual standards I have always admired; they are rather the expression of certain helplessness when I should say how much it captures the meaning of words and sentences in natural language. And I am quite willing to change my mind when convincing arguments are put forward.

The second general comment is very different. For a satisfactory analysis of a linguistic phenomenon, it should be clear what has to be accounted for; this includes the relevant phenomena as well as the problems which their analysis has raised in the past and still raises. The ground for the analysis must be properly laid. This requires a broad inspection of the field, here of counterfactuality; Sections 2–4 in the target paper are devoted to this aim. What is said there is the touchstone against which any analysis has to be checked. I have tried to put together what I thought to be relevant for the purpose; surely, it does not cover everything that deserves discussion. In some cases, this was clear in advance; in others, it became clear to me when I read the comments. I particularly regret the almost complete omission of two research domains in which counterfactuality plays an important role; these are counterfactual reasoning in psychology, which was not dealt with at all in the target paper, and fictional literature, which is only briefly touched upon. A final deplorable omission is texts whose sentences don't describe a "reality" against which they can be matched, such as legal texts like laws, contracts, or court opinions; in a way, they are neither factual nor counterfactual, and typically, they use the indicative. As I have argued elsewhere (Klein 2000b), they often lead to a more adequate view on well-known and intensively studied phenomena, whose analysis often lives on a scarce diet of examples and observations. I encourage the reader to go through some of the numerous *if*-phrases in this or in any other text.

The comments raise so many issues that it is impossible to treat them all; so, I had to be selective. All comments by and large agree with what is said in the

"setting the ground" part; this is surprising because some of the points made there are serious challenges for earlier analyses. Four comments – Hinterwimmer, Grønn, and to a lesser extent, Fabricius-Hansen and Zakkou – see bigger or smaller problems with my analysis. The other four comments primarily take up some points and elaborate on them; this is very much in the spirit of what was said in the last sentence of the target paper – that my proposals should be seen as something that "opens new vistas to an understanding of how counterfactuals work". My replies to these four are much shorter than those to the first four. This is not because I would find them less interesting, but because there is less to say if people are largely agreed. There are some repetitions in the replies; I thought this makes them better to read than when stuffed with many disruptive cross-references and thus producing a sort of unreadable hypertext.

2 Stefan Hinterwimmer: no conditionalization without restriction

2.1 Truth conditions

Stefan Hinterwimmer argues that my analysis does not yield the correct truth conditions. That is a substantial objection. In the first version of his argument, the subtleties of the topic-comment structure are omitted for clarity of exposition. His example is (1SH), and he writes that my analysis leads to the truth conditions in (2SH):

(1SH) If Mary was at Paul's party, she had a lot of fun.

(2SH) Some situation containing Mary that is (a) temporally located at the time of Paul's party and (b) part of the actual world is a situation of Mary having a lot of fun. Possibly, it is also a situation of Mary attending Paul's party.

I agree that (2SH) does not capture the truth conditions of (1SH). But it does not correspond to my analysis, either. I quote from the target paper:

(48WK) 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]*. [italics added, WK]
   b. 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.

Note that the *if*-clause is not "the topic" of the main clause; it is a syntactically optional addition to the topic-features of the main clause. Note, furthermore, that in *if* p, the IND-marking only says that its topic situation belongs to the actual world; it does not say that this topic situation has the comment-features of p and thus all properties of p. Whether that is the case, is open, but the assertion is restricted to cases in which it is. A speaker may well say that the situation talked about belongs to the actual world or to some nonactual world. But that does not deprive him of the right to "iffy" that, i.e., to indicate that to his mind, it is open whether that situation has the comment features or not. In other words, someone can talk about a world with a particular situation without being committed to the opinion that this world indeed includes this situation with its comment features. This is just as *It is unclear whether Mary was at Paul's party* describes a situation in the actual world and indicates that according to the speaker, it is open whether the actual world contains such a situation.

So, leaving the topic-comment-features aside for a moment, (48WK) yields for (1SH):

(1)   In a world which contains a situation with properties [Mary PAST ACTUAL be at Paul's party], Mary had a lot of fun.

It is not asserted that in that world – here the actual world – Mary was at Paul's party; and about a world, in which she was not, nothing is asserted at all. This is just like *As a child, I was often unhappy* does not assert anything about how I am right now.

Before including the topic-comment aspect, I would like to establish what was just said with a naïve test of the reader's semantic intuitions which he or she connects with the *if*-clause of (1SH) alone when uttered on some occasion:

(2)   If Mary was at Paul's party

Here are my own intuitions. First, the expression is neither true nor false; it is an adverbial phrase, and adverbial phrases have no truth value. Second, the speaker describes a situation with the descriptive properties [Mary PAST ACTUAL be at Paul's party]. Third, it is not asserted that the situation talked about has these properties, quite to the opposite, it is explicitly marked that this is open: it could be that Mary was at Paul's party; it could be that she was not at Paul's party. This would be quite different without *if*; thus, it is this word that marks the undecidedness (in contrast to, for example, *when Mary was at Paul's party* or *because Mary was at Paul's party*). The speaker may do this because he does not know whether she was there or not; or because she knows it but does not want to reveal that knowledge; in either case, it is marked as open. So, with (2), the speaker describes a world as a world that contains a situation with properties [Mary PAST ACTUAL be

at Paul's party]; he does not say that this world (for example the actual world) is such a world. And this immediately leads to the semantic analysis (3):

(3)   The adverbial phrase *if xyz* means: In a world which contains a situation with the properties [xyz]

This is not a big theory. It just describes the semantic intuitions which I connect with an *if*-clause. Now, I am not a native speaker, but I suppose that most native speakers share these semantic intuitions. What do you, a native speaker or not, think? We could try to translate (3) into a formal language, but I do not see what the added value would be.

Let us now turn to the information-structural differentiation. It does not make sense to ask whether the description of a situation is true unless it is clear which situation is thus described. How can we know what the situation talked about is? This may become clear from the context (*I don't know whether Mary was at Paul's party. But if she was, …*). Thus (2) is about a situation with properties [Mary PAST ACTUAL], and the properties which that situation may have or not [be at Paul's party]; I call the former "topic features" and the latter "comment features", respectively. The topic features of (1) help to identify the topic situation of the *if*-clause, and the latter describe which other properties the topic situation may have or not have. In an *if*-clause, the speaker refrains from taking a stand on that. This differentiation is no longer a semantic intuition but an assumption that may not be shared by others; to me, it sounds indispensable.

Consider now the second part of (1SH) alone, also uttered on a particular occasion:

(4)   Mary had a lot of fun.

Unlike an adverbial clause, it has a truth value; it is not marked as undecided but as true according to the speaker. Is it really true? As said above, this question is meaningless, unless it is clear which situation it is about (or do you know whether it is true?). Under the simplifying assumption above, the topic features of (4) are [Mary PAST ACTUAL], and the comment features are [be at Paul's party]. But there are very many situations with those features; therefore, we need more specific information. Again, this information may be – and normally does – come from the context. But it may also be partly provided by an initial adverbial phrase, as in (5):

(5)   a.   In Pontefract, Mary had a lot of fun.
              b.   As soon as Ira stopped singing, Mary had a lot of fun.
              c.   If Mary was at Paul's party, she had a lot of fun.

In all of these cases, the initial adverbial phrase says more about the topic situation of the main clause than just [Mary/she PAST ACTUAL]; in (5a), it specifies its "topic place", which is not explicitly mentioned in the main clause at all. In (5b), it

specifies its "topic time" above and beyond [PAST]. In the entire communicative context, that may be enough to identify the topic situation. In (5c), the additional information is also provided by an adverbial clause. But it is left open whether its topic situation with [Mary PAST ACTUAL] also has the comment features [be at Paul's party]. If yes, it contributes to the topic features of the main clause above and beyond [Mary PAST ACTUAL]. If not – that is, if Mary was not at Paul's party –, not. The meaning of (5c) alone is, therefore, "in a world which contains a situation with features [Mary PAST ACTUAL be at Paul's party]". The assertion is thus confined to a topic situation with these features and the features [Mary PAST ACTUAL] (plus the contribution of the context). That topic situation is said to have the properties [she have a lot of fun].

If we pack all of this together, we get a refined version of (3), which now includes the role of topic features and comment features:

(6)   In a world in which a situation with the topic features [Mary PAST ACTUAL] has the comment features [be at Paul's party], a situation with the topic features [Mary PAST ACTUAL] has the comment features [have a lot of fun].

This is just as in the case of (5a), except that the contribution of the adverbial phrase with topic function is different:

(7)   In Pontefract, a situation with the topic features [Mary PAST ACTUAL] has the comment features [have a lot of fun].

At the risk of being too repetitive, let me rephrase this in a slightly different way. An assertion states that according to the speaker, some situation has certain properties. Which situation, which properties? In *If Mary was at Paul's party, she had a lot of fun*, this is as follows (I omit the topic-comment distinction within the adverbial phrase):

(8)   A situation with the features
       – Mary is the subject
       – it is in the past
       – it is in the actual world
       – it belongs to a world which includes a situation with the features [Mary PAST ACTUAL be at Paul's party]
       has the properties [have a lot of fun]

In *If Mary was at Paul's party, did she have a lot of fun?*, the analysis is the same except that it is not asserted that the situation talked about has these properties but it is asked whether it has them.

In Stefan Hinterwimmer's full argument, he describes my analysis as follows:

(3SH)   Some situation s_1 containing Mary the time of which is the time at which Paul's party took place is in a world w in which a situation s_2 containing Mary the time of which is the time at which Paul's party took place is a situation of Mary attending Paul's party and s_1 is a situation of Mary having a lot of fun, where w is the actual world.

I do not think that this is the same as (6), and I also think that, unlike (6), it does not capture the meaning of *If Mary was at Paul's party, she had a lot of fun*. It misses the requirement "in a world, in which a situation with the topic-features of [p] *has the comment-features of [p]*".

Hinterwimmer's main interest is in indicative conditionals, but he also briefly extends this argument to counterfactual conditionals in (5SH). Since his objection is essentially the same, I do not take it up here.

2.2 Other points

2.2.1 Existence

In connection with (48WK), cited above, I said in the target paper about the undecideds of the *if*-clause: "Whether that is the case, is open, but the assertion is restricted to cases in which it is. This is just as in *Blue kangaroos are easy to detect*, the assertion is restricted to blue kangaroos, but it is not asserted that there are any blue kangaroos." Hinterwimmer comments on this:

> First, while the sentence may not directly assert the existence of blue kangaroos, their existence in the actual world is presupposed according to my intuitions when the speaker utters the sentence in the indicative. Consequently, a speaker who does not believe in the existence of blue kangaroos or is at least undecided on the matter has to use the subjunctive version *Blue kangaroos would be easy to detect*.

I agree that, if we consider the indicative assertion to be true, we are inclined to assume that there are blue kangaroos in the actual world. But why is this so? Nothing in the NP *blue kangaroos* itself suggests, or even requires, that they exist. There is no contradiction in the assertion *There are no blue kangaroos*. Here, *blue kangaroos* don't have topic status, and this may be the reason for the intuition. But that cannot be the full explanation, cf. *Blue Kangaroos are often mentioned in Leichhardt's travel account* or *Blue kangaroos I have never seen*. In neither case is it necessary that blue kangaroos really exist. Note, furthermore, that in the subjunctive version, *Blue kangaroos would be easy to detect*, we also are inclined to assume that they exist – in the nonactual world talked about. So, I share

Hinterwimmer's intuition. I do not believe, however, that it reflects a presupposition; it seems rather the result of several interacting factors that lead to that intuition.

2.2.2 *If*-clauses as comment features

Another point which Hinterwimmer addresses concerns "the existence of cases where the *if*-clause is the focus, not the topic of a sentence". That point is briefly discussed in connection with exx. (40WK) and (41WK). The semantic analysis of *if*-clauses, as stated in (3) above, does not say anything about the status which they have in the information structure of the sentence: an *if*-clause can contribute to the topic features of the main clause, it can also contribute to its comment features (not the focus – this notion is not used at all![2]). It may well be that the topic situation of the main clause is only specified by the context and perhaps other elements within it; in that case, the *if*-clause restricts the comment part. This is just like with other adverbial phrases, cf. *In München, I often eat Weisswürste* versus *I often eat Weisswürste in München*. Intuitively, the difference is often hard to grasp; but that is a general problem of information structure and has nothing to do with *if*-clauses in particular. Note, furthermore, that initial position is not the only marker of topic status (in the sense, in which I use this notion). In *I only eat Weisswürste in München*, it is crucial whether the final fall – a typical assertion marker – is before *in München* or within *München*. In the former case, *in München* is perceived as de-stressed (and *Weisswürste* as stressed) in the latter case, *in München* is perceived as stressed.

2.2.3 Ebert et al. (2014)

The target paper hardly discusses other analyses; that would have required a book. As to Ebert et al. (2014) in particular, its focus is on *if*-clauses, and counterfactuality comes in only marginally. This is not the place to discuss their analysis in detail, it is by far too complex; so, I will only address a few points which seem problematic to me. Hinterwimmer sums up their analysis as follows:

> [Our analysis] builds on Schlenker's (2004) assumption that *if*-clauses denote maximal pluralities of possible worlds where the proposition denoted by the *if*-clause is true and which are otherwise as close to the world of evaluation as possible. We argue in that paper that in indicative as well as subjunctive conditionals with the pro-form *then* in the main clause, the

---

2 As an aside, I find the best-elaborated analysis of "focus", alternative semantics, fundamentally problematic because in a sentence like *Mary was at Paul's party*, **all** elements (except perhaps *at*) have alternatives, not just those which we feel somehow to be "focused".

maximal plurality of possible worlds denoted by the respective *if*-clause is the aboutness topic (in the sense of Reinhart 1981) of the sentence.[3]

For the antecedent, this is spelled out as follows (Ebert et al. 2014: 355 (footnote omitted)):

(6CE)   The antecedent of an NC [normal conditional, WK] is interpreted as a (semantically plural) definite description referring to the maximal plurality of possible worlds *W* that
1. is compatible with everything the speaker knows in the world of evaluation w_0, and that
2. makes the proposition expressed by the antecedent true.

I have four questions and three remarks. Here are the questions:

(i)   In Section 2.1, I sketched my semantic intuitions regarding *If Mary was at Paul's party*. Does their analysis, as cited in the summary above, capture those intuitions? Or are my intuitions false?

(ii) How does it eventually differ from what I say in simple words in (3) or (6), except for the maximal similarity requirement with all its problems (see end of Hinterwimmer's comments)?

(iii) Is the pro-form *then* in any way crucial for the function of *if*-clauses? In Ebert et al (2014), it is primarily used to separate "normal conditionals" from "biscuit conditionals" (briefly mentioned in the target paper, exx. (32WK) and (33WK).

(iv) How does it account for the intuitions that are traditionally associated with counterfactual and other conditionals, such as restrictedness, consequence, reality status, probability, remoteness, (non)factuality (see Section 2.2 of the target paper), and for the problems that were discussed in Section 3 there?

The three remarks are:

(v)   The idea that *if*-clauses are the "topic" or the "theme" of a sentence is very old; in fact, it can be traced back to Georg von der Gabelentz to whom we owe the idea of "aboutness". But I do not think that an assertion like *If the weather is fine, the party is in the garden.* asserts something about what the weather is like; it rather asserts something about the party, more precisely, where its place is. The *if*-clause somehow restricts that assertion.[4]

---

3 Schlenker's assumption and Lewis' observation, on which it is based, are discussed with respect to "definiteness" in my reply to Grønn, see Section 3 below.
4 In 1980, I happened to have many discussions with Tanya Reinhart about her aboutness topics. And while I got some intuitive grip, I never understood what precisely they are. Isn't *Mary had fun at Paul's party* about the party, as well?

(vi)   The definition of the antecedent requires that the maximal plurality of possible worlds is compatible with what the speaker knows in the evaluation world. I do not think that the meaning of an *if*-clause (or any other linguistic expression) should be made dependent on what a particular speaker knows; the knowledge of speakers varies enormously. And does the interpretation ("is interpreted") by the interlocutors really vary with the knowledge of the speaker?

(vii) The third remark concerns the difference between "normal conditionals" and "biscuit conditionals". I think it is more complex than assumed in Ebert et al. (2014). In German *Wenn ich mich recht erinnere, dann sind noch Biscuits im Schränkchen*, there is a (optional) pro-form *dann*, but the availability of the biscuits does not depend on the speaker's recollections. (I suppose that this also holds for English *If I remember correctly, then there are biscuits in the sideboard*, but my intuitions here are less certain.) Or take *If the Madrid players were wearing red shirts, then Madrid won the final* is perfectly fine, although the red shirts is hardly a condition for their victory. But that conditional makes perfect sense when if the speaker does not know whether Madrid won but knows that the team with red shirts won. Note, furthermore, that biscuit conditionals normally have no counterfactual counterpart: *If you were hungry, there would be biscuits in the sideboard* or *If I remembered correctly, there would be biscuits in the sideboard* sound odd. Under my analysis, biscuit conditionals are a special case of an assertion restriction, and their different behavior in some respects has nothing to do with the semantics of the *if*-phrase, which is the same throughout, but with the particular type of restriction on the assertion. In *If I remember correctly, …,* there is a restriction to those worlds which correspond to my recollections.

This ends my reply to Hinterwimmer's comments. I guess we are not entirely agreed. But I found the comments really interesting and challenging.

3 Atle Grønn: justifying tense and mood morphology in counterfactuals

I was touched to read that both of us set out to work on tense with our joint friend Arnim von Stechow. As said in Section 1, this attempt led in my case to an increasing skepticism as regards the value of formal approaches to natural language semantics; in Atle Grønn's case, the cooperation brought forth concrete results, and a good part of his comments on the target article is based on them. My

reply begins, however, with a different point. I do not agree at all with his opinion that puts my analysis on a par with the Lewis-Stalnaker view and the Schlenker and the Ebert et al. analysis:

> A reasonable paraphrase of Klein's analysis of counterfactuals of the form *if P, Q* seems to be "the topic situation *P* has the comment feature *Q*".

> A certain parallel – allegedly a shared nonmonotonic behavior – between definite descriptions in the individual domain and conditionals in the world domain, was noted in passing already by Lewis (1973). Schlenker (2004) and Ebert et al. (2014) worked out a formal analysis in the spirit of Lewis' observation. Klein's target paper can also be seen as an implementation of this idea.

While the first sentence is a strong simplification of what I believe, its spirit is correct. But I see no parallelity with respect to "definite descriptions in the individual domain and conditionals in the world domain", as considered by Lewis and elaborated by Schlenker and by Ebert et al. The underlying idea of these and many other approaches is (with considerable variation in detail) what Lewis wrote in (1973b:424–425), repeated here for ease of reading from the target paper):

> (7DL) A counterfactual conditional p >> q is true at the actual world iff (a) there are no (accessible) worlds at which p is true or (b) if some (accessible) world at which p and q are true is closer to the actual world than any world at which p is true and q is not.

In Sections 2.3.2–2.3.5 of the target paper, I raised four serious problems against this view. I am not aware of any approach along this line that could solve these problems; in fact, I am not aware of any attempt to solve them. A potential exception in one respect is Arregui (2008, 2009) who, like me, casts doubt on the idea of overall similarity between worlds. Her motivation is somewhat different, but the net result is in both cases that the notion of "global similarity" is not very useful here. Her solution is to operate with a "local similarity" relation, in which the features that are relevant for the comparison are provided by a matching past situation of the actual world. I do not believe that this can solve the problem, first, because I do not believe that pastness plays an important role in counterfactuality (this will be discussed below), and second, because in many very natural cases, there is no situation that could serve as a base of comparison, cf:

(9)   If I went to Isa's wedding next week, she would be surprised.

In the actual world, there is no situation in the past, at which I went to Isa's wedding next week.

Let us turn now briefly to Schlenker's idea of definite *if*-clauses (as to Ebert et al., see also my reply to Stefan Hinterwimmer's comment). His core idea picks up an occasional observation in Lewis (1973a). It is stated as follows (Schlenker 2004, 417f.):

> For many years, Lewis's observation went largely unnoticed. The present paper is an attempt to revive it, and to take it quite literally: we suggest that *if*-clauses are simply definite descriptions. […]

> Following Lewis's intuition, we will suggest that *if* should be seen as the form taken by the word *the* when it is applied to a description of possible worlds.

Lewis's idea originates amidst his attempts to get a grip on the concept of maximal similarity (see Lewis 1973a, Section 5.3, in particular, p. 115). He sees a certain parallelity between *if*-clauses and noun phrases like *the xyz*. The latter is what he calls "contextually definite descriptions". Among a contextually given set of entities that fall under the description *xyz*, some catch someone's attention more than others. The most salient is definite, it is **the** *xyz*. That may also be applied to worlds, and then, it becomes another turn on the theme of what is maximally similar. The argument which Lewis gives is "alternating sequences" like *If Otto had come, it would have been a lively party; but if both Otto and Anna had come it would have been a dreary party; but if Waldo had come as well, it would have been lively; but … .* (ibid, p.10), which according to him are a mark of counterfactual conditionals as well as of *the xyz* noun-phrases.[5]

    This parallelity, if it indeed exists, seems to be a weak argument for definiteness. But this is not the place to examine neither the original idea nor Schlenker's implementation of it. I do not see how an analysis of counterfactuality along this line can help to solve the problems discussed in Sections 2.3.2–2.3.5 of the target paper or the various semantic intuitions which are traditionally associated with (counterfactual) conditionals. Now, this was not Schlenker's intention; it is rather an attempt to capture certain properties of *if*-clauses: they are definite descriptions of possible worlds. So, I leave aside what the proposal means for an analysis of counterfactuality and concentrate on that point. I see four problems:

(a) Traditionally, the concept of definiteness relates to the referents of noun phrases, rather than to the referents of adverbial phrases, such as *if*-phrases. In Schlenker's analysis, these latter referents are worlds, with *if* (or its analoga) as the counterpart of *the* (or its analoga). Does this mean that we have to

---

5 Such alternating sequences are no hallmark of counterfactual conditionals, cf. *If a house is very old, it is valuable; if a very old house has a rotten roof, is not valuable; if a very old house has a rotten roof, but also a huge garden, it is* etc. etc. It is even not a hallmark of conditionals, cf. *A very old house is valuable; an old house with a rotten roof is not valuable: an old house with a rotten roof and a huge garden is valuable,* etc. etc. In general, the much-discussed problems with antecedent strengthening seem to me a fruit of the "original sin" – the idea to base the analysis of natural language conditionals in one way or another on material implication, where the notion of antecedent strengthening indeed makes sense.

deal with definite and indefinite adverbial phrases, and if so, what is the indefinite counterpart to an *if*-phrase?

(b) There is no generally accepted analysis of definite noun phrases, not even for a single language like English. It is somewhat embarrassing for us linguists that there is no agreement on what the most frequent word in the best-studied language of the world, name *the*, exactly means. There are two major notions which, in varying forms, guided the discussion over the last 100 years: these are "uniqueness", exemplified, for example, by Bertrand Russell, and "familiarity", exemplified, for example, by Hans Kamp or Irene Heim. One might also add David Lewis' "salience", mentioned above; but for that, it might hold what I once heard him say in one of his wonderful talks (Cambridge 1973): "we gained safety by saying close to nothing" (relating to "cases", about which adverbs quantify). Intuitively, we understand that something catches our attention in some context; but that seems to be a very fuzzy concept. (As a little biographical note, I once pointed out to David Lewis, during a walk in Konstanz in 1979, that definite noun phrases in German and English show a number of differences; he thought about that for about a minute and then stated that these are probably idiosyncrasies).

I think that singularity and familiarity, in one way or another, are indeed at the core of *the*-noun-phrases and their counterparts in other languages, although there are a number of (nonidiosyncratic) cases in which they do not work. One example which I like particularly well is "Goldbach's conjecture" (for a detailed discussion of these and other cases, see Klein 2000b):

(10)   Every even integer greater than 2 is the sum of two prime numbers.

The referent of *the sum of two prime numbers* is neither unique nor familiar, let alone salient.

   How do familiarity and singularity fare in the case of *if*-clauses? If I go through the numerous examples in the target paper or in the comments, I do not get the impression that these properties – the **defining** properties of NP-definiteness – play any substantial role. But I may be wrong, of course.

(c) Here is another appeal to semantic intuitions: in which sense are the *if*-clauses in the following examples or the worlds about which they speak definite?

(11)   a. If you ever were in Aix, you know what I mean.
         b. If you had ever been in Aix, you would know what I mean.
         c. If I had never been in Aix, I would never have met Isa.

My intuition, admittedly vague, is rather that in none of them, the world(s) referred to or the situation(s) referred to are definite. In fact, I even don't know on which ground to argue here.

My remarks so far concern Atle Grønn's comments on how the analysis in the target paper is related to some other approaches. Let me turn now to his own proposal, the main points of which are "(1) the fake past tense morphology, (2) the temporal control from the main modal into the *if*-clause and (3) the duality of the counterfactual necessity modal *would* and the counterfactual possibility modal *could*". And his view on these three points is as follows (one footnote omitted):

> The position I will defend here is the following: the main tense of a counterfactual conditional with simple past morphology – the *would-could* conditional in English – is a semantic present, while constructions with two layers of past morphology – the *would-could have* conditionals – are (typically) shifted to a past time by the matrix time shifter *have*.

Unlike Atle Grønn, I do not believe that past tense morphology plays any substantial role in the analysis of counterfactuality. In the target paper, this point is only briefly discussed (Section 2.1). But since "pastness" is an ingredient, even a core ingredient, in many recent approaches, I will elaborate a bit on this issue. In a nutshell: I think one should not base the analysis of an ambiguous form on that one of its meanings which it does not have in counterfactual constructions.

In their careful survey of counterfactuality marking in 43 languages, Van linden and Verstraete (2008: 1865) found that in about one-third of them, the expression of counterfactuality is somehow connected to pastness marking. They sum up their findings as follows:

> Contrary to popular belief, neither past tense nor imperfective aspect is a universal feature in the combinations of markers used to signal counterfactuality: the only type of element that is found in every combination is a modal element marking some type of potentiality, which can be combined (i) with past tense markers, (ii) with a combination of past tense and aspectual (perfect or perfective) markers, or (iii) just with aspectual markers.

Let me second their finding by a brief look at a few languages which are typically treated in studies of counterfactuality. I begin with Latin, the language which not only shaped much of our thinking about grammar up to and including modern linguistics, but which also was for many of us, including me, the first encounter with grammatical notions. In Latin, there is a clear difference between forms that express "pastness" and forms that express "irrealis":

(12)		Indicative ("realis")	Subjunctive ("irrealis")
	a.	Si taces, philosophus manes	si taceas, philosophus maneas
		if you keep silent, you remain a	if you kept silent, you would
		philosopher	remain a philosopher

   b.   Si tacueras, philosophus   Si tacuisses, philosophus
         manseras   mansisses
         if you had kept silent, you had   If you had kept silent, you would
         remained a philosopher   have remained a philosopher

(This picture is simplified; thus, the pluperfect *tacueras* is the indicative counterpart to *tacuisses*, whereas the Latin imperfectum, which may be considered the closest counterpart of the English simple past, is *tacebas*; either way, there is a clear difference).

The Romance languages simplified this system in various ways. Old French, for example, maintained the difference for a long time. But in the XIIth century, the protasis form was more and more replaced by a form which corresponds to the "imparfait", whereas the apodosis form was replaced by what is now called "conditionnel", a new form that is akin to the French future: *si tu te taisais, tu resterais un philosophe*. So, pastness could play a role for the protasis, whereas the apodosis form resembles more a future tense. None of these forms has a tense meaning.

In German, the protasis verb, as well as the apodosis verb in counterfactual conditionals, is regularly different from past tense forms. The most common way is this:

(13) a.   Wenn du kommen würdest, würde ich mich freuen.
          If you came, it would please me.
      b.   Wenn du gekommen wär(e)st, hätte ich mich gefreut.
          If you had come, it would have pleased me.

(13a) is used when the "event", the addressee's coming, includes or follows the speech time (in this example probably the latter); (13b) is used, if it precedes the speech time.[6] There is no past tense meaning whatsoever in the finite verb. There is, however, the possibility to use the combination of a finite auxiliary and past participle. In that case, the time talked about is not the "time of the event" itself but its "post-time" – the "*have come* time"; thus, the "event time" must be earlier (this, and also the differences to English, are discussed in some detail in Klein 2000a and – much better – in Klein 2010).

For a limited number of verbs, it is also possible to use the "Konjunktiv II", a form that is historically derived from an old optative (like English *would*); it resembles but also clearly differs from the Präteritum (the historical counterpart to the English simple past). This choice is only available for the protasis; so, instead of

---

6 As I have argued elsewhere, I do not believe that tense expresses a relation between the speech time and the event time; but that does not matter for the present point.

Wenn du kommen würdest, it is possible to say *wenn du käm(e)st*. Usage varies here; but whenever the *würde*–construction is not used, the simple form must be different from the Präteritum. Since the Konjunktiv II and the Präterium are the same when the latter is formed by the suffix *-te* (the counterpart to English *-ed*), the former is avoided in counterfactuals. People would look very baffled, when you said *Wenn ich kochte, schmeckte es dir nicht*. I do not know why this is so, but it seems plausible that it is done in order to avoid ambiguities. Whatever the reason might be, German does not use past tense forms for counterfactuality.

Given all of this, it would never have occurred to me that the protasis form of English counterfactuals is a "past tense". What looks like past tense, is just one reading of an ambiguous form, an ambiguity that like many others is due to the morphological decay (blessed be it!) of English. This is also the position of the Cambridge Grammar of English, the ultimate reference work nowadays (Huddleston and Pullum 2004). They use the term "preterite" for forms like *(he) came, called, was, had,* etc., and say (p. 85):

The preterite has three distinct uses illustrated in [29]:

[29]	i.	She always took her dog with her.	[past time]
	ii.	If he took the later plane tonight, he wouldn't have to rush.	[modal remoteness]
	iii.	Kim said I took things too seriously.	[backshift]

The last use is not relevant here. Modal remoteness is their term for counterfactuality or irrealis; past time is temporal, in the most typical case, it means "event precedes the speech time". So, we simply deal with ambiguity, without any relevance for the analysis of counterfactuality or the analysis of tense. Note that the English present perfect cannot be used at all in a counterfactual protasis, although it also places the event before the speech time and in that respect corresponds to the simple past.

In my analysis, I follow the simple idea that there is an ambiguity as so many others. I do not use the notion of "modal remoteness". Instead, I believe that these verb forms, when used in the protasis (and only there!), as well as the form *would* + infinitive indicate that the situation talked about does not belong to the actual world. So, my question to Atle is this: what is the gain, if we somehow manage to link the alleged pastness to the irrealis meaning, which these forms have in the protasis of a counterfactual conditional?

There is, of course, a kind of "pastness" which comes in with the forms such as *would have left* or *would have been dead*. But this has nothing to do with the past tense, in the sense of a deictic relation to past, present, or future. It results from the fact that *have* + verb does not relate to a temporal interval with properties described by the verb, but to an interval after such an interval, to its "post-time".

I mentioned that already for German, it also applies analogously to English. Consider (14):

(14)   a. Tomorrow at ten, Ira will leave   vs. Tomorrow at ten, Ira will have left
        b. Yesterday at ten, Ira seemed to   vs. Yesterday at ten, Ira seemed to
           leave   have left.

In (14a), the time talked about is after the speech time, as indicated by the adverbial phrase and the finite element *will*. In the first variant, this is the time of Ira's leaving. In the second variant, it is the time after his leaving; the leaving itself may be in the future, it may be around the speech time, and it may even precede it (as in *I am not sure whether Ira has left already; but tomorrow at ten, he will definitely have left*). In (14b), the time talked about precedes the speech time; in the first variant, the leaving time is the time talked about, and in the second variant, it is even "more in the past".

Let me conclude my reply to Atle Grønn's comments with a brief look at another aspect of *would* + verb. In my impression, the research tradition on counterfactuals (and many other linguistic phenomena) has an unfortunate predilection for a small menu of examples that are discussed time and again; so, here is an authentic one. On August 24, 1935, congressman Rich stated (US Congressional Record 1934/35: 14,664):

> If I would do that, I would do more detriment to America than has ever been done in the history of this Nation.

An average speaker of English, when asked to explain what the congressman put into words here would probably say something like: "Well, he said about an imaginary action of his in the future that it would be very, very bad, and therefore, he will vote nay." And he might add: "Actually, he should have said 'If I did that, …', because that is better English." But he would not say that the two versions mean something different – it is just a stylistic difference. In fact, a Google search of the sentence *If I would do that* yields more than 1.2 million citations (May 8, 2021). Should *would* + infinitive in the protasis also be treated as a (fake) past like *did*? At the speech time, the imaginary action of congressman Rich was clearly in the future, and so was the imaginary detriment to the great Nation. Mr. Rich voted "nay", but he was outvoted, and maybe that indeed led to the Dust Bowl, which took off in 1935.[7]

---

7 The basic difference between *would* and *could* in the main clauses – a point also addressed by Grønn – is to my mind this: *would* says "situation does not belong to the actual world", whereas *could* says "situation belongs to a possible world" – both in contrast to the default case which is that the situation belongs to the actual world.

4 Cathrine Fabricius-Hansen: reflections on counterfactuals

Fabricius-Hansen begins her comments with a general remark:

> 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 McCready 2007; Kasper 1992).

To which I have two replies. First, I am almost religiously convinced that an analysis should be as simple as possible; in particular, it should not make highly specific assumptions about the syntactic structure or the semantic properties of an expression, unless they are absolutely necessary. The "technical" elaborations I have seen in connection with counterfactual expressions often do that. I did not want to develop and to sell my own little syntax as a specimen of one of the numerous frameworks that are on the market. My intention was rather to base the analysis as much as possible on standard views which you find in any good descriptive grammar. Second, my touchstone for the value of other accounts is what is listed in Sections 2.3 and 3 of the target paper. I did not find any which can handle what is said there. This holds, for example, for all those which are in the "material/strict/variably strict implication" tradition, or for all those which operate with "maximal similarity", or for all those which are somehow based on "pastness" (see the reply to Grønn above). But maybe I overlooked one.

Let me turn now to the concrete questions.

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

The mood says that the topic situation is in the actual or in a nonactual world, respectively, and the *if* marks as undecided whether that topic situation has the comment features or not; it is, as it were, a "de-assertion marker". So, in *if it was raining*, it is marked as doubtful whether a past situation (*was*) in the actual world (indicative!) has the features [be raining]. And when this adverbial phrase is used in topic function within the main clause, it expresses that the topic situation of the main clause is in a world (here the actual world) which indeed also includes a situation with the comment features [be raining]. The *if*-clause is an adverbial phrase, and it has no truth value. But if the whole sentence *If it was raining, Ira stayed at home* is true, then the actual world must, of course, be such that it

contains a situation with the properties [PAST be raining]. Accordingly for the subjunctive counterpart.

4.2 What is (in) 'the topic situation'?

More concretely, the question is:

> […], 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?

Yes to the first question, and no to the second. I believe that many, if not all descriptive features of a clause can act as topic features as well as comment features, although there is a certain preference for some of them ("topic time, topic place, topic world, and topic entity (often encoded as grammatical subject)"). So, in *Without my lawyer, I don't say a word*, the initial adverbial phrase helps to describe the situation talked about. The fundamental difference to other notions of the topic is that it is always **the whole situation** about which something is said. This does not exclude that only a few elements of the entire situation are made explicit – for example, by an initial adverbial or by the tense and the mood component of the finite verb which specify the topic time and the topic world, respectively.

Since this point is really dear to my heart, let me repeat a mantra: **It does not make sense to ask whether an assertion is true or false, unless it is clear which situation is talked about, and that is specified by information in the assertion itself and by contextual information.** In very simple words, an assertion says that some situation has certain properties; therefore, it must somehow be made clear what that situation is and what the added properties are. This holds analogously for other sentence or clause types, for example for the "undecidedness marker" *if*. How this exactly works in different languages is a vast field, which – in spite of so much work on information structure over the last years – looks like an impenetrable jungle to me. I once tried to look into this jungle for French (Klein 2012) – but alas, things are much more complicated than said there.

4.3 Against or not against the facts?

The label "counterfactual" suggests that something that is said or assumed is "against the facts", i.e., does not hold in the actual world. This is indeed the typical case, but there are many counterexamples. In simple words, my explanation is as

follows. A counterfactual expression like *The lemming would be dead* expresses that a situation in a nonactual world has the properties [lemming be dead]. About the actual world, nothing is asserted at all. In particular, it is open whether a situation with these properties also obtains in the actual world. Why then the clear preference for the "against-the-fact" understanding? It mirrors the fact that we normally talk about the actual world; if someone deviates from that, there should be a reason. The most plausible reason is that if the speaker knew that the lemming is dead, he would probably say that. This is just as in *Yesterday, Ira was not drunk*, we assume that a speaker has a reason to restrict the claim to *yesterday*, and most likely, this reason is that according to the speaker, Ira is usually drunk and yesterday was an exception. Fabricius-Hansen writes "I am not sure [that WK's explanations] differ essentially from accounts to the effect that counterfactuality is an implicature (see e.g. [...] Arregui 2021 and references therein)". I do not think that counterfactuality itself is an implicature; what may be implicated in a way is what things are like in the actual world. Note, however, that just talking about a (conventional or conversational) implicature does not suffice because it does not explain the clear preference for the nonfactual understanding. Or maybe it does, depending on how the implicature is concretely spelled out.

While I have a clear opinion on this issue, I have no satisfactory reply to another problem that Cathrine Fabricius-Hansen raises in this connection:

> 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)?

The weak version – to which I am inclined to subscribe – is the former: there is no cross-world identity of situations; *Ira would be dead* and *Ira is dead* talk about two different situations with the same descriptive properties. This is not fully satisfactory, however; my intuition is that this may well concern the same Ira; this is just as *Today, Ira was here* and *Yesterday, Ira was here*, normally talk about the same Ira, albeit at different times. So, there is a cross-time identity for the persons and objects (see the beautiful first paragraph of Lewis 1986). On the other hand, we would probably say that the two sentences talk about two different situations, one today and one yesterday, with the same descriptive features [Ira PAST be here]. The problem which lurks here in the background is: what constitutes identity above and beyond the "attributes" which something, a person, an object, a situation, has. In the target paper, this problem is discussed in connection with the Billy Holliday song *If I were you, I would love me my whole life through*. Is this the same person or not? The Scholars had the maxim: *materia est principium individuationis*. But

situations have no matter, and there are many immaterial objects. Suppose you are asked: *Imagine two squares that are exactly alike*! On which ground do we consider them as individuals? A Helium atom has two protons with exactly the same properties – what renders them individual? I leave these questions to people who are more intelligent than me.

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

Fabricius-Hansen points out that I do not say very much about the role of the context. True, I lamented it myself. Human communication is fundamentally an interplay between information that is "in the words" and information from other sources – what is said before or after, situational context, world knowledge of the interlocutors. In the target paper, I tried to focus on what is specific to counterfactuality and counterfactual constructions. I assume that in a piece of discourse that includes counterfactuals expressions, the usual principles of textual coherence are operative as anywhere else. So, in *Mary went to Paul's party. She had a lot of fun* (cf. the reply to Hinterwimmer), we assume that *she* relates to Mary, although in principle, it could refer to some other person mentioned before. Similarly, we normally assume that she had a lot of fun at Paul's party, although that is not explicitly said, either, etc. What is specific, however, is the shift from actual to nonactual and vice versa. These shifts are indeed very common, and (6CFH) is a good example. Another case is arguments like *If Isa were in the garden now, we would see her from here. Now, we don't see her. Thus, she is not in the garden.* Phenomena like these would be a fertile field for further studies. In particular, such studies could shed more light on the somewhat mysterious question of why counterfactuality is so important and efficient in human reasoning.

5 Julia Zakkou: variations on Anderson's conditionals

Zakkou's comment systematically addresses the problem which Fabricius-Hansen raised in her third question. She confines the discussion to subjunctive conditionals and argues that

> at least *would have* conditionals […] indefeasibly convey that the antecedent is false (call this the minority view, to be distinguished from the majority view according to which the conditionals in question merely defeasibly convey that the antecedent is false). More concretely, I argue that *would have* conditionals have the antecedent's counterfactuality as a

non-at-issue content that belongs to the conditional's semantics (broadly construed) rather than its pragmatics. Even more concretely, I suggest that *would have* conditionals have the antecedent's counterfactuality as a semantic presupposition, but depending on one's background theory, one could also model it as a conventional implicature.

Now, the antecedent is an *if*-clause and thus an adverbial phrase; as such, it is neither true nor false. But for the purpose of the discussion, we may take the embedded clause as the antecedent and leave the *if* aside. The argument is thus that in *if Ira had been in the garden*, it is presupposed (or perhaps implicated) that Ira was not in the garden; it is left open whether this also holds for the "nonpast" version *if Ira were in the garden*. I found Zakkou's examples and arguments really interesting and an innovative contribution to the discussion. Nevertheless, I do not agree. In what follows, I will discuss some of her arguments.

Zakkou begins with Anderson's initial example. He wrote (Anderson 1953: 37):

> In the investigation of Jones' death, a doctor might say, "If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show ". Now in this context, the doctor's statement would probably be taken as lending support to the view that Jones took arsenic – it would certainly not be held to imply that Jones did not take arsenic. Such examples indicate that it is incorrect to say that we can infer the falsity of the antecedent from a true subjunctive conditional in the past (or any) tense.

Zakkou now adds a factual statement about the actual world and writes:

(4JZ) If Jones had taken arsenic, he would have shown the same symptoms he actually showed. So, he didn't take arsenic.

> This reasoning sounds strange. If, however, *would have* conditionals indefeasibly conveyed that their antecedents are false, (4JZ) should sound fine […].

Thus, the falsehood of the *if*-clause should be defeasible. Zakkou considers this argument line as flawed and gives some reasons for that. But Anderson's argument is much simpler than the apparent oddity of (4JZ), as the quoted passage from Anderson shows. In order to refute it, one would have to show that the correctness of the entire subjunctive conditional indeed implies the falsity of the antecedent. I agree with Zakkou, though, that what for (4JZ) to be fine "would additionally be needed is that the first sentence provides a reason for what is presented in (4JZ) as a conclusion." The following reasoning, she continues, sounds fine:

(5JZ) If Jones had taken arsenic, he would have shown the same symptoms he actually showed. So, he took arsenic.

And again, she gives some arguments against this reasoning. Note, however, that *So, he took arsenic* need not be true; it is something that is warranted by

plausibility considerations. At a court trial, a lawyer could easily knock it down by saying: "Well, but experts have demonstrated that the same symptoms can also be caused by fresh French fries in combination with old wasabi. And precisely that he had for lunch." The net result is, therefore: the antecedent can be true, and it can be false.

In her Section 3, Zakkou examines some "so-far neglected considerations that seem to support the claim that *would have* conditionals indefeasibly convey that their antecedents are false. I don't consider any of them conclusive, but I hope that they will further the debate." They concern explanative and concessive conditionals. She writes (two footnotes omitted):

> Consider now the *would have conditional* […] (18JZ):
>
> (18JZ) If Jones had taken cocaine, he would have had fun.
>
> And compare it to the following:
>
> (19JZ) Since Jones had taken cocaine, he would have had fun.
>
> (20JZ) Because Jones had taken cocaine, he would have had fun.
>
> These sentences sound strange. Consider also the following variation of (18JZ):
>
> (21JZ) If Jones had taken cocaine, he would still have had fun (he would have had fun nonetheless).
>
> And compare it to the following:
>
> (22JZ) Although Jones had taken cocaine, he would have had fun.
>
> (23JZ) Even though Jones had taken cocaine, he would have had fun.
>
> These sentences sound strange as well.

Her explanation is this:

> (19JZ) and (20JZ) as well as (22JZ) and (23JZ) sound strange because we expect them to convey both that the respective first part is false and that the respective first part is true. The former expectation is due to the subjunctive marking, the latter expectation arises from the semantics of the explanative and concessive conjunctions.

I think the reason for the oddity of these sentences is much simpler: The verb form in the antecedent is interpreted as indicative pluperfect, while the verb form in the consequent is considered as past counterfactual. If we replace the antecedent forms with the nonambiguous "*would* version", the sentences are no longer strange:

(15)   Since Jones would have taken cocaine, he would have had fun.

(16)   Because Jones would have taken cocaine, he would have had fun.

Adverbial phrases with *since* or *although* normally do not take a subjunctive; but it is possible, and if so, the strangeness disappears.

I think that the interpretation of pluperfects as "fake pasts" is also at the heart of some other types of conditionals that Zakkou considers in the remaining section. Interestingly, she concludes her discussion as follows (last para of Section 3):

> [I suspect that] (40JZ) [= *Only if Jones had taken arsenic, he would have shown the same symptoms he actually showed*] sounds okay because we apply a repair strategy and interpret 'had taken' as tense rather than fake tense; in other words, we interpret the subordinate clause as plainly being about the past.

I think that is correct, although I would not call it a repair strategy; it is just one way to use an ambiguous form.

In sum, I do not concur with Zakkou's suggestion "that *would have* conditionals have the antecedent's counterfactuality as a semantic presupposition" or that it is part of their semantics that the antecedent is false. But her arguments are really original, and I think more arguments of this sort are needed, when we want to understand why the antecedent as well as the entire conditional in subjunctive mood are sometimes considered as "factual" and sometimes as "nonfactual", with the latter being the normal case.

6 Isabel Repiso: universal parameters yielding counterfactuals

> The Mind, that Ocean where each kind
> Does straight its own resemblance find;
> Yet it creates, transcending these,
> Far other Worlds, and other Seas;
> Annihilating all that's made
> To a green Thought in a green Shade.
> Andrew Marvell *The Garden*

This comment and the next one deal with an aspect of counterfactuality that is not treated in the target paper – the psychology of counterfactual expressions. The reason is not that I consider this aspect irrelevant, quite the opposite. It is rather what the famous lexicographer Samuel Johnson said when he was asked why some words are missing in his dictionary: "Ignorance, sheer ignorance!"

Psychology is a vast field. The particular area on which Repiso dwells is the remarkable capacity of the human mind to create alternative worlds, to make up alternatives to factual events. We dream to be heroes, to meet our dream partner, real in the dream; we invent "alternative facts", we build castles in the air and sell them to others. We use that very capacity also for scientific reasoning, as in reductio at absurdum proofs in mathematics (*If 221 were a prime number, it would not divide by seventeen. It divides by seventeen. So, it is not a prime number*). The same cognitive capacity lets Tevje the Dairyman sing *If I were a rich man, diddle diddle da* and it would let him, the simple man, reason *If I were a rich man, I would be happy. I am not happy. Thus, I am not a rich man.* It is irrational, and it is rational. Tevje probably does not reason like that, he is fiction anyway. But clearly, counterfactual reasoning plays an eminent role in decision-making and judgments in everybody's daily life. What has psychological research revealed about counterfactual reasoning in daily life? Repiso sums up four core findings:

A. In daily life, upward counterfactuals – i.e., assessments resulting from the comparison of reality to better scenarios – are more frequent than downward counterfactuals – i.e., assessments coming from the comparison to worse scenarios – (Dray and Uphill 2009; Summerville and Roese 2008). In other words, people are more inclined to imagine how things could have been better as opposed to worse.
B. The prominent mechanism used to undo negative outcomes consists in restoring the normal value of a variable rather than introducing unlikely occurrences (Kahneman and Tversky 1982).
C. Negative outcomes are judged more regrettable when failed by a small margin (Kahneman and Miller 1986; Kahneman and Tversky 1982).
D. In two independent events, people tend to blame the agent responsible for the most recent event (Miller and Gunasegaram 1990).

Each of these four principles sounds plausible to me. Someone who is happy with his life, his job, her friends is less inclined to think about alternatives than someone who is unhappy, has a lousy job, and has no friends at all. It is surprising, though, that principle A (and perhaps the others as well) should be confined to alternatives in the past ("how things could have been better"), because that seems to run against the commonly held view that in the past, everything was better, including the future. But many also ruminate forever about the day when they stepped into the wrong train and why they did that and whether they can travel back and change trains.

The four principles should not be at variance, and indeed, Repiso sums up A and B under a more general principle:

> The salient role of upward counterfactuals (principle A) and the mechanism of undoing negative outcomes (principle B) can be summarized into the following:
>
> *If p* [RESTORING normal value of a variable], *(then) q* [PREVENTING negative outcome]

But that would mean that the "normal" state of affairs is the one that prevents negative results. Not my experience. Maybe principle A only reflects the just-mentioned experience that at some point in the past, something went wrong, and one would like to undo it, such that the course of events would have been a better one. I cannot answer that (sheer ignorance); so, I would rather ask Isabel three questions:

(a) Is the value reset to "normal state" or to "favorable state"?
(b) Why is principle A restricted to "how things could have been better" and doesn't include "would be better"?
(c) Do principles A – D indeed capture "counterfactual reasoning in daily life", or do they only mirror particular preferences in tight experimental settings?

A second problem that Repiso addresses is the necessity to distinguish between what the interlocutors[8] believe the actual world is like and what it is actually like. This is an important point; I think that the perennial discussion about Ernest Adams' famous "Oswald examples" (*if Oswald hadn't shot Kennedy, Kennedy would be alive today* versus *if Oswald didn't shoot Kennedy, Kennedy is alive today* in Adams' original version from 1970) is often misguided because no clear distinction is made between the linguistic meaning of these two sentences, on the one hand, and the contribution of the interlocutors' world knowledge, on the other.[9] Any judgment on what is true and what is not true is based on an inter-action of these two information sources. If I know that someone is a notorious liar, I judge that his assertion *I just bought a Porsche* is false, unless I also know that he indeed just bought a Porsche. But that knowledge does not affect the truth value of the sentence *I just bought a Porsche*. We do not want to say that the truth of an assertion depends on what a particular person believes or knows. Principles A – D are based on the interaction of linguistic meaning and world knowledge. They do not describe the linguistic meaning of counterfactuals but the way in which people reason with their help. This also holds for the notion of warrant, which plays an important role in my analysis of counterfactuality. As Repiso

---

8 It is not only the speaker's belief which is relevant, but also the listener's, in general: the beliefs of interlocutors.
9 I believe the indicative example is simply an enthymeme, in which the premise *Oswald did not shoot Kennedy* is omitted because it is provided by world knowledge. That is not the case in the subjunctive variant, because there is no such unstated premise.

correctly states, the warrant is something that guides our judgment; it is not a part of the semantics of counterfactual expressions, and I appreciate her reference to earlier studies in which this idea plays a role.

In the last section of her comments, Repiso examines how "heuristic counterfactuals (HC)", i.e., those which apply principles A and B, relate to semantic intuitions that are traditionally assigned to counterfactuals: restrictedness, consequence, reality status, probability, remoteness, (non)factuality. She considers HCs like these:

i. Conditionals under the form *If p (then) q*
ii. Whose finiteness marking are subjunctive verbal forms
iii. Whose propositional content satisfies the following schema: *If p* [RESTORING normal value of a variable], *(then) q* [PREVENTING negative outcome]

Not all traditional intuitions play a role in such HCs. That should not come as a surprise, because (iii) goes beyond the mere linguistic side. Repiso also points out that in other languages (but also in English), counterfactuality often comes with the indicative. This is correct. In fact, one might argue that it is the normal case to leave unmarked whether something is said about the actual world or about some nonactual world: there is "marked counterfactuality" and "unmarked counterfactuality". The best example is fictional literature, where it is often unclear whether something in a novel is about the actual world or about the nonactual world created by the author (see target paper, Section 2.3.5). Another salient example is deontic texts like laws or contracts, which are typically in the indicative; the sentence *La recherche de la paternité est interdite* (Napoleon's major contribution to the Code Napoléon) does not depict something in the actual world; at best, one could say that it changes the actual world. The third type of nonactual indicatives is pretending plays: *you are Napoleon – I am Josephine*. And then, there are often indicatives in counterfactual reasoning. If this is correct, and it is correct, then one still could say that the subjunctive indicates that the world talked about is a nonactual world. But there are other options, for example, the indicative, seen as the unmarked form for several moods (see footnote 17 in the target paper).

The field of counterfactuality has been plowed for many years. But one of the shortcomings of this long tradition is a certain tunnel vision, and although the analysis I propose tries to broaden the view, it is still very restricted. A deeper look at the way in which people use counterfactuality in their daily reasoning – the way in which they use it, not the way in which they should use it – surely helps to gain a more accurate understanding of this remarkable achievement of the human mind.

7 Eugenia Kulakova: taking a psychological view on another way to look at counterfactuals

What was said above also applies to Kulakova's psychological perspective, except that it does not address reasoning in everyday contexts but the processing of counterfactual conditionals in real time. I will take up three of the various points she makes, and then close with a suggestion.

7.1 Counterfactuals are gradual

Fauconnier (1994) has described a gradient of nonactual expressions with the strongest being negations (*He did not win the Palme d'Or*), followed by wishes (*He wishes he had won the Palme d'Or*), counterfactuals (*If he had won the Palme d'Or then he would have rejoiced*), and modals (*He could have won the Palme d'Or*).

There is certainly a gradient between these (and perhaps other) expressions. But what exactly is that gradient? Is it the amount of remoteness which is traditionally associated with counterfactuality, and if so, on which scale do we measure this distance? Is it the similarity between the actual world and the world in which there is a situation with the properties [he win PAST the Palme d'Or]? That is reminiscent of the "minimal revision" idea of the Stalnaker-Lewis-line; but just as little as it is plausible for counterfactual conditionals, in particular, it is plausible for the degree of nonactuality. Whether the Milky Way has one billion stars more than it actually has or seven billion stars less, makes a substantial difference; but it does not matter for the gradient between the expressions above. Nor would it matter if the speed of light were, say, twice as fast as it actually is; a law of nature would be substantially different, but that is not crucial here. Is it the likelihood that he won the Palme d'Or in the actual world – another of the traditional intuitions of counterfactuality? Certainly not, because it may have been extremely unlikely that he won it ("if he had won the big lottery"). Or is it the likeliness that he indeed won the Palme d'Or, if one of these sentences (as a whole) is true? To my mind, that is the most plausible interpretation of such a gradient, but even that one is problematic. First, on which scale is the likelihood determined; second, in *He wishes he had won the Palme d'Or* dramatically changes if the subject is not the same, as in *Chabrol wishes that Wenders had won the Palme d'Or*. In sum, I think that there is a soupçon de vérité in the idea of a gradient, but it is certainly not easy to pin down.

7.2 Polarity reversal and negation

> A commonality of the various counterfactual expressions is that they convey an implicit
> negation of the information they explicitly formulate, a characteristic that has been labeled
> polarity reversal (Van linden and Verstraete 2008). This feature signals that in spite of the
> factually false supposition conveyed by their surface form, the speaker is aware that the
> opposite is in fact the case.

That is in a way correct, but there is also a fundamental difference between this
"implicit negation" and the truth reversal of a real negation: in a real negation, the
positive and the negative sentence, as in *He won the Palme d'Or* and *He did not win
the Palme d'Or*, cannot both be true nor can they both be false. In a counterfactual,
e.g., *If he had won the Palme d'Or*, it is possible that he won it, and it is possible that
he did not win it, where the latter is the preferred understanding. Note that the
truth reversal of a real negation only holds under topic consistency; it is perfectly
possible that *He won the Palme d'Or* is true for 1982 and false for 1985. A coun-
terfactual, however, speaks about a nonactual world, and thus, it does not say
anything about what is the case in the actual world: there is no topic consistency by
definition.

7.3 Dual meaning

In their excellent survey on online counterfactual language processing, Kulakova
and Nieuwland write:

> Cognitive and linguistic theories of counterfactual language comprehension assume that
> counterfactuals convey a dual meaning. Subjunctive-counterfactual conditionals (e.g., 'If
> Tom had studied hard, he would have passed the test') express a supposition while
> implying the factual state of affairs (Tom has not studied hard and failed). […] Here, we
> review the available studies that examine online counterfactual language comprehension
> through behavioral measurement (self-paced reading times, eye-tracking) and neuro-
> imaging (electroencephalography, functional magnetic resonance imaging). While we
> argue that these studies do not offer direct evidence for the online computation of coun-
> terfactual dual meaning, they provide valuable information about the way counterfactual
> meaning unfolds in time and influences successive information processing. (Kulakova and
> Nieuwland 2016a: 49)

And in their own informative study on the role of pragmatic skills in the inter-
pretation of counterfactuals, they write:

> Counterfactual thought allows people to consider alternative worlds they know to be false.
> Communicating these thoughts through language poses a social-communicative challenge
> because listeners typically expect a speaker to produce true utterances, but counterfactuals
> per definition convey information that is false. (Kulakova and Nieuwland 2016b: 814)

The idea of a dual meaning, as defined here, hits a point but also misses a point: counterfactual conditionals are not false by definition; in fact, a great deal of the research tries to define their truth conditions (see, e.g., Section 2.2 of the target paper or Hinterwimmer's comments and my replies). They just do not assert something about the actual world. The question, therefore, is: How is it possible for an interlocutor to draw conclusions from something said to be true in that non-actual world to what is true in the actual world? Typically, the cognitive operations of the interlocutor lead to the assumption that a situation like the one that is true (!) in the nonactual world (the world in which Tom has studied hard) is false in the actual world. The Stalnaker-Lewis idea of maximal similarity (or minimal revision) is an attempt to characterize these operations, my idea of "the same warrant" is another. None of these, however, is phrased in terms of factual psychological operations. It would be interesting to see whether that is possible.

A major complication is, as already said above, that something said to be true in some nonactual world can also be understood as true in the actual world. If someone says *Your Hamster would have died anyway*, then the assumption is clearly, that the hamster is dead, and not that he did not die (see target paper, exx. (5WK) and (31WK), and Section 4 above). So, the cognitive processes may lead into one direction, the preferred one, but also into the opposite direction. Why that? Particles such as *as well, also, anyway* may steer our mind toward a particular interpretation. But I doubt that this is the whole story.

7.4 A suggestion on my part

Over the many years I have spent at the Max-Planck-Institute for Psycholinguistics in Nijmegen, I learned two things about the relation between linguists and psychologists: first, that the constraints of clean experimentation render it difficult and often impossible to live up to the sophistication of linguistic analyses; and second, that linguists like to refer to the results of psycholinguistic experiments if they confirm their theory, but pass tacitly over them if they contradict it. I suspect that this also applies to me; but here is a possible exception.

The semantic analysis of a word or a larger expression in a natural language is ultimately based on the intuitions which its speakers connect with it (on what else?). How is this with the little word *if*? There are several options on the market, for example:

(a)   Introducing a clause of condition or supposition (the protasis of a
        conditional sentence).
        On condition that; given or granted that; in (the) case that; supposing that;
        on the supposition that.

This is the main definition of the best English dictionary in the (actual) world, the Oxford English Dictionary.

(b)   *if* has no meaning proper but only serves to indicate that in *if* p, it restricts an overt or covert operator

This is the view of Angelika Kratzer's influential paper on conditionals (first 1986, revised in Kratzer 2010).

(c)   […], we suggest that *if*-clauses are simply definite descriptions. […] we will suggest that *if* should be seen as the form taken by the word *the* when it is applied to a description of possible worlds.

This is the view of Philippe Schlenker (2004: 417) as well as Ebert et al. (2014); see my reply to Grønn in Section 3.

(d)   *if* in *if* p marks that according to the speaker, it is undecided whether p is true or not

This is my own view. A bit more precise: According to the speaker, it is undecided whether the topic situation of p has the comment features of p. In a way, this is the counterpart to assertion marking, which indicated that the topic situation has the comment features; *if* is an "undecidedness marker" or a "de-assertion marker".

    Can these four views be checked in a psycholinguistic experiment, offline or online? If so, I would not promise to accept the result, but I would not pass it over with silence, either. And is there any way to check whether my semantic intuitions of what *if Mary was at Paul's party* (see 1SH) means are shared by others? Such results of processing studies might indeed change our view on what *if*-clauses express.

8 Verstraete and Luk: shaking up counterfactuality: even closer to the linguistic facts

This comment (like the following one by Van linden) is close to my own views of counterfactuality. Therefore, my reply will be much shorter, and to most points, I simply agree. In what follows, I will focus on the notion of polarity reversal which I find very appealing but which also raises problems, followed by two remarks on smaller points.

    In Van linden and Verstraete (2008) – a paper to which I owe a lot –, polarity reversal is not defined but introduced by means of some examples (p. 1866):

> These [i.e., the following] conditional structures are counterfactual because they involve a reversal of the polarity marked in the structure: part of the interpretation of the structure with positive polarity in (1VLa) is that the police troops were not in fact sent in, and that as a consequence people were killed, while part of the interpretation of its counterpart with negative polarity in (1VLb) is that troops were sent in, and that they were able to prevent the killing.

> (1VL)   (a)   If they had acted and sent in enough police troops, says the report, the bloody episode could have been prevented.
>   (b)   If they hadn't acted and sent in enough police troops, says the report, the bloody episode could not have been prevented.

> In addition to the conditional structures illustrated in (1VL), many languages also have counterfactual constructions in simple clauses [examples omitted here]. As with conditional constructions, the interpretation of these structures involves a reversal of polarity.

In the examples, sentences without negation stand for positive polarity, sentences with a negation stand for the negative polarity. This is very intuitive. But to my mind, polarity reversal of counterfactuals cannot mean the familiar "truth reversal" by a negation marker such as *not* in English: *The defendant would have been found guilty* and *The defendant was not found guilty* relate in very different ways to positive *The defendant was found guilty*. The difference becomes immediately clear when we compare them when given in reply to the question *Was the defendant found guilty*? or the question *What was the verdict of the court?* The difference is even more salient when another person confirms the *not*-reply: *The defendant was not found guilty. – Quite right, the defendant would have been found guilty*. This is not to deny that there is something true in the idea of a polarity reversal, but it needs elaboration (see also Klein 2018).

My own analysis of the polarity reversal effect is that counterfactual expressions do not assert anything about the world as it is – the factual, real, actual world. It speaks about a situation in a nonactual world, a world which is imagined, and it says that in that imaginary world, the defendant was found guilty. From there, we may come to the conclusion that in the actual world, he was not found guilty. But the inference to *He was found guilty* is possible in some contexts, too. As to truth reversal by negation markers like *not*, it is important to keep in mind that it only makes sense when positive and negative sentences speak about the same situation, i.e., under "topic consistency". The sentences *It was raining* and *it was not raining* may simultaneously be correct, when the first case relates to the weather in Marseille and the second to the weather in Bergen. This is quite obvious, but it is completely ignored in the studies of negation from Antiquity to modern linguistics (cf. the detailed analysis in Klein 2018). In a counterfactual like *It would have been raining*

and its indicative counterpart *It was raining*, there is – under my analysis – no topic consistency by definition because they relate to situations in different worlds. This is my preferred view, although the possibility of cross-world identity of situations is a tricky one (see the reply to Fabricius-Hansen in Section 4.3).

A point where I agree with Verstraete and Luk (and already Van linden and Verstraete 2008) is the need to distinguish between various types of modality that may interfere with the mere notion of counterfactuality in a clause, even in a single verb ("double layer of modality"). In the target paper, I did not look at modal verbs like *could, should, might* (and one might add *must*, which can be used in a deontic and in an epistemic sense), because they introduce a world of problems on their own. So, I have nothing to say here about these problems, either. But I would add a few words on the optionality of the protasis, which Verstraete and Luk discuss in Section 3. I indeed considered it as syntactically optional. This has primarily to do with the fact that in all languages which I happen to know sufficiently well, the protasis is an adverbial phrase, and adverbial phrases are usually syntactically optional (except in predicative uses like *Ira was in the garden,* except again in clearly elliptical cases). I would not go so far as to say, however, that the restriction on the topic situation of a main clause can always be achieved by other means, like nonclausal adverbial phrases or by context alone; in that sense, protases might sometimes be semantically mandatory. As Roman Jakobson put it: "Languages differ essentially in what they must convey and not in what they may convey." Incidentally, while simple *would*-sentences without a protasis are clearly possible, they seem to call much more for additional information (for example a following *but xyz*) than simple sentences with other modal verbs, cf. *Ira would have come* versus *Ira should/could/might have come*. But my intuitions might or could but not should be deceptive here.

9 Van linden: a usage-based approach to counterfactuality: optionality of the apodosis

As probably every linguist, I firmly believe in Gricean or Gricean-type maxims. But I also have a problem with them – they are so general that I feel you can plug any hole with them and that they do not lend themselves to falsification. Thus, it is quite correct when Van linden writes:

> In fact, while examples like *If Isa were in Berlin, she would be in Marseille* may pop up in philosophical discussions (perhaps in their metalinguistic use), they are immediately felt to be 'unnatural' in everyday conversations because they are semantically inconsistent, and the speaker is thus thought to violate the principle of manner […]

But to my mind, it is not only unnatural but false. Moreover, a good linguistic analysis should also be able to explain why they are semantically inconsistent. This, however, is only possible if we can check what precisely is expressed against what we know about the world – the actual world, in which people cannot be at different places at the same time. We must be able to check whether something can be true in the actual world. Or to put it differently, and thus avoiding the notions of truth and falsity: we must in principle be able to check whether some situation in the actual world can have the properties which the speaker assigns to it. To this end, he or she must understand the sentence correctly, on the one side, and have access to the actual world, on the other. The latter is not a linguistic issue, but the former is. I do not think that a polarity reversal analysis is in any way different here; under this approach, the interlocutor must understand "that the police troops were not in fact sent in" in (1VL) – see my reply to Verstraete and Luk above.

My own final analysis in (48WK) is not phrased in terms of true and false but in what the speaker expresses and what the listener hopefully understands. That can be interpreted in two ways. In the Billie Holliday example *If I were you, I would love me my whole life through* (52WKb), the communicative message that she wants to convey is certainly "I advise you to love me". But this message is phrased in a very particular, highly sophisticated linguistic form – in a counterfactual which the listener must have understood in order to derive the speaker's intention from it. She could convey her intention in a much simpler form, of course; she could simply say *I advise you to love me*. But it would not be such a beautiful play with what is the case and what is not the case and what should be the case if things were different in the way described by the words.

The potential optionality of the apodosis, which Van linden discussed in Section 2, was something totally new to me (and probably to most who work on counterfactuals). Clearly, there are elliptical reductions like *If Casillas had played, Madrid would have lost, and if not, not.* or *If Casillas had played, Madrid would have lost, and if not, as well.* But these are purely structural phenomena that have nothing to do with counterfactuality in particular. I find Van linden's analysis of the examples convincing throughout, and I have only two small annotations. They both concerns examples like *If only he hadn't been so drunk* (from 3VL) which "have been described as conventionally expressing the speaker's unfulfilled wish concerning a past situation". This is related to Isabel Repiso's cases of heuristic principle A, except that they are not in an experimental setting but in authentic discourse (or well-invented authentic discourse). And like in Repiso's cases, I wonder whether the phenomenon is restricted to past situations, and if so, why. In *If she would only love me!*, there is also an unfulfilled wish – but it is not in the past. The other annotation is rather a question: Why does the effect disappear if the particle *only* is omitted? Normally, *only xyz* means "*xyz* and no alternative to *xyz*".

That does not work here. The easiest way out is to say that *if only* is a set phrase, a construction like *let alone*. But first, that idea works for *If only he hadn't been so drunk* but not for *If he would only love me!* And second, even if it is a set phrase, this does not dispense the linguist from the obligation to specify what that phrase means as a lexical item.

10 Concluding remarks

Linguists primarily want to find out what the semantic and morphosyntactic properties of counterfactuals are. Psychologists are primarily interested in the role of counterfactuals in human reasoning. As Kulakova puts it: "Why do people use counterfactuals and what do they try to communicate with them? Why does it matter to us whether a counterfactual is true or not?" These are good questions.

I do not think that there is a single reason. What motivated Dante to put the gruesome fabrications of his merciless mind into the sublime verses of the *Inferno*? Why did counterfactual "fictions" play such an eminent role in many areas of human thought, as Vaihinger (1923) convincingly demonstrated a century ago? They still do. What we believe about global climate changes in the years to come is entirely based on computer models. But climate is a phenomenon of unheard of complexity, and any computer modeling requires counterfactual simplifications. One of them is the common reduction of the three-dimensional atmosphere to a two-dimensional grid – i.e., to flat earth (see Prather and Hsu 2019). No scientist believes that the earth is flat – but this counterfactual assumption is useful, and so it is made.

But not only writers and scientists permanently operate with counterfactual assumptions. It is an essential component in everybody's daily reasoning. Why that? I believe it allows us to reason by analogy if there is no other way. If we want to know whether an assertion holds, we must check the situation described by it. If we want to know whether the assertion *In Pontefract, it was raining* is correct, we may have no access to the situation so described. But we may have – or have had – access to a situation in the actual world which is "sufficiently similar", and we may draw conclusions from there to the situation in Pontefract. "Sufficiently similar" does not mean "maximally similar"; the religion of the people at these two places, important as it may be, does not matter (?). The "warrants" of our judgment must be the same – those factors which tell us what the weather is like. In that case, there is a shift within the actual world. But our mind allows us to leave the actual world and to create situations in a nonactual world as we want it to be like. That allows us to judge the case and to issue the verdict, as it were, *in effigie*. And if the warrants on

which we base our judgment are the same in that world and in the actual world, we are entitled to draw conclusions about what is the case in the actual world.

I much enjoyed this discussion. If we continued it, it would be marvelous to continue it, but if not, it would also be marvelous to continue it.

References[10]

Arregui, Ana. 2008. On the role of past tense in resolving similarity in counterfactual. In Atle Grønn (ed.), *Proceedings of SuB12*, 17–31. Oslo: ILOS.

Arregui, Ana. 2009. On similarity in counterfactuals. *Linguistics and Philosophy* 32. 245–278.

Ebert, Christian, Cornelia Ebert & Stefan Hinterwimmer. 2014. A unified analysis of conditionals as topics. *Linguistics and Philosophy* 37. 353–408.

Klein, Wolfgang. 2000a. An analysis of the German perfekt. *Language* 76. 358–382.

Klein, Wolfgang. 2000b. Was uns die Sprache des Rechts über die Sprache sagt. *Zeitschrift für Literaturwissenschaft und Linguistik* 118. 115–149.

Klein, Wolfgang. 2010. On times and arguments. *Linguistics* 48. 1221–1253.

Klein, Wolfgang. 2012. The information structure of French. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure*, 95–126. Berlin: de Gruyter.

Klein, Wolfgang. 2018. On the scope of negation. In Wolfgang, Klein (ed.), *Looking at language*, 253–282. Berlin: De Gruyter.

Kulakova, Eugenia & Mante Nieuwland. 2016a. Understanding counterfactuality: A review of experimental evidence for the dual meaning of counterfactuals. *Language and Linguistics Compass* 10. 49–65.

Kulakova, Eugenia & Mante Nieuwland. 2016b. Pragmatic skills predict online counterfactual comprehension: Evidence from the N400. *Cognitive, Affective, & Behavioral Neuroscience* 16. 814–824.

Lewis, David. 1986. *On the plurality of worlds*. Oxford: Blackwell.

Prather, Michael & Juno Hsu. 2019. A round Earth for climate models. *Proceedings of the National Academy of Sciences* 116. 19330–19335.

Schlenker, Philippe. 2004. Conditionals as definite descriptions. *Research on Language and Computation* 2. 417–462.

---

10 Publications that are already quoted in the target paper or in the comments are not listed here again. My apologies for referring so much to my own work; but the core ideas in the target paper are rooted in it.