

# Counterfactual Skepticism and Multidimensional Semantics

H. Orri Stefánsson

[www.orrstefansson.is](http://www.orrstefansson.is)

June 9, 2017

## Abstract

It has recently been argued that indeterminacy and indeterminism make most ordinary counterfactuals false. I argue that a plausible way to avoid such counterfactual skepticism is to postulate the existence of primitive modal facts that serve as truth-makers for counterfactual claims. Moreover, I defend a new theory of ‘might’ counterfactuals, and develop assertability and knowledge criteria to suit such unobservable ‘counterfacts’.

## 1 Introduction

Counterfactuals seem indispensable to practical deliberation, ordinary discourse and philosophical theorising. Recently, however, it has been argued that most ordinary<sup>1</sup> counterfactuals are false (Hájek [ms](#)). Suppose my partner is emptying the dishwasher, and after putting a plate into the cupboard I tell her:

If you had dropped the plate it would have fallen to the floor. (1)

In this circumstance, my partner and I would both agree that (1) seems intuitively and

---

<sup>1</sup>A counterfactual is not ordinary if it has a consequent that is necessarily true, or if its consequent is logically or physically necessitated by its antecedent, or if it has a true antecedent. A counterfactual can be ordinary and true, according to the argument in question, if its consequent is qualified by e.g. ‘probably’, ‘almost certainly’, and so on.

straightforwardly true. But things are not always as they intuitively seem. For instance, our currently best physical theory, quantum mechanics, apparently entails (according to a quite widespread interpretation)<sup>2</sup> that there is a positive chance that had my partner dropped the plate it would have flown off sideways (Hawthorne 2005). Therefore, I can truly say to my partner:

If you had dropped the plate it might have flown off sideways. (2)

But if the plate flies off sideways then it does not fall to the floor. Similarly, if it *might* have flown off sideways then it seems false that it *would* have fallen to the floor. After all, if it might have flown off sideways then it might *not* have fallen to the floor. Hence, it seems that (1) must be false. But quantum mechanics allows for similar manoeuvres for almost *any* counterfactual we ordinarily take to be true. Therefore, we seem to be forced to accept the counterintuitive conclusion that most counterfactuals we take to be true are actually false.

The problem is even worse: It seems that most ordinary counterfactuals are false even if quantum mechanics is false. For instance, the antecedent in (1) does not specify the way in which my partner drops the plate. But there are ways in which one could precisify the antecedent of (1) such that the plate would not fall to the floor. For instance, my partner might provocatively and playfully drop the plate with one hand just to catch it with the other. Therefore, it seems true that:

If you had dropped the plate you might have caught it before it reached the floor. (3)

But again, it seems that if (3) is true then (1) must be false. After all, if you might have caught the plate before it reached the floor then it might not have fallen to the floor. And since most counterfactuals that we ordinarily utter and take to be true have antecedents that are similarly underspecified, it seems that, with or without the aid of quantum mechanics,

---

<sup>2</sup>As a referee for *Erkenntnis* points out, quantum mechanics is deterministic according to some interpretations of the theory, such as Bohmian mechanics. However, the arguments that I am addressing assume a non-deterministic interpretation of quantum mechanics, and therefore I will do so too.

we reach the counterintuitive conclusion that most counterfactuals we take to be true are actually false (Hájek [ms](#)).

The aim of this paper is to respond to this skeptical challenge by defending a theory of counterfactuals that is consistent with the truth of the counterfactuals that we ordinarily accept. I will suggest that the presumption of primitive modal facts, which I will call *counterfacts*, that serve as truth-makers for counterfactual claims, offers a plausible response to counterfactual skepticism. The main reason I offer in favour of the existence of such counterfactuals is simply that there are many counterfactuals, that we intuitively take to be true, but whose truth seems threatened by arguments like those above, *if* we assume that their truth-makers are part of the ordinary, non-modal facts. In contrast, these counterfactuals could well be true, despite these arguments, if their truth is grounded in primitive counterfactuals.

Others have hinted at a similar solution to the skeptical problem (e.g. Hawthorne [2005](#)). But, in addition to defending this position in more detail than what has previously been done—which includes explaining how one can have evidence of such primitive counterfactuals—what is distinctive about my proposal is that I base it on a recent *multidimensional possible world semantics for conditionals* (Bradley [2012](#), Stefánsson [2014](#)), which makes reference to primitive modal facts that serve as truth-makers for counterfactuals.

It might be worth acknowledging right away that my proposed response to counterfactual skepticism both makes assumptions and has implications that some will find counterintuitive (and raises hard questions that my paper does not answer). For instance, some may find the idea that there are primitive counterfactuals to be unintuitive, while others will have a hard time accepting the implication that, say, there is a fact of the matter as to *how* the plate would have been dropped, had it been dropped. However, it should be kept in mind that any response to the above arguments for counterfactual skepticism is bound to have counterintuitive implications and/or premises. After all, what is so interesting about these arguments, is that they show that some intuitive premises lead to a very counterintuitive conclusion, namely, the conclusion that most of the counterfactuals we take to be true are actually false.<sup>3</sup>

---

<sup>3</sup>Quantum mechanics is admittedly not an 'intuitive premise'. However, the intuitive premise is that we

The rest of this paper is structured as follows. In the next section I discuss the skeptical challenge in more detail and explain how it raises special problems for the classical Stalnaker-Lewis theory of counterfactuals. Section 3 provides some preliminary considerations in favour of primitive counterfactuals. In Section 4.1 I describe the multidimensional semantics for conditionals, which I use to formalise my proposed solution to the skeptical problem. In Section 4.2 I argue that abandoning the Humean supervenience thesis is a justifiable cost of responding to the skeptical challenge. In Section 4.3 I discuss how the theory I suggest accounts for the seeming inconsistency between ‘would’ counterfactuals such as (1) and the corresponding ‘might-not’ counterfactuals, such as those entailed by (2) and (3). Finally, Section 4.4 explains how one can have evidence of primitive counterfactuals and develops an assertability criterion for counterfactuals based on the proposed framework.<sup>4</sup>

## 2 The skeptical challenge

We have seen that the truth of ordinary counterfactuals seems, on one hand, undermined by the fact that their antecedents are not specific enough to ensure the truth of their consequents. On the other hand, chancy laws and the indeterminism postulated by quantum mechanics (according to some interpretations of the theory) seem to undermine the truth of ordinary counterfactuals irrespective of how their antecedents are specified. Following Alan Hájek (ms), I will call the first problem *the challenge from indeterminacy*, the second problem *the challenge from indeterminism*.

---

philosophers should accept the physical theory that more or less all physicist accept.

<sup>4</sup>I am not the first to respond to counterfactual skepticism. To take just a few examples, David Lewis (1979), Anthony Gillies (2007), Robbie Williams (2008), Jonathan Ichikawa (2011), Hannes Leitgeb (2012a, 2012b), Sarah Moss (2012, 2013), Moritz Schulz (2014), and Karen Lewis (2016) all offer responses to arguments for counterfactual skepticism. While these proposals are certainly very interesting and worthy of a detailed discussion, I will have to set them aside in this paper, due to lack of space. Hence, I cannot claim to have shown that my response to counterfactual skepticism is the best one on offer. Instead, I simply offer it as a plausible candidate response to such skepticism.

## 2.1 The challenge from indeterminism

Suppose I have a fair coin in my pocket that will never be tossed. Is the following counterfactual true?

If the coin had been tossed it would have landed heads. (4)

Arguably it is not. The coin has the same chance of landing heads as tails. So it seems no more true that it would show heads than tails. But it cannot be true that it would show heads and that it would show tails. Therefore, (4) seems false, and so does the counterfactual that the coin would have landed *tails* if tossed.

As Hájek (ms) points out, the above conclusion also seems to hold if the coin is biased. Say the coin has a 0.9 chance of landing heads up if tossed. Would that make (4) true? It seems it would not. For there is still a 0.1 chance that the coin would have shown tails if tossed. So the coin *might* have landed tails up if tossed, which seems inconsistent with the claim that it would have shown heads. And it seems we would reach the same conclusion no matter how biased the coin were in favour of heads.

It is instructive to evaluate (4) by the orthodox, Stalnaker-Lewis theory of counterfactuals (Stalnaker 1968, Lewis 1973). In what follows,  $\Box \rightarrow$  denotes the counterfactual would-operator, such that  $A \Box \rightarrow B$  denotes, for any propositions  $A$  and  $B$ , that if  $A$  were the case then  $B$  would be. Setting aside Stalnaker and Lewis' disagreement over what the latter called 'Stalnaker's assumption',<sup>5</sup> their theory is that  $C \Box \rightarrow B$ , where  $C$  denotes any *non-contradictory* proposition, is true at world  $w$  just in case  $B$  is true at all the  $C$ -worlds that are *most similar* to  $w$  (where ' $C$ -worlds' is short for 'worlds where  $C$  is true'). Now if the similarity ordering is to be grounded in the ordinary language concept of 'similarity', as both Stalnaker and Lewis wanted, then since the coin has some objective chance of landing tails, it seems that among the coin-toss-worlds that are most similar to the actual world there are some worlds where the coin lands tails. Therefore, it is false<sup>6</sup> that the coin would have landed heads if tossed.

---

<sup>5</sup>The assumption is that for any world  $w$  and any antecedent  $A$  there is a exactly one  $A$ -world most similar to  $w$  (Lewis 1973: 78).

<sup>6</sup>Strictly speaking, counterfactual (4) is not false but *indeterminate* on Stalnaker's theory (1968, 1984). In fact,

If quantum mechanics is true, then, on a popular interpretation of the theory, *any* ordinary counterfactual is like (4), in the sense that there is some chance that its consequent would be false if its antecedent were true. Recall counterfactual (1) from the introduction: ‘If you had dropped the plate it would have fallen to the floor.’ Now according to a common interpretation of quantum mechanics, it is truly indeterministic whether the plate would have flown off sideways or not. But that means that some worlds where the plate behaves this way are no more different from the actual world than the worlds most similar to actuality where it does not behave this way: These counterfactual worlds *only* differ in whether the plate flies off sideways or not (and in facts that are caused and/or entailed by this difference<sup>7</sup>). But they are identical in all other respects. Therefore, among the plate-dropping-worlds most similar to actuality, there are some where it flies off sideways, which means that, according to the orthodox Stalnaker-Lewis theory, it is not true that the plate would have fallen to the floor if dropped.

It may be tempting to argue that the chance of the plate flying off sideways is so small that we can justifiably ignore it in most contexts (see e.g. Lewis 2016). But, on the face of it at least, it seems that treating low probability counterfactual events as being false would also have counterintuitive consequences (Hájek *ms*). For any particular point on the floor, the chance that the midpoint of the plate first touches that point if dropped is incredibly small (apparently zero). But since that is true of any point, it seems to follow, from the above type of reasoning, that the plate would not have landed on *any* point if dropped.<sup>8</sup> Hence, treating low probability counterfactual events as false does not ensure the truth of the counterfactual (1): “If you had dropped the plate it would have fallen to the floor.”

In sum, unless we have an argument against what seems to be the most widespread the same is true of most of the counterfactuals that I have been discussing so far. Nevertheless, Stalnaker’s version of the similarity semantics entails that most ordinary counterfactuals are not true, by the argument under discussion, which seems counterintuitive enough, for ordinary speakers do take counterfactuals like (1) to be true.

<sup>7</sup>Examples of such caused or entailed differences include: the pattern of photons reflected off the plate, the location of the gravitational field around the plate, where the plate is located, and so on. (Thanks to a referee for *Erkenntnis* for encouraging me to clarify this.)

<sup>8</sup>This argument relies on a principle that is often called Agglomeration, which says that if  $A \Box \rightarrow B_i$ ,  $A \Box \rightarrow B_j$ , ..., are all true, then  $A \Box \rightarrow B_i \wedge B_j \wedge \dots$  is true. One might try to avoid this counterintuitive consequence by arguing that the principle is only satisfied within a context (see e.g. Lewis 2016), and that the context shifts once one considers all the consequents together. For a different treatment of Agglomeration, see Leitgeb (2012a, 2012b).

interpretation of a physical theory that almost all physicists accept, then it may seem that we have no choice but to accept that most ordinary counterfactuals are false.

## 2.2 The challenge from indeterminacy

The challenge from indeterminacy is based on the observation that the antecedents of ordinary counterfactuals are generally too underspecified to ensure the truth of their consequents. For instance, it seems true that (3), ‘If you had dropped the plate you might have caught it before it reached the floor,’ since that you drop the plate does not rule out that you drop it just in order to catch it. But the truth of (3) seems to make it false that had you dropped the plate it would have fallen to the floor.

According to Lewis’ theory, we are simply mistaken in thinking that (3) is true on what seems to be the sentence’s most natural interpretation. We could make the sentence come out true by interpreting it as having an epistemic modal in the consequent (as Lewis 1973 briefly considered). But that does not seem to be the only interpretation under which the sentence is true. I take it that my partner was free to choose whether she dropped the plate just to try to catch it again or to have it fall to the floor. Surely in the former case, there was a physical possibility that she would catch the plate. Therefore, I contend, (3) is true even if we interpret it as expressing a physical (but counterfactual) possibility. A similar argument can be made against Stalnaker’s suggestion (which I discuss in more detail below), that we read any ‘might’ counterfactual as a ‘would’ counterfactual with an epistemic modal operator in front of it.

## 2.3 Might and woulds

So far I have been appealing to the intuitive clash between a ‘would’ counterfactual and the corresponding ‘might-not’ counterfactual. The most influential treatment of ‘might’ counterfactuals—namely, David Lewis’ (1973)—takes this intuitive clash as given and defines a ‘might-not’ counterfactual as the negation of the corresponding ‘would’ counterfactual:

**Might 1** (Lewis).  $A \diamond \rightarrow \neg B \doteq \neg(A \square \rightarrow B)$

where  $A \diamond \rightarrow \neg B$  means that if  $A$  were the case then  $B$  might not have been. In other words, Lewis defines ‘might’ counterfactuals (‘mights’ for short) such that if the plate might not have fallen to the floor if dropped, then it is false that it would have fallen to the floor if dropped.

If Lewis’ definition accurately captures the relationship between ‘mights’ and ‘would’ counterfactuals (‘woulds’ for short), then the truth of ‘mights’ like those discussed above does indeed show that most ordinary ‘woulds’ are false. And intuitively, his definition does seem to capture something important about this relationship. After all, the following utterance certainly sounds strange, if not straightforwardly inconsistent:

If you had dropped the plate, it would have fallen to the floor; however, had  
you dropped the plate, it might not have fallen to the floor. (5)

Is the proposition expressed by (5) inconsistent? If it is, then Lewis’ definition is appropriate. However, it could be argued that while one cannot appropriately *utter* (5), it does express a consistent proposition. Stalnaker (1984) and DeRose (1999), for instance, argue that ‘might’ counterfactuals express epistemic possibilities and suggest we define these as:

**Might 2** (Stalnaker-DeRose).  $A \diamond \rightarrow \neg B \doteq \diamond_e \neg(A \square \rightarrow B)$

where  $\diamond_e(C)$  means it is *epistemically possible* that  $C$  (according to some person).

If this definition captures the relationship between ‘mights’ and ‘woulds’, then the truth of the ‘might’ counterfactuals discussed so far does not undermine the truth of the corresponding ‘would’ counterfactuals. For instance, that the plate would have fallen to the floor if dropped is perfectly consistent with me (the speaker) taking it as an epistemic possibility that it would not have done so. Nevertheless, this definition is consistent with the thought that I could not appropriately utter a sentence like (5), since by doing so I am, on this view, simultaneously representing myself as knowing (1) to be true and taking the falsity of (1) to be an epistemic possibility.

There are at least two problems with this solution to the skeptical problem. First, some of the ‘mights’ that seem to undermine the truth of ordinary ‘would’ counterfactuals are not just epistemic possibilities. For instance, the strange quantum events we have been



discussing are *physically* possible. So (2), ‘If you had dropped the plate it might have flown off sideways’, expresses a physical possibility, contra the Stalnaker-DeRose theory, according to which such ‘mights’ express *merely* epistemic possibilities.

Second, even if we set aside the first problem, and accept the Stalnaker-DeRose theory of ‘mights’, we are left with the problem that both indeterminism and indeterminacy seem to make ordinary counterfactuals false on the classical Stalnaker-Lewis similarity semantics. For instance, even if we ignore the intuition that (2), ‘If you had dropped the plate it might have flown off sideways,’ is inconsistent with (1), ‘If you had dropped the plate it would have fallen to the floor,’ we are left with the problem that amongst the worlds where you drop the plate that are most similar to actuality there are some worlds where the plate flies off sideways (assuming indeterminism and an intuitive understanding of ‘similarity’). Therefore, adding the Stalnaker-DeRose theory of ‘mights’ to the similarity semantics is not enough to make (1) come out true.

In Section 4 I defend a theory of counterfactuals that differs from the Stalnaker-Lewis theory in that it can explain how ordinary counterfactuals like (1) can be true, despite the possibilities entailed by quantum mechanics and notwithstanding their unspecified antecedents. In Section 4.3 I suggest an interpretation of ‘mights’ for the proposed theory, and argue that it does better than the Stalnaker-DeRose definition. But first, I will provide some preliminary reasons in favour of the existence of primitive counterfactuals.

### 3 Primitive counterfactuals and CEM

It seems to me that one of these sentences must be true:

- (★) If Bizet and Verdi had been compatriots, then both would have been French.
- (†) If Bizet and Verdi had been compatriots, then they would not have been both French.

But now consider what could make either (★) or (†) true. It seems clear that no fact that we could actually discover would determine which of these is true. Moreover, many people

have the intuition, which I will not argue against, that there is no fact that we could even in principle discover which makes either counterfactual true. In other words, the facts of our world simply do not determine whether (★) or (†) is true. Some philosophers have argued that since the truth of neither (★) nor (†) is entailed by what is actually the case, both sentences are false. So these philosophers, most famously David Lewis, have denied a principle known as *Conditional Excluded Middle* (CEM), which states that for any propositions  $A$  and  $B$ , either  $A \Box \rightarrow B$  or  $A \Box \rightarrow \neg B$ .<sup>9,10</sup>

In contrast, my (and many others’) intuition remains strong that had Bizet and Verdi been compatriots, then both would have been French, or they would not have been both French had they been compatriots. Moreover, various theoretical arguments can and have been given in favour of CEM. For instance, the principle is needed to explain the validity of certain inferences involving quantified counterfactuals (see e.g. Williams 2010). Furthermore, rational confidence judgements in counterfactuals seem to validate CEM. Consider for instance the interaction between our confidence in the counterfactual ‘if we were to toss the coin it would come up heads’ and ‘if we were to toss the coin it would come up tails’.

Nevertheless, it is hard to deny that there is no principled way to decide whether Bizet and Verdi would both have been French or not had they been compatriots. I suggest that we make this observation compatible with the truth of CEM by postulating primitive modal facts that serve as truth-makers for counterfactuals. So either it is a fact that both would have been French had they been compatriots or it is a fact that not both would have been French had they been compatriots. But since these facts are not entailed by the ordinary, non-modal facts, we can accept the intuition that there is no fact we could in principle discover that would tell us which disjunct is true. In the next section I will present a formal model that appeals to such facts, and explain why I take them to provide the basis for a response to counterfactual skepticism.

---

<sup>9</sup>Although some people have taken the Bizet-Verdi example to undermine CEM, and cite Lewis as authority (see e.g. Joyce 1999), a close reading of Lewis’ discussion of the example shows that he took it to provide a counterintuitive implication of his theory. In fact, he went so far as admitting that what his theory entails about the Bizet-Verdi example “sound[s] like a contradiction” (1973: 80)!

<sup>10</sup>I will assume a standard (two valued) logic throughout this paper. Hence, it follows from CEM—for any  $A$ ,  $B$ :  $A \Box \rightarrow B \vee A \Box \rightarrow \neg B$ —that (at least) one of  $A \Box \rightarrow B$  and  $A \Box \rightarrow \neg B$  is true. (I thank XY for encouraging me to make this assumption explicit.)

Some authors, in particular Alan Hájek, have argued that chanciness and indeterminism provide counterexamples to Conditional Excluded Middle. Hájek cites a story from Robert Stalnaker (1984: 164-5)—who famously defended CEM (e.g. in Stalnaker 1981, 1984)—in support of this view:

Consider the following two contrasting stories:

(1) Tweedledee and Tweedledum tossed a fair coin, but before they could see how it landed someone picked it up and ran away with it. Tweedledee is convinced that it landed heads, Tweedledum that it landed tails. Neither has any reason for his belief, but each still feels quite certain. Neither belief is justified, but one of them—we will never know which—is surely correct.

(2) This time someone ran off with the coin before it was tossed. Having no other coin, Tweedledee and Tweedledum argue about how it would have landed if it had been flipped. Tweedledee is convinced that it would have landed heads, Tweedledum that it would have landed tails. Again, neither has a reason—they agree that the coin was a normal one and that the toss would have been fair.

This time, there is little inclination to say that one of them must be right. Unless there is a story to be told about a fact that renders one or the other of the counterfactuals true, we will say that neither is.

As indicated, I will tell a story about a type of fact that could render true counterfactuals like those in the above example. Nevertheless, I think that Stalnaker and Hájek are right in finding both Tweedledee's and Tweedledum's insistence that they know how the coin would have landed to be unreasonable. For according to the story I will tell, the fact that determines who among them is right is an unobservable modality. Moreover, in the case under consideration, due to the fairness of the coin, neither could reasonably (assuming that they are aware of the coin's fairness) cite any evidence in favour of their view. In Section 4.4, I will argue that the same is not true of all counterfactuals.

However, the absence of an observable fact that could determine whether Tweedledee or Tweedledum is right in the latter part of the above example does not mean that neither of

them *is* right. Suppose that Tweedledee and Tweedledum are engaged in a heated argument about whether or not they are brains in a vat. In that case there would also be no observable fact that could determine who is right. However, (almost) everyone would agree that in that case, one of them is right. The problem, if epistemic skeptics are correct, is just that we cannot know who of the two is right.

So, I contend, either the coin would have landed heads if tossed or it would not have landed heads if tossed. But if the coin is unbiased and the toss is truly chancy, then there is no ordinary, non-modal fact that determines which way it would have landed. Instead, I suggest, there is a primitive counterfactual, that is not entailed by (nor supervenes on) the ordinary facts, but is part of the fundamental structure of reality.

I should emphasise that while I think the presumption of primitive counterfactuals explains why Conditional Excluded Middle holds despite examples like those of Bizet and Verdi and Tweedledum and Tweedledee, CEM is not needed for my response to counterfactual skepticism.<sup>11</sup> The vindication of CEM would provide a very strong response to counterfactual skepticism, since it would show that for any propositions  $A$  and  $B$ , it is either true that  $A \Box \rightarrow B$  or it is true that  $A \Box \rightarrow \neg B$ . However, the presumption of primitive counterfactuals does, even without CEM, provide a potential response to counterfactual scepticism. For as we shall see, primitive counterfactuals make it possible that *some* ordinary (and unhedged; see fn. 1) counterfactuals are true, in spite of the arguments from indeterminism and indeterminacy.

## 4 A new solution

According to the theory of counterfactuals that I will develop in this section, both ordinary, non-modal facts and counterfactuals are primitive and part of the fundamental structure of reality. To make sure the distinction between these types of facts is clear: Ordinary (non-modal) facts concern what *actually* happens or is the case, for instance, the fact that grass is green, the fact that I am now writing a paper on counterfactuals and the fact that it is

---

<sup>11</sup>In fact, we could generalise the multidimensional semantics, which I present in the next section, to allow for failure of CEM, by not insisting that unique possible worlds (always) represent how things would be under the various suppositions.

currently not raining. (Occasionally, when the context allows, I will simply refer to these as *facts*.) The counterfactuals, on the other hand, concern what *would*, *could* or *might* have happened, for instance, the fact that had I missed the bus this morning I would have been late for a meeting, the fact that I could have decided to skip the meeting, and the fact that had I done so the people I was meeting with might have been cross with me.

In addition to explaining how Conditional Excluded Middle can be true, despite counterfactuals like those discussed in last section (where, recall, it seems that neither disjunct is made true by the non-modal facts), primitive counterfactuals provide a potential response to counterfactual skepticism.<sup>12</sup> For such primitive modal facts could make it true that the plate would have fallen to the floor if dropped, even though (given indeterminism) the non-modal facts are consistent with the plate having flown off sideways if dropped. Similarly, while someone dropping the plate does not, say, rule out that they catch it again, it could, given the existence of primitive counterfactuals, be true that had they dropped the plate they would not have caught it. In other words, if counterfactual claims are made true or false by primitive counterfactuals, then the arguments from indeterminacy and indeterminism do not suffice to show that they are all false.<sup>13</sup>

The idea that counterfactuals are primitive truth-makers for counterfactual claims raises the pressing question of whether *all* counterfactuals have such primitive and non-modal truth-makers. For instance, if one accepts a principle that is often called *Strong Centring*—which is implied by what Lewis (1973) calls a (strongly) *centered system of spheres*—then  $A \wedge B$  entails  $A \Box \rightarrow B$ .<sup>14</sup> But then if  $A \Box \rightarrow B$  is true in virtue of  $A \wedge B$  being true, is this counterfactual really also made true by some primitive counterfactual? Similarly, if  $T$  denotes a tautology, then it might seem that no counterfactual is needed to make  $A \Box \rightarrow T$  true.

I see three ways of responding to the above issue, all of which are consistent with my

---

<sup>12</sup>Another benefit of postulating primitive counterfactuals, is that it makes sense of statements like ‘I don’t know what would have happened had this or that been the case’. As a referee for *Erkenntnis* points out, people often make such statements, which suggests that they are committed to there being a fact of the matter that they are ignorant of.

<sup>13</sup>Note that this theory of counterfactuals does not ensure that the counterfactuals in question are true, nor tell us which counterfactuals are true. In other words, the theory satisfies what Sarah Moss (2012) calls *semantic humility*: it delivers the truth *conditions* for counterfactuals without telling us the truth *values* of counterfactuals like those discussed throughout this paper.

<sup>14</sup>See Walters (2009) and Walters and Williams (2013) for some arguments for Strong Centring.

general proposal. First, one could say that all counterfactuals have primitive counterfactuals as truth-makers,<sup>15</sup> and simply deny that no counterfactual is needed to make  $A \Box \rightarrow B$  true when  $A \wedge B$  is true (and similarly deny that no counterfactual is needed to make  $A \Box \rightarrow T$  true). Second, one could hold that all counterfactuals have primitive counterfactuals as truth-makers and moreover hold that the truth of the above two counterfactuals is *overdetermined*. For instance,  $A \Box \rightarrow B$  is true in virtue of some counterfactual and in virtue of the fact that  $A \wedge B$  is true. Finally, one could say that only some counterfactuals have primitive counterfactuals as truth-makers. This third response raises the question which counterfactuals have primitive counterfactuals as truth-makers and which do not. Fortunately, Hájek (ms) provides us with a very natural answer to this question: The ordinary and commonly accepted counterfactuals that Hájek argues are false are made either true or false by primitive counterfactuals, whereas those non-ordinary counterfactuals that even Hájek accepts as true (see fn. 1) do not have such primitive modal truth-makers.<sup>16</sup> This third suggestion also raises the question of *why* only some counterfactuals have primitive counterfactuals as truth-makers. Responding to this question might be the hardest problem we would have to solve to make this third suggestion work. In any case, since all three responses are consistent with my proposal, and I have no strong argument in favour of one over the other, I will remain agnostic between them.

#### 4.1 Multidimensional semantics

Before discussing in more detail the metaphysics of my response to counterfactual skepticism, let me briefly discuss the semantics on which it is based.<sup>17</sup>

---

<sup>15</sup>This response raises another pressing issue, namely, how one can have evidence for the truth of any counterfactuals. I address this issue in Section 4.4.

<sup>16</sup>Counterfactuals with qualifiers such as ‘probably’ or ‘almost certainly’ in their consequent, which Hájek (ms) admits can be both true and ordinary, have primitive truth-makers on this view, since they are made true by the high chance of their consequent in some counterfactual world.

<sup>17</sup>It might be worth emphasising that I do not take this semantics to provide independent, direct evidence in favour of my thesis that there are primitive counterfactuals. (I thank a referee for *Erkenntnis* for pressing me on this issue.) The semantics assumes that there are primitive counterfactuals but should not be taken to be an argument in favour of such facts. However, it might be worth pointing out that this semantics has certain benefits that are independent of its potential response to counterfactual skepticism. For instance, it entails (given rather weak assumptions) Stalnaker’s thesis for indicatives and (a version of) Skyrms’ thesis for counterfactuals while avoiding the triviality results by David Lewis (1976) and others (see Bradley 2012 and Stefánsson 2014 for details). Moreover, the semantics has proven very useful in extending decision theory to counterfactual prospects, in a way that avoids various well-known ‘paradoxes’ of expected utility theory that result from the

Just as the theory of counterfactuals I endorse maintains a separation between facts and counterfactuals, so the semantics on which it is based maintains that there are different truth-makers for factual claims and counterfactual ones. Ordinary possible worlds (or, strictly speaking, the ordinary facts they represent) are truth-makers for factual claims. For instance, the claim

John spent the Christmas of 2013 in prison. (6)

is made true by ‘possibly actual’ worlds where John did spend Christmas 2013 in prison. However, possibly *counterfactual* worlds (or, strictly speaking, the counterfactuals they represent) are truth-makers for counterfactual claims. For instance, the claim

John wouldn’t have had to spend Christmas 2013 in prison if he hadn’t bragged  
about the robbery at the pub. (7)

is made true by counterfactual worlds where John didn’t brag at the pub and didn’t spend the Christmas of 2013 in prison.

Moreover, just as (6) is true if John spent Christmas 2013 in prison in the actual world—i.e., the possible world that represents how things actually are—so (7) is true if John didn’t spend Christmas 2013 in prison in the world that is counterfactual under *the supposition that John didn’t brag at the pub*—i.e., the possible world that represents how things would be under that supposition.

To illustrate this thesis, consider a simple model based on the set  $W = \{w_1, w_2, w_3, w_4\}$  of four possible worlds and the corresponding set of its subsets, including the propositions  $A = \{w_1, w_2, w_3\}$ ,  $\neg A = \{w_4\}$ ,  $B = \{w_1, w_2, w_4\}$  and  $\neg B = \{w_3\}$ . Relative to the set of possible worlds  $W$ , a supposition induces a set of possible counterfactual worlds. The supposition that  $A$ , for instance, induces the set of counterfactual  $A$ -worlds:  $W_A = \{w_1, w_2, w_3\}$ . To keep things simple, let’s restrict our attention to a single supposition, the supposition that  $A$ . The

---

fact that reasonable decision-makers are often influenced by what could have been in their evaluation of the desirability of what actually occurs (Stefánsson 2015, Bradley and Stefánsson 2017). Hence, the fact that my response to counterfactual skepticism can be formulated using a formal semantics that has proven successful in dealing with other problems, could be taken to provide some *indirect* evidence for my proposal.

set of elementary possibilities is then given by a subset of the cross-product of  $W$  and  $W_A$ , as shown in Table 1.<sup>18</sup>

Worlds	Supposed $A$ -worlds		
	$w_1$	$w_2$	$w_3$
$w_1$	$\langle w_1, w_1 \rangle$	$\langle w_1, w_2 \rangle$	$\langle w_1, w_3 \rangle$
$w_2$	$\langle w_2, w_1 \rangle$	$\langle w_2, w_2 \rangle$	$\langle w_2, w_3 \rangle$
$w_3$	$\langle w_3, w_1 \rangle$	$\langle w_3, w_2 \rangle$	$\langle w_3, w_3 \rangle$
$w_4$	$\langle w_4, w_1 \rangle$	$\langle w_4, w_2 \rangle$	$\langle w_4, w_3 \rangle$

TABLE 1: POSSIBILITY SPACE

Each ordered pair  $\langle w_i, w_j \rangle$  represents an elementary possibility: that  $w_i$  is the actual world and  $w_j$  the counterfactual  $A$ -world. Sets of such possibilities are propositions. Factual propositions are given by rows of the table. The proposition that  $A$ , for instance, is the set of possibilities in the first, second and third row of the table, since  $A$  is true at the first world of each pair in these rows. Conditional propositions, on the other hand, are given by columns of the table. The proposition that if  $A$  then  $B$ , for instance, is the set of possibilities in the first and second column of the table, since  $B$  is true at the second world of each pair in these columns. Conjunctions, disjunctions and negations of propositions are given by their intersections, unions and complements.

The model can easily be generalised to multiple suppositions and a greater number of possible worlds. Instead of pairs representing the elementary possibilities,  $n$ -tuples (i.e., ordered sets of worlds) would then represent these elementary possibilities:  $\langle w_i, w_j, w_k, w_l, \dots, w_n \rangle$ . For instance, we could stipulate that the first world in each tuple denotes a possibly actual world, the second world a possibly counterfactual world under the supposition that  $A$ , the third world a possibly counterfactual world under the supposition that  $\neg A$ , the fourth world a possibly counterfactual world under the supposition that  $B$ , and so on. And then proposi-

<sup>18</sup>In this framework, centring requires that if  $w$  is actual and  $A$  is true at  $w$ , then  $w$  is also counterfactual under the supposition that  $A$ . Adding this constraint to the semantics amounts to removing a few of the pairs in Table 1. For instance, if centring holds, then  $\langle w_1, w_2 \rangle$  is not a possibility, nor is  $\langle w_2, w_1 \rangle$ , and so on. For a further discussion of centring in relation to the multidimensional semantics, see Stefánsson (2014).



tions are sets of  $n$ -tuples; for instance, the factual proposition  $A$  is the set of  $n$ -tuples where  $A$  is true at the first world, the conditional  $A$  then  $B$  the set of  $n$ -tuples where  $B$  is true at the second world, and so on.

Let's however focus on the simple 4-world model. Coming back to John and his time spent in prison, let  $A$  be the proposition that John brags about his robbery at the pub and  $B$  the proposition that John spends Christmas 2013 in prison. Then if any of the pairs in the third row in Table 1 represent how things actually are and would have been if  $A$ , then it is true that John did *not* spend Christmas 2013 in prison (since  $\neg B$  is true at the first element of these pairs). However, if any of the pairs in first two columns correctly represent how things are and would have been if  $A$ , then it is true that *had* John bragged about his robbery at the pub he *would have* spent Christmas 2013 in prison (since  $B$  is true at the second member of these pairs).

## 4.2 Realism vs. Humean supervenience

Recall that we seemed forced to conclude from the arguments from indeterminism and indeterminacy that it is false that (1) the plate would have fallen to the floor if dropped. However, the multidimensional semantics can make room for the truth of this counterfactual notwithstanding the two challenges. On this semantics, it is true that had my partner dropped the plate it would have fallen to the floor just in case the plate falls to the floor in the world that is counterfactual under the supposition that she drops the plate. Whether or not this is true at that particular world is not determined by a similarity ordering of possible worlds. So (1) can be true even though amongst the plate-dropping-worlds most similar to actuality there are some worlds where the plate flies off sideways. Nor is the truth of counterfactuals determined by the ordinary (non-modal) facts of the actual world. For instance, in the simple 4-world model in Table 1, the second elements of each world-pair serve as truth-makers for counterfactuals whose antecedent is  $A$ , by determining what would be true under the supposition that  $A$ . And as the model illustrates, the same potentially actual world can be paired up with different potentially counterfactual worlds. In other words,

Humean supervenience—the thesis that all truths supervene on non-modal truths—is not assumed, which means that (1) can be true even though the (non-modal) facts of the actual world do not suffice to determine that the plate would have fallen to the floor if dropped.

The multidimensional semantics is flexible enough to accommodate different metaphysical views, such as those concerning the realism or not of the counterfactuals. For instance, we could take them to be simply a convenient way of representing a person's hypothetical beliefs rather than something that is really a person-independent feature of reality. Similarly, a Humean supervenience constraint can consistently be added to the framework (see Stefánsson 2014). I want to suggest, however, that the counterfactuals are real, and are, just like the ordinary facts, part of the fundamental structure of reality but not entailed by the ordinary facts. These are really two theses: on one hand, realism about counterfactuals, and, on the other, anti-Humeanism about counterfactuals. But as I explain below, I think that it is hard to maintain the first thesis without accepting the latter.

Let's first get clear on what realism about counterfactuals entails. There is a fact of the matter as to whether John did or did not spend Christmas 2013 in prison. (If you think there is some vagueness concerning what counts as Christmas, then change this to: There is a fact of the matter as to whether John did or did not spend every minute of December 25<sup>th</sup> 2013 in prison.) Denying this would amount to a radical anti-realism that, I think, nobody endorses.<sup>19</sup> Similarly, I am suggesting that there is a fact of the matter as to whether John would or wouldn't have spent Christmas 2013 in prison if he had (and if he hadn't) bragged about his robbery at the pub.<sup>20</sup> Moreover, just as there is a fact of the matter as to whether my partner did or did not drop the plate, so there is a fact of the matter as to whether or not the plate would have fallen to the floor had she dropped it.

Why accept realism of this type? Intuitively, it seems that the type of counterfactuals

---

<sup>19</sup>The idea that all there 'really' is are atoms and their spatiotemporal arrangements doesn't mean denying that there is a fact of the matter as to whether John did or didn't spend Christmas 2013 in prison, since even on such reductive views, it either is or isn't a fact that the atoms are (or were) arranged such that the macro event in question supervenes on them.

<sup>20</sup>As a referee for *Erkenntnis* points out, this suggestion is only plausible if we think that there is a (single) fact of the matter as to *how* John would have bragged had he done so (e.g. whether he whispered it to a discreet friend or shouted it so that everyone could hear). Some will find that implausible, but it is a bullet I am willing to bite, especially since any response to counterfactual skepticism will have some counterintuitive features (as I explained in the introduction).

that we use so widely in various types of reasoning are propositions with truth values, as opposed to e.g. expressions of our belief revision policies (as e.g. Edgington 2008 holds). For instance, it is hard to understand how counterfactuals can have probabilities if they lack truth values (Hájek 2014). Moreover, notwithstanding the challenge from indeterminism and indeterminacy, it seems equally intuitive that our beliefs in, or acceptance of, counterfactuals should aim at the *truth*,<sup>21</sup> as opposed to merely acceptability, assertability, probability, etc. I want to take these intuitions at face value. But to do so we need to assume realism about those counterfactuals that serve as truth-makers for the counterfactuals we accept.

Given the multidimensional semantics, the realism I am suggesting means that for any proposition  $A$  and world  $w_i$ , there is a fact of the matter as to whether  $w_i$  is or is not counterfactual under the supposition that  $A$ . Perhaps the vocabulary is a bit misleading, since ‘supposition’ is usually used to describe a hypothetical attitude of some agent rather than what would really be the case if some condition held. However, I am suggesting that there is a type of fact, a *counterfact*, that determines whether or not  $w_i$  is counterfactual under the supposition that  $A$ .

Let’s see then how this leads to anti-Humeanism. Suppose now that John actually bragged about the robbery at the pub and as a result ended up in prison. It seems implausible that whether or not John would have gone to prison if he hadn’t bragged at the pub is completely determined by the ordinary (non-modal) facts of the actual world.<sup>22</sup> After all, the antecedent is consistent with him bragging about it somewhere else, and also consistent with him deciding to keep his mouth shut. Either way, whether he would have been put in prison if he hadn’t bragged is not determined by the (non-modal) facts. Similarly, since my partner didn’t drop the plate, and since the world is apparently indeterministic, the (non-modal) facts of the world don’t determine whether the plate would or would not have

---

<sup>21</sup>Just as one can aim at doing what is right even without direct access to moral, or deontic, facts, I think one can aim at believing the truth about counterfactuals even though one has no direct access to counterfactuals.

<sup>22</sup>Someone like Lewis, who was committed to both a similarity semantics and Humean supervenience, might disagree (as a referee for *Erkenntnis* points out), and for instance argue that the actual world fully determines which of the following two worlds is more similar to actuality: one where John doesn’t brag about the robbery and goes to prison or one where he doesn’t brag about the robbery and manages to stay out of prison. Perhaps such an argument can be made. However, it seems clear that we would then need to employ some rather counterintuitive conception of similarity (contrary to Lewis’ initial hope).

flown off sideways had she dropped it.

Now the thesis of Humean supervenience holds that all true counterfactuals are entailed by (and thus supervene on) the (non-modal) facts. But these latter facts do not determine whether John would or wouldn't have gone to prison if he hadn't bragged about his robbery at the pub. Therefore, if we accept Humean supervenience, it seems we should either say that it is false that John would have been caught, and also false that he wouldn't have been caught, if he hadn't bragged at the pub (as Lewis 1973 might do), or we should say that it is indeterminate whether he would or wouldn't have been caught if he hadn't bragged at the pub (as Stalnaker 1984 might do). In contrast, a natural way to make room for the idea that one of these counterfactuals is true and the other is false is to deny Humean supervenience, I think, since this truth is not entailed by the non-modal facts.

Similarly, it seems that a natural way to make room for the possibility that it is true that the plate would have fallen to the floor had my partner dropped it and false that it would have flown off sideways, is to deny Humean supervenience, since due to the indeterminism of our world, the (non-modal) facts neither entail that the plate would or wouldn't have flown off sideways.<sup>23</sup> Moreover, these facts do not determine precisely *how* she would have dropped the plate had she done so; in particular, whether or not she would have dropped it with one hand just to catch it with the other. So again, if we want to say that it is true that the plate would have fallen to the floor if dropped, then we should, I think, deny Humean supervenience.

Perhaps due to the influence of David Lewis, many people seem to take the Humean supervenience thesis to be sacred. However, I think that inconsistency with the thesis should not count too heavily against my proposal. I will only mention one reason why. A number of people have argued that Humean supervenience is inconsistent with quantum mechanics, which, at least on one popular understanding of it, seems to postulate the existence of

---

<sup>23</sup>As a referee for *Erkenntnis* points out, there might be ways of making the truth of the plate-counterfactual (1) consistent with Humean supervenience; for instance, by offering a (plausible) Humean supervenience rendering of Lewis' (1979) idea of quasi-miracles. While I grant that perhaps some such argument can be made, I must admit that I do not find it very plausible. In particular, I do not see how such a response could be made to the argument from indeterminacy, since there does not seem to be anything miraculous about, say, my partner catching the plate after dropping it.

irreducible modalities (see e.g. Teller 1986, Maudlin 2007, Schaffer 2010 and Weatherson 2015). Lewis (1987) was well aware of the problem, but claimed that there was still an interest in defending Humean supervenience against *philosophical* (as opposed to empirical or scientific) arguments. Moreover, not everyone agrees that the Humean supervenience thesis is inconsistent with quantum mechanics (see e.g. Miller 2013 and Esfeld 2014). I will not attempt to enter this debate, but simply contend that as long as it is unclear whether modern physics is consistent with the Humean supervenience thesis, I need not be too worried by the fact that my theory of counterfactuals violates the thesis.

To sum up: By driving a wedge between facts and counterfactuals, the theory of counterfactuals I am proposing makes room for the possibility that counterfactuals can be true even when their truth is not determined by the ordinary, non-modal facts, and even though their antecedents are not specific enough to ensure the truth of their consequents.

### 4.3 Might and would again

So, the theory of counterfactuals I am suggesting is consistent with the truth of the counterfactuals we tend to accept. But what does that mean for the corresponding ‘might-not’ counterfactuals? It seems true that (2) the plate might have flown off sideways if it had been dropped, and, similarly, that (3) my partner might have caught it before it fell to the floor. Therefore, it seems the plate might not have fallen to the floor if it had been dropped. (And similarly for other ordinary counterfactuals that we take to be true.) But isn’t this inconsistent with the claim that (1) the plate *would* have fallen to the floor if dropped?

I have already rejected the Stalnaker-DeRose view that the ‘mights’ we have been discussing express merely epistemic possibilities, since (2), for instance, expresses a physical (but counterfactual) possibility. So I cannot argue that (1) and (2) can both be true on the grounds that the first expresses how things actually would be but the second expresses that some agent believes it to be possible that (1) is false. However, I think the failure of supervenience in the multidimensional model provides an explanation of how (1), (2) and (3) can be simultaneously true, and, more generally, makes room for a theory of ‘mights’ that deals

well with the problems raised by the arguments for counterfactual skepticism. Whether it is all-things-considered a good theory of ‘mights’ is another question, that falls outside the scope of this paper.

I want to suggest that sometimes a ‘might’ counterfactual,  $A \diamond \rightarrow B$ , expresses that *the (ordinary, non-modal) facts of the actual world don’t determine that B would have been false had A been true.*<sup>24</sup> So for instance, it is true that (2) ‘If you had dropped the plate it might have flown off sideways,’ since the ordinary facts do not determine that the plate would not have flown off sideways. More generally: *It might have happened since it is not determined that it would not have happened.* However, on this understanding, it could still be the case—given the failure of Humean supervenience—that the plate *would* have fallen to the floor if dropped. For it is possible that, while the (non-modal) facts of the actual world don’t determine that the plate would have fallen to the floor if dropped, the plate nevertheless does fall to the floor in the world that is counterfactual under the supposition that the plate is dropped.

Let’s compare this with propositions about the future. Given indeterminism, either (almost) no contingent and categorical<sup>25</sup> proposition about the future is true, or some are true despite the corresponding ‘might-not’ being also true. Personally I find the latter view more plausible. So while it is true that when I have lunch an hour from now, my plate might suddenly fly off sideways, the truth probably is that it will not behave this way (where ‘probably’ is subjective). In other words, the (subjectively) probable scenario is: It is true now that the plate will not fly off sideways even though it might do so (where ‘might’ is not just epistemic). Similarly, the plate might have flown off sideways if dropped but probably it would not have.

What should we say then about the *indeterminacy* cases? Recall that it seems true that (3) ‘If you had dropped the plate you might have caught it before it reached the floor.’ And intuitively, that seems inconsistent with (1), because if the plate might have been caught before it fell to the floor then it might not have fallen to the floor. Again, I suggest we interpret the ‘might’ in this case as expressing that the (non-modal) facts of the world don’t

---

<sup>24</sup>This is not to deny that ‘might’ counterfactuals sometimes express epistemic possibilities.

<sup>25</sup>That is, no proposition stating that this or that *will*, as opposed to *might* (or will *probably*), happen.

determine the way in which my partner would have dropped the plate had she done so. Had she dropped the plate she might have caught it before it fell to the floor: *It might have happened since it is not determined that it would not have happened.* However, it could still be the case that she doesn't catch the plate before it falls to the floor in the world that is counterfactual under the supposition that she drops the plate. Therefore, it could be true that she might have caught it but would not have done so. And again, this fits well with how we tend to think about propositions about the future: Probably, it is true that I will not drop my plate at lunch even though I might do so (where, again, the 'might' is not just epistemic but the probability is).

So on this understanding, 'mights' and 'woulds' are not duals, as they are in Lewis' theory. Some may find that to be a drawback of my proposal. Recall the following expression (5) that I discussed in section 2.3:

If you had dropped the plate, it would have fallen to the floor; however, had you dropped the plate, it might not have fallen to the floor.

This expression is certainly odd. Recall that Stalnaker and DeRose think that what is wrong with it is that the speaker represents herself as knowing a proposition while taking it to be an epistemic possibility that the proposition is false. But their theory cannot be quite right, I argued, since the 'might' counterfactual in question expresses more than a mere epistemic possibility. In the next subsection I make a proposal that, like Stalnaker and DeRose's, entails that the tension in (5) is one of pragmatics rather than semantics. But unlike Stalnaker and DeRose's, my suggestion does not require that the 'mights' in question are epistemic. The proposal combines the so-called *knowledge norm of assertion* with a particular contextualist understanding of knowledge.

#### **4.4 Knowledge and assertability**

According to the knowledge norm of assertion, one represents oneself as knowing a proposition when one asserts it. Hence, one should not assert a proposition that one does not know (Williamson 2000). While the idea is not universally accepted, it has proven successful

in explaining Moore-type paradoxes, that concern sentences that are semantically consistent but pragmatically infelicitous. Take the sentence:

It is raining but I don't believe it is. (8)

Sentence (8) is semantically consistent—that is, there is a possible world where it holds true. But the sentence is pragmatically infelicitous. There are (at least) two ways to explain this, given the knowledge norm of assertion. First, since one should not assert 'It is raining' unless one knows it (and knowledge entails belief), the speaker has, if the second part of the sentence is true, violated a norm of assertion, by asserting that it is raining without believing that it is. Second, a person who asserts (8) simultaneously represents herself as knowing (first part of sentence) and as not knowing (second part of sentence) that it is raining.

Hájek (ms: 66) however takes issue with the knowledge norm of assertion:

Arguably, one represents oneself merely as believing something when one asserts it, rather than knowing it. When one really does want to represent oneself as knowing something, explicitly saying so is not redundant: "The pub closes at midnight. Trust me: I know that it does." And it's not the difference of now representing oneself as knowing that one knows—that's too fancy for the folk, who often speak this way (even outside pubs).

Unlike Hájek, I do not think that examples like these undermine the knowledge norm of assertion. If anything, they confirm it. First, it seems to me that what the speaker is doing, when, after asserting a proposition, she adds that she knows it, is to assure the listener that she is following the norms of assertion; and, more generally, adhering to conversational norms. Second, compare the above conversation with one where the speaker, who claims that the pub closes at midnights, adds that she *believes* it does. If Hájek's conversation were a counterexample to the knowledge norm of assertion, then the latter conversation would be a counterexample to the 'believe norm of assertion' that Hájek tentatively endorses. But the second conversation seems to me to provide evidence for the knowledge norm of assertion.



When one adds that one believes a proposition after having asserted it, one (typically) seems to be admitting that one should not categorically assert the proposition, as evidenced by the fact that one would normally add something like ‘*at least I believe it*’ (which suggests that one may not know, but *only* believes, what was asserted). So when adding that one believes a proposition one has asserted, one seems to be hedging one’s assertion, which suggests that one is admitting that one had not completely followed the norms of assertion.

A lot has been written about the knowledge norm of assertion, and I do of course not claim to have provided an ultimate argument for it. But let’s accept it for now. In particular, let’s accept, for the sake of the argument, that when one asserts a counterfactual one represents oneself as knowing it. However, it is worth pointing out that my argument below would also work for something like a ‘high confidence norm of assertion’, if the threshold for ‘high confidence’ is sufficiently context dependent.

But how can one know a counterfactual if its truth-makers are not part of the ordinary, non-modal facts? Below I suggest a simple model of how one could know a counterfactual even if my theory of counterfactuals is correct. This model of knowledge may be too simplistic, but my aim is just to demonstrate that there are at least somewhat plausible, and certainly very traditional, views on knowledge according to which one could know counterfactuals, even though their truth-makers are not part of the ordinary, non-modal facts. On this model, an agent, whose credence is represented by  $Cr$ , knows a counterfactual,  $A \Box \rightarrow B$ , *only if* the following conditions hold:

1.  $Cr(A \Box \rightarrow B) \geq \delta$  for some  $\delta$  close to 1.
2.  $Cr(A \Box \rightarrow B)$  is ‘justified’
3.  $A \Box \rightarrow B$  is true

If we take belief to be credence above a threshold—a view often called the *Lockean Thesis* (Foley 1992)—then the satisfaction of the above three conditions is both necessary and sufficient (for some  $\delta$ ) for knowledge according to the traditional view that knowledge is

justified, true belief.<sup>26</sup> However, many philosophers think that to avoid Gettier-problems (Gettier 1963), a fourth condition is needed for the conditions to collectively suffice for knowledge. In other words, the idea would be that an agent knows  $A \Box \rightarrow B$  whenever she satisfies the above three and some additional condition (e.g. ‘sensitivity’ or ‘safety’) that is, roughly speaking, meant to rule out believing something for the wrong reason.<sup>27</sup> I will not discuss this fourth condition in more detail, nor try to defend a view like this on knowledge. Rather, I will simply suggest it as one traditional view on knowledge according to which one can, as I will argue, know counterfactuals even though their truth-makers are not part of the non-modal facts.<sup>28</sup>

The two most pressing questions raised by the above view on knowledge of counterfactuals, are how to understand the threshold  $\delta$  and what it takes for a credence assignment to a counterfactual to be justified. The only thing I want to say about  $\delta$  is that it should be context dependent and always at least close to 1.

I suggest that a minimum requirement for a credence assignment to a counterfactual to be justified is that it satisfies a version of Skyrms’ thesis, which, as mentioned above (fn. 17), is entailed by the multidimensional semantics (given some rather weak assumptions). Informally, the thesis states that one’s credence assignment to a counterfactual should equal one’s expectation of the conditional chance of the counterfactual’s consequent given its antecedent.<sup>29</sup> More formally, let  $\cup_i\{CH_i\}$  be a partition of *chance propositions*, such that, for

<sup>26</sup>Note however that my arguments works even if we are, say, eliminativists about belief and e.g. take knowledge to consist in having justifiably high credence (perhaps for the right reasons) in a true proposition. Similarly, as already mentioned, my argument works if we ignore knowledge altogether, and replace the knowledge norm of assertion with something like a high confidence norm of assertion.

<sup>27</sup>See Ichikawa and Steup (2014) for an overview and discussion.

<sup>28</sup>On the face of it, it might seem that someone who accepts this type of view on knowledge in general would have to give up Conjunction Closure (and Agglomeration) for knowledge: a person could know  $A$  and  $B$  without knowing  $A \wedge B$  (and she could know  $A \Box \rightarrow B$  and  $A \Box \rightarrow C$  without knowing  $A \Box \rightarrow B \wedge C$ ). Those who are willing to give up Conjunction Closure for belief (such as Foley 1992) and think that belief is a necessary condition for knowledge will not have a problem with this implication of the view under considerations. But many others may find this quite counterintuitive. A possible response is to exploit the context dependence of  $\delta$  to ensure closure. I will have to set these issues aside in this paper.

<sup>29</sup>One might well ask (as a referee for *Erkenntnis* does) *why* there should be this relationship between chance and primitive counterfactuals. The formal (but perhaps unsatisfying) explanation is that once we add to the multidimensional semantics what Bradley (2012) calls the ‘Principal Suppositional Principle’—which says that the credence we assign under the supposition of  $A$  should correspond to our expectation of conditional chance given  $A$ —and in addition the assumption that our credence in chance propositions is independent of *counterfactual* (but not evidential) suppositions, the semantics entails Skyrms’ thesis, and thereby builds a bridge between rational credence in chances and counterfactuals. For these assumptions to be plausible, we need to assume, I think, that

instance proposition  $CH_i$  says that function  $Ch_i$  is correct about the objective chance. I will be agnostic about how to understand chances, other than assuming that they represent an objective feature of the world rather than the uncertainty of some agent, and that they evolve (in the sense that a proposition describing an event that has already taken place or not has either chance 1 or 0), which is needed if we want counterfactuals to satisfy both (this version of) Skyrms' thesis and Modus Ponens.<sup>30,31</sup> Then the thesis in question states that:

**Skyrms' thesis.** *For any  $A, B$ :*

$$Cr(A \Box \rightarrow B) = \sum_i Cr(CH_i) \cdot Ch_i(B | A)$$

The fact that the semantics in question entails Skyrms' thesis explains how we can have evidence<sup>32</sup> for counterfactual claims even when their truth-makers are not part of the non-modal facts: Any method for gathering evidence of (conditional) chance—in particular, statistical hypothesis testing and observing frequencies—can give us evidence of counterfactuals. Similarly, Skyrms' thesis explains how one can have reasons for being confident in the truth of counterfactuals, even when their truth-makers are primitive counterfactuals. To take a concrete example, a sufficiently large number of observations where a plate that is sufficiently similar to plate  $p$  is dropped and falls to the floor, in circumstances that are sufficiently similar to circumstance  $c$ , provides evidence for the claim that plate  $p$  would have fallen to the floor (call this proposition  $A$ ) if dropped in circumstance  $c$  (call this proposition  $B$ ).<sup>33</sup> Therefore, such frequencies provide a reason for being confident that the conditional chance of  $B$  given  $A$  is high, and thus for being confident that  $A \Box \rightarrow B$ . Hence, there is nothing mysterious about the fact that we can have evidence for, and reason for being confident in, counterfactuals, chances are primitive too (just like the counterfactuals).

<sup>30</sup>This means that the thesis I am discussing is, strictly speaking, not Skyrms' thesis, since his thesis referred to 'prior propensities' rather than evolving chances. See e.g. Brian Skyrms (1981).

<sup>31</sup>It might be worth mentioning that I assume conditional chances to be primitive rather than defined in terms of unconditional ones, for instance since I want conditional chances to be defined in some cases where the condition has zero or undefined chance (see Hájek 2003).

<sup>32</sup>By 'evidence' for  $C$  I here simply mean anything that *should* increase one's credence in  $C$ . If an agent learns something that increases her credence that the conditional chance of  $B$  on  $A$  is high, then by Skyrms' thesis, she has learned something that should increase her credence in  $A \Box \rightarrow B$ . Hence, she has gained evidence for  $A \Box \rightarrow B$ .

<sup>33</sup>This is not because finite frequencies *are* chances, but because they often provide good estimates of chances (as e.g. Hájek 1996 points out).

even though their truth-makers are not part of the non-modal, directly observable facts.<sup>34</sup>

So for a person's credence assignment to a counterfactual to be justified it has to equal her expectation of the corresponding conditional chance. Further requirements presumably have to be satisfied as well. For instance, credence functions may need to be sufficiently sensitive to evidence, perhaps they have to be based on reasonable priors, and so on. A lot has been written on what it takes for credences to be justified, and I will remain agnostic about the details, except for requiring that credence in counterfactuals satisfy the above version of Skyrms' thesis.

To sum up, I am suggesting that one should not assert a proposition unless one knows it, and that one knows a counterfactual if one has sufficiently high credence in its truth, if the credence is justified, if the counterfactual is true, and if some fourth 'anti-luck' condition is satisfied. To take an example, consider the counterfactual (1), 'If you had dropped the plate it would have fallen to the floor.' If one's high expectation of the conditional chance of the counterfactual's consequent given its antecedent is grounded in, say, observing the relative frequency with which plates sufficiently similar to the one in question fall to the floor when dropped (in circumstances sufficiently similar to the one in question), then one's expectation might be justified and high for the right reason. But then if one has *sufficiently* high credence in the counterfactual, and the counterfactual is in fact true, then one knows it, according to this view.

How does this help us explain the seeming clash between a 'would' counterfactual and the corresponding 'might-not' counterfactual? I contend that a (true) 'might-not' counterfactual can change the context such that the corresponding 'would' counterfactual becomes unassertable. But note that I do not need to embrace the radical contextualist view (e.g. defended by Lewis 2016) that the 'might-not' changes the meaning and/or truth value of the corresponding 'would'.

---

<sup>34</sup>One might worry that such evidence will never warrant, for any conceivable context, being *certain* that a counterfactual is true. (And some might even take this to show that one cannot know a counterfactual.) All I want to say about this worry is that I do not see the need for certainty. For any contingent proposition, what matters is whether one can justifiably be confident in its truth, not whether one can be certain that it is true.

Recall again the following expression (5):

If you had dropped the plate, it would have fallen to the floor; however, had you dropped the plate, it might not have fallen to the floor.

I am claiming that (5) is semantically consistent. But it is clearly not assertable. There are (at least) two ways to explain this, given the above suggestion.<sup>35</sup> First, it seems plausible that the utterance of the ‘might-not’ makes improbable and strange events salient, and thus changes the context such that the credence threshold  $\delta$  increases. If  $\delta$  increases sufficiently by the acknowledgment of the ‘might-not’, then one no longer *knows* the ‘would’, and hence should not assert it. Perhaps  $\delta$  becomes 1, or arbitrarily close to 1, once the ‘might-not’ has been uttered, which could explain why a sentence like (5) is unassertable.

Secondly, acknowledging a ‘might’ counterfactual could change what chance hypothesis one can justifiably ignore (in a sense to be explained below). Recall from the introduction counterfactuals (1) and (2) respectively:

If you had dropped the plate it would have fallen to the floor.

If you had dropped the plate it might have flown off sideways.

In normal conversational and epistemic contexts, one can arguably be justified in ignoring quantum events like a plate flying off sideways. One way to understand this more formally, is that one can, in such contexts, justifiably leave such events outside the algebra over which one distributes one’s subjective probabilities; the idea being that the algebra in question represents the domain of discourse. And the same could be said about chance hypotheses that give a positive chance to plates flying off sideways conditional on being dropped. But then one can, in normal contexts, be justifiably fully confident in a counterfactual like (1), according to Skyrms’ thesis; which means that the counterfactual is maximally assertable (assuming that it is true) in normal contexts.

However, it is not clear that one can justifiably ignore quantum (and other strange but

---

<sup>35</sup>Note that both explanations also work for a contextualist *high confidence* norm of assertion.

possible) events once they have been introduced. And one certainly cannot justifiably ignore such events while simultaneously acknowledging their possibility. Therefore, one cannot, in a context where one has uttered a counterfactual like (2), leave all chance hypotheses that give a positive chance to plates flying off sideways conditional on being dropped outside the algebra over which one distributes one's subjective probabilities. But by Skyrms' thesis, this entails that in contexts where quantum events have been introduced, one cannot be justifiably fully confident in a counterfactual like (1) (assuming that one should assign the hypotheses in question a positive probability once they have been introduced into the domain of discourse<sup>36</sup>). Hence, in a context where one asserts (2), it seems plausible that one's justified credence in (1) cannot be as high as  $\delta$  (which is assumed to be very close to 1 in such contexts), which would explain why the two counterfactuals are not co-assertable; and would similarly explain why one cannot appropriately assert (5).

In conclusion, I hope to have delivered what was promised. I have argued that the assumption of primitive modal facts provides a potential response to counterfactual skepticism, and I have shown that a recent multidimensional possible world semantics for conditionals, which assumes such primitive modalities, makes such a response plausible. Finally, I have explained how one could appropriately assert and have evidence for counterfactuals, even if their truth-makers are not to be found in the actual world.

## References

Bradley, R. (2012). Multidimensional possible-world semantics for conditionals. *Philosophical Review*, 121(4):539–571.

Bradley, R. and Stefánsson, H. O. (2017). Counterfactual desirability. *British Journal for the*

---

<sup>36</sup>Note that this does not require assuming that Regularity holds in general; i.e., the assumption need not be that one cannot acknowledge *any* possibility without assigning it a positive probability. (I thank a referee for *Erkenntnis* for encouraging me to clarify this point.) In particular, the assumption that one should assign the hypotheses in question a non-zero probability once they have been brought into the domain of discourse, is consistent with the view that one can acknowledge that, say, some infinite sequence of heads/tails combinations is possible while still assigning it a zero probability (see e.g. Williamson 2007).

*Philosophy of Science.*

DeRose, K. (1999). Can it be that it would have been even though it might not have been? *Noûs*, 33(1):385–413.

Edgington, D. (2008). Counterfactuals. *Proceedings of the Aristotelian Society*, 108(1):1–21.

Esfeld, M. (2014). Quantum Humeanism, or: physicalism without properties. *Philosophical Quarterly*, 64(256):453–470.

Foley, R. (1992). The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly*, 29(2):111–121.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23(6):121–123.

Gillies, A. S. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360.

Hájek, A. (1996). “Mises redux” – redux: Fifteen arguments against finite frequentism. *Erkenntnis*, 45(2):209–227.

Hájek, A. (2003). What conditional probability could not be. *Synthese* 137, 137(3):273–323.

Hájek, A. (2014). Probabilities of counterfactuals and counterfactual probabilities. *Journal of Applied Logic*, 12(3):235–252.

Hájek, A. (ms). Most counterfactuals are false. Available at: [http://philosophy.anu.edu.au/sites/default/files/Most%20counterfactuals%20are%20false.1.11.11\\_0.pdf](http://philosophy.anu.edu.au/sites/default/files/Most%20counterfactuals%20are%20false.1.11.11_0.pdf) (last visited on October 12, 2015).

Hawthorne, J. (2005). Chance and counterfactuals. *Philosophy and Phenomenological Research*, 70(2):396–405.

Ichikawa, J. (2011). Quantifiers, knowledge, and counterfactuals. *Philosophy and Phenomenological Research*, 82(2):287–313.

Ichikawa, J. J. and Steup, M. (2014). The analysis of knowledge. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition.

- Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.
- Leitgeb, H. (2012a). A probabilistic semantics for counterfactuals. Part A. *The Review of Symbolic Logic*, 5(1):26–84.
- Leitgeb, H. (2012b). A probabilistic semantics for counterfactuals. Part B. *The Review of Symbolic Logic*, 5(1):85–121.
- Lewis, D. (1973). *Counterfactuals*. Blackwell.
- Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85(3):297–315.
- Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, 13(4):455–476.
- Lewis, D. (1987). Introduction. In *Philosophical Papers, Vol. 2*. Oxford University Press.
- Lewis, K. S. (2016). Elusive counterfactuals. *Noûs*, 50(2):286–313.
- Maudlin, T. (2007). *The Metaphysics Within Physics*. Oxford University Press.
- Miller, E. (2013). Quantum entanglement, Bohmian mechanics, and Humean supervenience. *Australasian Journal of Philosophy*, 92(3):1–17.
- Moss, S. (2012). On the pragmatics of counterfactuals. *Noûs*, 46(3):561–586.
- Moss, S. (2013). Subjunctive credences and semantic humility. *Philosophy and Phenomenological Research*, 87(2):251–278.
- Schaffer, J. (2010). Monism: The priority of the whole. *Philosophical Review*, 119(1):31–76.
- Schulz, M. (2014). Counterfactuals and arbitrariness. *Mind*, 123(492):1021–1055.
- Skyrms, B. (1981). The prior propensity account of subjunctive conditionals. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Ifs: Conditionals, Belief, Decision, Chance, and Time*. D. Reidel Publishing Company.



- Stalnaker, R. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory*. Blackwell.
- Stalnaker, R. (1981). A defence of conditional excluded middle. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Iffs: Conditionals, Belief, Decision, Chance, and Time*. D. Reidel Publishing Company.
- Stalnaker, R. (1984). *Inquiry*. MIT Press.
- Stefánsson, H. O. (2014). Humean supervenience and multidimensional semantics. *Erkenntnis*, 79(6):1391–1406.
- Stefánsson, H. O. (2015). Fair chance and modal consequentialism. *Economics and Philosophy*, 33(1):371–395.
- Teller, P. (1986). Relational holism and quantum mechanics. *British Journal for the Philosophy of Science*, 37(1):71–81.
- Walters, L. (2009). Morgenbesser’s coin and counterfactuals with true components. *Proceedings of the Aristotelian Society*, 109(1pt3):365–379.
- Walters, L. and Williams, R. (2013). An argument for conjunction conditionalization. *Review of Symbolic Logic*, 6(4):573–588.
- Weatherson, B. (2015). Humean supervenience. In Loewer, B. and Schaffer, J., editors, *A Companion to David Lewis*. John Wiley & Sons.
- Williams, R. G. (2008). Chances, counterfactuals and similarity. *Philosophy and Phenomenological Research*, 77(2):385–420.
- Williams, R. G. (2010). Defending conditional excluded middle. *Noûs*, 44(4):650–668.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.
- Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, 67(295):173–180.