Might-counterfactuals and the principle of conditional excluded middle[*]

Ivar Hannikainen
University of Sheffield

Abstract

Owing to the problem of inescapable clashes, epistemic accounts of might-counterfactuals have recently gained traction. In a different vein, the might argument against conditional excluded middle has rendered the latter a contentious principle to incorporate into a logic for conditionals. The aim of this paper is to rescue both ontic might-counterfactuals and conditional excluded middle from these disparate debates and show them to be compatible. I argue that the antecedent of a might-counterfactual is semantically underdetermined with respect to the counterfactual worlds it selects for evaluation. This explains how might-counterfactuals select multiple counterfactual worlds as they apparently do and why their utterance confers a weaker alethic commitment on the speaker than does that of a would-counterfactual, as well as provides an ontic solution to inescapable clashes. I briefly sketch how the semantic underdetermination and truth conditions of might-counterfactuals are regulated by conversational context.

Keywords

Inescapable clashes, counterfactuals, Lewis-Stalnaker, possible worlds, semantic underdetermination.

1. Introduction

Consider the following conjunction:

$$(\varphi \Diamond\!\!\rightarrow \psi) \;\&\; (\varphi \Box\!\!\rightarrow \chi \vee \varphi \Box\!\!\rightarrow {\sim}\chi)$$

[*] Thank you to Marta Jorba, Sergi Oms, Miguel Ángel Sebastián and all the attendees for fostering such a productive and stimulating environment at the 13th Taller d'Investigació en Filosofia. I am grateful also to Javier Vilanova Arias for his terrific guidance in writing this paper and Antonio Blanco Salgueiro, Luis Fernández Moreno, Gonçalo Santos and an anonymous reviewer for precious feedback on diverse versions of this paper.

Disputatio, Vol. IV, No. 30, May 2011

Lewis famously upheld, as a consequence of his account of the comparative similarity relation between possible worlds, that the second conjunct need not be true. In so doing, he denied the *principle of conditional excluded middle* (CXM) and committed to saying things like:

> It is not the case that if Bizet and Verdi were compatriots, Bizet would be Italian; and it is not the case that if Bizet and Verdi were compatriots, Bizet would not be Italian (1973: 80).[1]

For Stalnaker (who understands might-counterfactuals as expressions of epistemic possibility), the first conjunct cannot be an ontic claim since whether 'If φ, then it might be the case that χ' is true or false depends on the speaker's epistemic status. Additionally, on account of Stalnaker's selection function, there is a single antecedent world by which to evaluate the truth of the consequent; and, therefore, there are no matters of fact about what *might* (or might not) have been the case, only about what *would* (or would not) have been the case. To me, this counts against Stalnaker's analysis: there must be matters of fact about what might counterfactually have been the case which might-counterfactuals serve to describe. Whether ψ might have been the case if it were the case that φ is an objective matter, and this being so is compatible with a semantics for conditionals (Stalnaker 1968, 1981) according to which χ either would have been the case if it were the case that φ or, if it were the case that φ, it would not have been the case that χ. The purpose of this essay, then, is to outline and defend an account of counterfactuals according to which CXM holds and might-counterfactuals express ontic, rather than epistemic, possibilities.

In Sections 2 and 3, I will introduce the topic by way of a historical review, looking at Lewis's and Stalnaker's views with regard to both might-counterfactuals and CXM, and then I will present the most developed epistemic account of might-counterfactuals (DeRose 1991, 1999). Next, in Section 4, I will lay out the major problem for ontic accounts that DeRose has furthered, the so-called *problem of inescapable clashes*, as well as his own solution. In the following section, Section

[1] Moreover, he claimed that $\sim(\varphi \,\Box\!\!\rightarrow \chi) \,\&\, \sim(\varphi \,\Box\!\!\rightarrow \sim\chi)$ does not contradict $\varphi \,\Box\!\!\rightarrow (\chi \vee \sim\chi)$. To be sure, for Stalnaker, on the other hand, 'Either if Bizet and Verdi were compatriots Bizet would be Italian, or Bizet would not be Italian if Bizet and Verdi were compatriots' is true.

5, I will advance my alternative explanation of the phenomenon of inescapable clashes. Lastly, in Sections 6 and 7, I will flesh out the account of might-counterfactuals that underlies my solution to the problem of inescapable clashes, try to answer some of the preliminary worries it could raise, and show how my ontic account of might-counterfactuals is compatible with CXM.

2. Might-counterfactuals and CXM in Lewis and Stalnaker

It'll be helpful to begin by briefly reviewing Lewis's and Stalnaker's semantics for conditionals paying special attention to how might-counterfactuals are defined and how CXM fares.

2.1. Lewis's duality thesis

Lewis's 1973 theory of conditionals is formulated in terms of a *comparative similarity relation*. Let the comparative similarity relation $C_i(j, k)$ mean that j is more similar to i than k is to i. A would-counterfactual, $\varphi \,\Box\!\!\rightarrow \psi$, is true iff there is a φ–world j such that ψ is true in j, and in all φ–worlds which are at least as similar to i as j. Crucially, the comparative similarity relation determines a weak total ordering which allows comparative similarity ties. In virtue of this feature, considering again the famous Bizet-Verdi example, the possible world(s) in which Bizet is French and the possible world(s) in which Verdi is Italian are tied in terms of comparative similarity. That is, there is a compatriot-world, c_1, in which 'Bizet is Italian' is true, but there is another compatriot-world, c_2, which is at least as similar to the actual world as c_1, in which 'Bizet is Italian' is false. Thus, the counterfactual 'If Bizet and Verdi were compatriots, Bizet would be Italian' is false. These very considerations make the counterfactual 'If Bizet and Verdi were compatriots, Bizet would not be Italian' false too. By conjoining these two false counterfactuals, we get Lewis's rejection of CXM.

Lewis put forth a straightforward and highly intuitive definition of might-counterfactuals in terms of would-counterfactuals known as the *duality thesis* (DT). It is straightforward and highly intuitive because it borrows the notions of necessity and possibility from modal logic and applies them to the setting of counterfactual conditionals in a rough-and-ready manner, *i.e.*:

[DT] $\varphi \Diamond\!\!\rightarrow \psi =_{df} \sim(\varphi \,\Box\!\!\rightarrow \sim\psi)$[2]

Lewis made the case that this definition of might-counterfactuals respects our ordinary usage of 'might' in counterfactual settings. Suppose I say 'If Lionel Messi had played for Real Madrid this season, Real Madrid might have won La Liga.' What I mean is that it is false that Real Madrid fails to win La Liga this season (that $\sim\psi$) in all possible worlds which (a) are at least as similar to the actual world as a world in which Lionel Messi plays for Real Madrid, and in which (b) Lionel Messi plays for Real Madrid. (Let's call worlds meeting these two criteria *relevant*). In other words, I am claiming that there is at least one relevant world which makes the conditional 'If Lionel Messi had played for Real Madrid this season, Real Madrid would not have won La Liga' false. If you thought I was mistaken, it seems likely that you should contradict me by saying: 'Even if Lionel Messi had played for Real Madrid this season, Real Madrid would not have won La Liga.'

2.2. Stalnaker's epistemic thesis

Evidently, DT is not available in Stalnaker's semantics for counterfactual conditionals. Recall that Stalnaker's truth conditions for conditionals (which I will adopt in my proposed analysis of the relation between might- and would-counterfactuals) are devised by using the *selection function* operator, f, and let $f(\varphi, i)$ be the selection function that picks out the possible world in which φ is true and which otherwise differs minimally from the base world, i. Then, $\varphi \,\Box\!\!\rightarrow \psi$ is true (/false) in i if ψ is true (/false) in the nearest φ-world $f(\varphi, i)$. It is a feature of Stalnaker's selection function that it operates under this so-called *Uniqueness Assumption*, according to which there is always at most a *single* φ-world at which to evaluate the truth of ψ. (For simplicity's sake, I will make this assumption too in the remainder of this paper. The analysis of might-counterfactuals I will develop is independent of the polemic about the similarity ordering of possible

[2] This definition is insufficient to deal with counterfactuals with impossible antecedents. For purposes of this paper, the analysis of ordinary language counterfactual conditionals, I will ignore the case of such counterfactuals and restrict the discussion to counterfactual conditionals with antecedents that describe possible states of affairs. The same point applies later in the paper to the provided definition of [ST]. Thanks to an anonymous reviewer for pointing this out.

worlds. Those worried about the ubiquity of comparative similarity ties, should help themselves to Stalnaker's 1981 appeal to supervaluations wherever I talk about the single nearest φ-world.) Seeing as for any φ, ψ will either be true or false at $f(\varphi, i)$, disjunctions like $(\varphi \Box\!\!\rightarrow \psi) \vee (\varphi \Box\!\!\rightarrow \sim\psi)$ will always be true; *i.e.*, CXM holds.[3] To see how DT and CXM are decidedly incompatible, consider the following argument known as the *might argument against conditional excluded middle* (Lewis 1973):

P1.	$\varphi \Diamond\!\!\rightarrow \psi = \sim(\varphi \Box\!\!\rightarrow \sim\psi)$	[DT]
P2.	$(\varphi \Box\!\!\rightarrow \psi) \vee (\varphi \Box\!\!\rightarrow \sim\psi)$	[CXM]
P3.	$\sim(\varphi \Box\!\!\rightarrow \sim\psi) \supset (\varphi \Box\!\!\rightarrow \psi)$	[from P2 and DF $\supset$]
P4.	$\varphi \Box\!\!\rightarrow \psi \supset \varphi \Diamond\!\!\rightarrow \psi$	[from DF $\Box\!\!\rightarrow$]
P5.	$\varphi \Diamond\!\!\rightarrow \psi \supset \varphi \Box\!\!\rightarrow \psi$	[from P1 and P3]
C.	$\varphi \Diamond\!\!\rightarrow \psi = \varphi \Box\!\!\rightarrow \psi$	[from P4 and P5]

Evidently, if we hold both DT and CXM, we arrive at the unhappy conclusion that might- and would-counterfactuals have the same truth conditions (which, it should be clear, is not faithful to their ordinary meaning in English). For the purposes of this paper, it being my aim to defend CXM, the most decisive consideration against DT is that, as seen above, it is incompatible with CXM. So, as an alternative to DT, Stalnaker 1981 proposes an epistemic view of might-counterfactuals, which is based on combining the semantics of 'might' outside conditional contexts with his analysis of would-counterfactuals.[4]

[3] Things change when supervaluations are considered (van Fraassen 1966). ψ will either be true or false for any possible valuation of φ. Thus, $(\varphi \Box\!\!\rightarrow \psi)$ and $(\varphi \Box\!\!\rightarrow \sim\psi)$ will each either be supertrue, superfalse or indeterminate, but the disjunction $(\varphi \Box\!\!\rightarrow \psi) \vee (\varphi \Box\!\!\rightarrow \sim\psi)$ will necessarily be true and CXM will remain valid.

[4] DeRose 1991, 1999, on whom I will focus in the bulk of this paper, painstakingly follows in Stalnaker's footsteps on this. However, while Stalnaker admits some non-epistemic uses of 'might' in counterfactual contexts (see his 1981: 99. 'But *might* sometimes expresses some kind of non-epistemic possibility. *John might have come to the party* could be used to say that it was within John's power to come, or that it was not inevitable that he not come'), DeRose 1999 thinks 'might' is *never* used to indicate non-epistemic possibility. For this reason, and because DeRose's account handles several types of uses of 'might' better than Stalnaker's does, my critique of the epistemic account of might-counterfactuals will focus on DeRose's account.

$$\varphi \Diamond\!\!\rightarrow \psi =_{\mathrm{df}} \langle e \rangle (\varphi \,\Box\!\!\rightarrow \psi)$$

He claims that ‘might’ outside conditional contexts indicates possibility and that the kind of possibility it typically expresses is epistemic possibility. In other words, ‘It might be the case that ψ’ means something like ‘What I know does not entail $\sim\psi$’ or ‘ψ is compatible with what I know’. Putting together the analyses of ‘might’ and of would-counterfactuals, Stalnaker’s thesis (ST) is that a might-counterfactual such as $\varphi \Diamond\!\!\rightarrow \psi$ as uttered by a speaker *S* means that nothing *S* knows obviously entails that ψ is false in $f(\varphi, i)$. Substituting the *definiens* above:

$$\textbf{[ST]} \qquad \varphi \Diamond\!\!\rightarrow \psi =_{\mathrm{df}} \sim K_S (\varphi \,\Box\!\!\rightarrow \sim\psi)$$

ST has a simple and famous rebuttal that Lewis (1973: 80-1) issued. Suppose I do not know what is in my pocket and I say ‘If I had looked in my pocket, I might have found a penny.’ The fact is that there is no penny in my pocket. This counterfactual is seemingly false, and DT explains why; *i.e.*, because ‘If I had looked in my pocket, I would not have found a penny’ is true. On ST, however, the counterfactual I uttered is equivalent to saying ‘It is compatible with what I know that if I had looked in my pocket, I would have found a penny,’ which is true. Thus, ST gives the wrong reading of counterfactuals such as the one in Lewis’s classic *Penny Case*.[5]

Hereafter I will dedicate my attention primarily to DeRose’s 1991, 1999 more honed and unswervingly epistemic account. The point here was merely to illustrate that the defender of CXM, like Stalnaker, will naturally favor an epistemic account of might-counterfactuals since it can be seamlessly coupled with the logical and semantic groundwork, put forth by Stalnaker in 1968, which upholds CXM. There is a *prima facie* conflict between giving an ontic reading to might-counterfactuals and preserving CXM to which it is the object of this paper to provide a peaceable resolution.

[5] Readers will be reminded of Stalnaker’s 1981 solution to the *penny case*. Indeed, Stalnaker provided a quasi-epistemic solution which dealt with this objection and which, in some ways, resembles the ontic account I will put forth.

3. DeRose and the problem of inescapable clashes

Keith DeRose thinks statements like ‘It is possible that *P,*’ ‘It might be the case that *P,*’ and, derivatively, might-counterfactuals express epistemic possibilities *all the time*. In DeRose 1991, he weaves through a range of cases in which a speaker for whom it is epistemically possible that *P* may felicitously assert that ‘It is possible that *P*,’ and he arrives at the following flexible proposal that statements of this kind are true iff:

(1) no member of the relevant community knows that *P* is false, and
(2) there is no relevant way by which members of the relevant community can come to know that *P* is false (1991: 593-4).

Substituting Stalnaker’s epistemic possibility operator, <*e*>, with DeRose’s analysis of epistemic possibility one gets:

$\varphi \Diamond\!\!\rightarrow \psi$ is true iff
(1) no member of the relevant community knows that $(\varphi \Box\!\!\rightarrow \sim\psi)$, and
(2) there is no relevant way by which members of the relevant community can come to know that $(\varphi \Box\!\!\rightarrow \sim\psi)$.

Or:

[ET] $\varphi \Diamond\!\!\rightarrow \psi =_{df} \sim K_{rc}\,(\varphi \Box\!\!\rightarrow \sim\psi)\ \&\ \sim\blacklozenge K_{rc}\,(\varphi \Box\!\!\rightarrow \sim\psi)$[6]

DeRose admits that the ‘relevant community’ and the ‘relevant way’ are vague notions but, by means of a variety of examples, shows that any greater specificity in giving truth conditions – of which the prototypical example is ST – too easily generates counterexamples.

3.1. The problem of inescapable clashes

DeRose 1999 has argued that an inescapable problem haunts non-epistemic accounts of might-counterfactuals.[7] While everyday coun-

[6] Think of ♦ as a highly specific modal operator that designates possible worlds which are accessible in the ‘relevant ways’ that DeRose has in mind.

[7] For different treatments of the problem of inescapable clashes, see also Eagle unpublished, Hawthorne 2005, and Williams 2010.

terfactuals are subject to this problem (see DeRose's baseball example [1999: 385-6]), Hawthorne 2005 has pointed out how quantum theory threatens to falsify any counterfactual conditional grounded in the principles of classical mechanics. Take the following commonsensically true counterfactual:

> **(W)** If I had dropped the plate, it would have fallen to the floor.

In a world governed by quantum mechanics, we must be prepared to accept that:

> **(M)** If I had dropped the plate, it might have flown off sideways.

Notice that (M) is a might-counterfactual, so that to deny (M) would be to insist that it is impossible for the plate to have flown off sideways and to refuse the conclusions of quantum theory. So, if we grant that (M) is true, this will lead us to believe that:

> **(M')** If I had dropped the plate, it might not have fallen to the floor.

It doesn't seem reasonable to agree to (M) while denying (M'). And once you agree to both (W) and (M'), you face the problem of inescapable clashes:

> **(W+M')** If I had dropped the plate, it would have fallen to the floor; nevertheless, the plate might not have fallen to the floor if I had dropped it.[8]

If one holds, as DT does, that (W+M') expresses an inconsistent proposition, one is forced to backtrack and choose between denying (W) – the *skeptical* option – and denying (M') – the *exclusionary* option – two counterfactual claims both of which have 'a good deal of initial plausibility' (DeRose 1999: 387). According to counterfactual skepticism, since (M') is a weaker commitment than is (W), one is encouraged to think of (M') as making (W) false. The counterfactual exclusionary strategy holds that the falsity of most ordinary counterfactuals is too high a price to pay to acknowledge the possibility of quasi-miracles. It is preferable to argue instead from the truth of the

[8] To be sure, conjunctions of the form $\varphi \ \Box\!\!\rightarrow \sim\psi \ \& \ \varphi \ \Diamond\!\!\rightarrow \psi$ are also instances of the phenomenon of inescapable clashes.

ordinary counterfactual to the falsity of the corresponding might-counterfactual. Thus, these theories of counterfactuals exclude remarkably low-probability outcomes and say things like 'If I were to roll a die a billion times, it's not the case that it might land tails every time' (Lewis 1979b, Williams 2008).

3.2. DeRose's escapism from pragmatic clashes

DeRose wants to say that both (W) and (M') are true; thus, that there is no semantic contradiction in (W+M') but that, nonetheless, there is pragmatic tension involved in utterances of (W+M') which explains why they are unassertible. He claims, moreover, that his account is singlehandedly capable of accommodating the truth of (W) and (M') while respecting the intuitive proscription against utterances of this kind. DeRose's pragmatic explanation of inescapable clashes is the following: In flat out asserting (W), the speaker represents herself as knowing that (W) while uttering (M') expresses the epistemic possibility for the speaker that (W) is false.

> Thus, what one says in asserting the second conjunct of [(W+M')], while it's perfectly consistent with what one says in asserting the first conjunct, is inconsistent with something one represents as being the case in asserting the first conjunct. This supports our sense that *some* inconsistency is responsible for the clash involved in asserting the conjunction, while, at the same time, happily removing that inconsistency from the realm of what's asserted: The conjunction asserted is itself perfectly consistent, but in trying to assert it, one gets involved in a contradiction between one thing that one asserts, and another thing that one represents as being the case (1999: 389).

DeRose concludes that ET is superior to DT in that it 'provides a way of avoiding the *really* nasty conclusion—that [(W)] is false' (1999: 390). Furthermore, he claims that other, non-epistemic theories are defective insofar as they define might-counterfactuals in a way that renders conjunctions like (W+M') the right thing to say in certain circumstances, and thereby succumb to the problem of inescapable clashes. DeRose discusses Heller's 1995 theory in Section 8 (1999: 395-6), and Lewis's 1986 'ambiguity thesis' in Section 9 (1999: 396-7). Briefly, Heller's theory claims that $\varphi \Diamond\!\!\rightarrow \sim\psi$ is true iff there is at least one *close enough* φ-world in which $\sim\psi$ is true. Thus, (W+M') is the right thing to say when ψ is true in the closest φ-worlds and false in some presumably farther yet close enough φ-world. Lewis's theory is that might-counterfactuals are ambiguous between DT and another reading according to which $\varphi \Diamond\!\!\rightarrow \psi$ is true iff some relevant φ-worlds are worlds where there is a non-zero chance of ψ being false (1986: 63-4). This theory makes (W+M') the right thing to say when ψ is true in the relevant φ-worlds but there is a non-zero chance of ψ being false in some of the relevant φ-worlds. The spirit of the definition of might-counterfactuals I will present in this paper is especially close to that of Heller's, insofar as I embrace and exploit his claim that might-counterfactuals admit 'gratuitous differences'. Contrary to DeRose's allegation that ontic accounts of might-counterfactuals cannot solve the problem of inescapable clashes, I will develop in the following section an account of the phenomenon of inescapable clashes that can be appended to ontic theories of might-counterfactuals in order to solve this problem.

4. The contextual-shift explanation of inescapable clashes

Consider the following discourse, from a high school principal talking to a teacher, featuring two italicized quantificational claims:

> Mr. D'Elia, I looked through the grade reports of your History 101 class this Spring. *Every single student failed the final exam.* You really ought to lower your expectations of undergraduate students. I understand that you want to promote excellence in the student body, but there are more effective ways to go about it. I remember Mr. Shillington. He's one of the best History teachers that ever came through Bumbletown High. He was rigorous but he always made sure the more promising and hard-working students were rewarded. Let me pull up his grade reports for History 101… Here they are! *10% of students passed the final exam.*

If we take the high school principal's quantificational claims out of context and conjoin them, we get:

> **(A+S')** Every single student failed the final exam, but 10% of students passed the final exam.

When we do this, we generate what has the appearance of a contradiction. But, of course, one would hardly say it *is* one since, in order to evaluate the italicized sentences, we must specify by extracting from the conversational context the (hitherto implicit) range over which quantification takes place and, when we do so, we see clearly that (A+S') amounts to a non-contradictory, and objectively verifiable proposition, which is true just in case

> every single student *in Mr. D'Elia's History 101* class failed the final exam, and 10% of students *in Mr. Shillington's History 101 class* passed the final exam.

The conjunction of these claims, (A+S'), is seemingly contradictory because there is a fairly prevalent pragmatic rule in natural language use, to do with anaphoric reference, according to which the range of quantification remains stable until it is explicitly set to a new range by the conversational context. In this specific instance, by anaphora, the second sentence would inherit the contextually-salient range of quantification over which the first quantificational statement holds. Whatever this class of 'students' turned out to be, the sequence or conjunction of these two quantificational sentences would lend itself to being interpreted as an obvious contradiction because the proposition expressed would be thought to be:

$$\forall x\ (Sx \rightarrow Fx)\ \&\ \exists x\ (Sx \rightarrow \sim Fx)$$

But, as seen, the range of quantification is left unspecified. There is nothing explicit in the sentence form of either quantificational statement to specify the range over which the proposition quantifies and thus nothing in the sentence form to establish definitively that this utterance involves a semantic contradiction. If the speaker of (A+S') were to insist that (A+S') is true on the grounds that every single student in Mr. D'Elia's History 101 class failed the final exam while 10% of students in Mr. Shillington's History 101 class passed the final exam, we would have to suppose there is something seriously wrong about his dominion of conversational pragmatics, but it would be odd to insist, beyond the unassertability of (A+S'), that (A+S') was false on these grounds.

I believe an analogous kind of pragmatic failure to make explicit a contextual shift takes place amid (W+M')-type conjunctions. A speaker can hold both (W) and (M') and utter them on separate occasions, as long as a contextual shift is adequately established between them, *e.g.*:

> Quantum mechanics, which I believe in, warns us about the possibility of highly erratic physical phenomena. For example, remember the plate I was spinning on my index finger yesterday. *If I had dropped the plate, it might not have fallen to the floor.* It could have flown off sideways instead. Yet, at the same time, commonsense and good ole' Newtonian mechanics tells me that it is oh-so very likely that the plate will shatter on the floor. You can hardly deny that. *If I had dropped the plate, it would have fallen to the floor.* I cannot be certain of it, but I bet it would happen even while I recognize that quantum oddities are possible.

What she cannot do felicitously is utter (W) and (M') in conjunction or in sequence without thereby *almost* invariably generating the impression of an obvious contradiction.[9] This explanation seems to me to satisfy DeRose's two criteria for an adequate solution to the problem of inescapable clashes, *i.e.*, the proposition expressed by (W+M') is not semantically contradictory, yet

(1) (W+M') is invariably unassertible.

DeRose might raise the same objection here that he raises against the non-DT version of Heller's 1995 view:

> to get a non-DT version of Heller's view, there should be contexts in which the range of [φ-]worlds relevant to the 'might' counterfactuals is different from (no doubt broader than) the range of worlds relevant to the 'would' counterfactuals. In such contexts, [(W+M')-type] conjunctions should be unproblematic. But there are no such contexts; these conjunctions always clash. So any non-DT version of Heller's view will succumb to the problem of inescapable clashes (1999: 396).

Indeed, on both Heller's and my view, $\varphi \;\Box\!\!\rightarrow \psi \;\&\; \varphi \;\Diamond\!\!\rightarrow \sim\!\psi$ can express a consistent, counterfactual proposition; namely, by describing an objective, counterfactual state of affairs across multiple φ-worlds. Thus, as DeRose says, (W+M') indeed should be unproblematic in such contexts. However, as I hope my account has explained, without the requisite contextual shift between (W) and (M'), the

[9] I say 'almost' because I think there are circumstances in which these conjunctions are assertible without the overt contextual shift, namely, when the function of discourse is *exploratory* (see the example on p. 29).

sentence form of (W+M')-type conjunctions renders them almost invariably unassertible.

I have here outlined the contextual shift explanation of inescapable clashes which, I believe, provides – by DeRose's own standards – an adequate account of the phenomenon. (1) The thought expressed is not semantically contradictory such that (M') does not contradict (W) and *vice versa*, thereby providing an alternative to counterfactual skepticism and counterfactual exclusion. Nonetheless, (2) said conjunctions invariably clash; they are never (or almost never) the right thing to *say*. Contrary to DeRose's presumption, the ontic camp can provide a solution to the problem of inescapable clashes.

As a sidenote, these conjunctions can, on my view, be the right thing to *think*, and this should be seen as an important advantage over DeRose's solution. As Eagle has pointed out, though DeRose's epistemic view does successfully dodge counterfactual skepticism and exclusion, it is subject to *weak counterfactual skepticism*: 'the thesis that, even if they are true, ordinary 'would' counterfactual claims cannot be known if the corresponding 'might' counterfactuals are known' (unpublished). This seems like a significant downfall for epistemic theories. It is counterintuitive to suppose that speakers cannot at a single time know both (W) and (M'): after all, a mature epistemic agent knows that she would not have won the national lottery had she picked some other number but that, of course, she just might have. On my account of might-counterfactuals, as on Heller's, one avoids this epistemic brand of counterfactual skepticism too.

5. Semantically underdetermined might-counterfactuals

Besides showing that the contextual-shift solution is an *adequate* ontic account, I will now argue that it is a *plausible* one, *i.e.*, that there are independent reasons to think that corresponding might- and would-counterfactuals are evaluated by taking into account different possible worlds (or sets of possible worlds), and thus that the antecedent of a counterfactual conditional, appearances notwithstanding, makes a different semantic contribution in the context of a might-counterfactual than it does in the context of a would-counterfactual. Consider the following three such reasons: first, multiple antecedent-worlds must be relevant to the evaluation of might-counterfactuals; second, semantic underdetermination allows multiple antecedent-worlds to be relevant to the truth-conditional evaluation of might-counterfactuals; and third, it is consistent with the discursive functions of might- and would-counterfactuals that speakers would use might-counterfactuals, but not would-counterfactuals, to appeal deliberately to the semantic underdetermination in the antecedent.

As I prefaced, might-counterfactuals, unlike would-counterfactuals, apparently must be evaluated by taking into account multiple antecedent-worlds. Stalnaker (1981: 91-5) defended, in relation to would-counterfactuals, that when a speaker asserts $\varphi \mathbin{\Box\!\!\rightarrow} \psi$, she purports to represent and describe a 'unique determinate possible world,' namely the φ-world selected by $f(\varphi, i)$, a possible world which other than accommodating the truth of φ differs minimally from the base world, i. This can be seen by considering the following dialog:

> X: President Carter would have appointed a woman to the Supreme Court last year if there had been a vacancy.
> Y: Who do you think he would have appointed?
> X: He wouldn't have appointed any particular woman; he just would have appointed some woman or other (Stalnaker 1981: 94).

X's response seems bad because the fact is that, if there had been a vacancy in the Supreme Court and President Carter had appointed a woman, he must have appointed some *particular* woman. (If X cannot name her, it is due to her epistemic limitations not to any insurmountable metaphysical vagueness.) This is, according to Stalnaker, because would-counterfactuals are evaluated in each case by taking into account the single nearest antecedent-world. Another consequence of the Uniqueness Assumption is that a speaker cannot go on to say that $\varphi \mathbin{\Box\!\!\rightarrow} {\sim}\psi$ without thereby contradicting herself as to what the nearest φ-world is like.

Now consider the Uniqueness Assumption with respect to might-counterfactuals. A speaker can felicitously and truthfully say things like 'If Messi had played for Real Madrid this season, Real Madrid might have won La Liga but Real Madrid might also have not won La Liga,' *i.e.*,

(M+M') $(\varphi \mathbin{\Diamond\!\!\rightarrow} \psi) \;\&\; (\varphi \mathbin{\Diamond\!\!\rightarrow} {\sim}\psi)$

If, here again, a single possible world were relevant to the truth conditional evaluation of the speaker's statement, it would be hard to see how her statement could be meaningful. But it seems that these statements are meaningful and sometimes true. So, if (i) Stalnaker is right, that would-counterfactuals are evaluated by taking into account only the single nearest antecedent-world, and (ii) seeing as (M+M')-type conjunctions, in which the contradictory consequents of two conjoined (and true) might-counterfactuals, are assertible, then it must be the case that a class of multiple antecedent-worlds is relevant to the truth conditional evaluation of might-counterfactuals. Thus, the antecedent of a counterfactual conditional must make a different semantic contribution in the context of a might-counterfactual than it does in the context of a would-counterfactual.

How is φ capable of diverging from its semantic contribution to a would-counterfactual? Moreover, how are *multiple* possible worlds at once worlds in which a proposition, φ, is true? It'll be easier to answer the questions in reverse order. Take the proposition 'Popes are not young.' Is this proposition true or false? Or indeterminate? If it is true of the actual world, what is it true *in virtue of*? If it is false or indeterminate, what possible state of affairs would render it true? There clearly are obstacles to determining what such state of affairs would be, and this is due to the inherent semantic underdetermination of this sentence. Firstly, the range of Popes over which this claim quantifies – whether it is all the Popes in history, most of them, only those with which the relevant community is acquainted, only those that are salient in the conversational context, *etc.*, – is unspecified. Secondly, there is no sharp boundary between *being young* and *not being young*, so the property that is predicated of Popes is underdetermined. Even while not knowing what it would take for 'Popes are not young' to be true, it seems obvious that any number of possible states of affairs could make it true (see Fine 1975). Even if the proposition were about a single, identifiable person and contained no vague predicates, *e.g.*,

> Jesulin de Ubrique's cape on the night of his professional bullfighting debut was red,

there would be no single state of affairs that this statement could truthfully report. Suppose we grant that this statement is true in the actual world given the particular cape's actual color. It is true too in those possible worlds in which it is a very slightly different shade of red, a slightly orangish shade of red, a slightly purplish or pinkish shade of red, *etc*. By the same token, virtually any proposition figuring in a counterfactual antecedent, φ, – 'I look in my pocket,' 'The pirates do not threaten the operation,' 'John does not suffer sudden cardiac arrest,' *etc.*, – is semantically underdetermined: it does not contain enough information to select a single φ-world for evaluation. Take 'Lionel Messi plays in Real Madrid this season.' There is a φ_1-world in which Messi plays three games for Real Madrid and gets injured, a φ_2-world in which Messi plays at least sixty minutes out of every league game for Real Madrid, a φ_3-world in which Messi plays for Real Madrid while Iniesta moves back to Albacete to live the simple life, and so on. In other words, in virtue of semantic underdetermination, multiple possible worlds fit the description in φ and all these φ-worlds are truthmakers for the proposition in a counterfactual antecedent, φ.[10] Since, in a would-counterfactual context, the semantic gaps in φ are 'plugged' by the requirement that the φ-world for evaluation be maximally similar to the base world, φ in a might-counterfactual is able to depart from its semantic contribution to a would-counterfactual.

The contention that semantic underdetermination is at work in the utterance and evaluation of might-counterfactuals should garner intuitive appeal by looking at the discursive functions of might- and would-counterfactuals. A speaker who believes with some degree of confidence that

> ψ would have been the case if it were the case that φ, *where only the nearest φ-world is a world in which φ is true*,

[10] I will not in this paper delve into what the sources of semantic underdetermination are, which I suspect is an empirical matter anyhow. However, here are some suggestions: the enrichment of the antecedent with presuppositions ('If Messi played for Real Madrid this season [*and Iniesta moved back to Albacete to live the simple life*], then…), the resolution of the underdetermination in similarity respects and relations ('If Sheffield were more like San Diego, then…', 'If I exercised [*1.5 hrs/day*] more, then…'), vague terms ('If Popes were generally younger, then…), and non-natural predicates ('If Jesulin de Ubrique's bullfighting-debut cape were not red, then…').

seems warranted in asserting the corresponding would-counterfactual ‘If it were the case that φ, then it would be the case that ψ’ because she thereby commits to it being true that

$$f(\varphi, i) \in \psi.$$

By contrast, a speaker who (is not confident that *ψ would have been the case if φ*, but rather) believes with some degree of confidence that

> ψ would have been the case if it were the case that φ, *where every member of the class of φ-worlds selected by the range of admissible precisifications of φ is a world in which φ is true*,

is prudent if he couches his claim in terms of a might-counterfactual, ‘If it were the case that φ, then it might be the case that ψ’ because he thereby makes only the far weaker commitment that

$$f(v_1(\varphi), i) \in \psi \ \vee\ f(v_2(\varphi), i) \in \psi \ \vee\ f(v_3(\varphi), i) \in \psi \ \vee \ldots f(v_{\mathrm{n}}(\varphi), i) \in \psi.$$

It should seem plausible now that multiple antecedent-worlds, which are irrelevant to the evaluation of the corresponding would-counterfactuals, are relevant to the truth-conditional evaluation of might-counterfactuals and semantic underdetermination explains how they are capable of so being: in the context of a might-counterfactual, the antecedent is semantically underdetermined with respect to the class of its truthmaking antecedent-worlds. This in turn enables speakers to use might-counterfactuals to talk, veridically and prolifically, about counterfactual possibilities.

6. The semantics of semantically underdetermined might-counterfactuals

In the previous section, I hinted at my truth conditions for might-counterfactuals, which I will here lay out explicitly. A given might-counterfactual,

$\varphi \Diamond\!\!\rightarrow \psi$ is true iff for *some* admissible precisification of φ, $v_k(\varphi)$, the nearest $v_k(\varphi)$-world, $f(v_k(\varphi), i)$, is a ψ-world; and

$\varphi \Diamond\!\!\rightarrow \psi$ is false iff for *every* admissible precisification of φ, $v_1(\varphi)$, $v_2(\varphi)$, $v_3(\varphi)$ … $v_n(\varphi)$, the nearest $v_n(\varphi)$-world, $f(v_n(\varphi), i)$, is a $\sim\psi$-world.

The proposed analysis of might-counterfactuals renders (M') consistent with (W) as seen in the contextual-shift solution to inescapable clashes. A counterintuitive consequence of this is that, contrary to what speakers seem prone to do (as seen in the *Lionel Messi Case* and Lewis's *Penny Case*), it is not generally valid to use (W) to falsify (M'), and *vice versa*. To see this, consider again the *Lionel Messi Case*. On Stalnaker's view of would-counterfactuals (which I adopt here), when we are asked to imagine what would have happened if Lionel Messi had played in Real Madrid this season, we conceive the counterfactual state of affairs most similar to the actual world with those minimal modifications made which are necessary to make Lionel Messi a member of Real Madrid's squad; while, on the proposed account of might-counterfactuals, when we are asked to imagine what *might* have happened if Lionel Messi had played in Real Madrid this season, we let our imagination run looser and conceive a range of counterfactual states of affairs with varying degrees of similarity to the actual world all of which about which 'Lionel Messi plays for Real Madrid this season' is true. This difference in the antecedent-worlds relevant for evaluation is precisely why certain possibilities *might* have been realized if such and such were the case that *would not* have been realized, and why the truth of (W) is consistent with the truth of (M').

One worry, which I will here attempt to assuage, easily engendered by this kind of proposal is that it is perhaps unclear what prevents all might-counterfactuals from being true. First of all, invariantly, counterfactual possibilities can be ruled out that presuppose inadmissible precisifications of the antecedent. If only *in*admissible precisifications of the antecedent select possible worlds in which the consequent is true, then the might-counterfactual in question is false. Here are some examples of invariantly false might-counterfactuals:

If I had looked in my pocket, I might not have looked in my pocket.
If the pirates had not threatened the operation, 2+2 might have been equal to 5.
If Lionel Messi had played for Real Madrid this season, *Mus musculus* (the house mouse) would be capable of prolonged levitation.
If John had not suffered sudden cardiac arrest, Kerry might have won the 2004 presidential election.

These might-counterfactuals are clearly false from the outset. There are no precisifications of what 'The pirates threaten the operation' means according to which, if the state of affairs described by such precisification held, it might be the case that '2+2 equals 5,' and *mutatis mutandis* for the rest of examples.

Much more frequently, might-counterfactuals can be falsified by a constraint on the admissibility of precisifications regulated by the conversational context. (Note that this conversational constraint must be a contingent matter so as to not incite the problem of the inescapable clashes.) Let's begin with an example in which the conversational context *slackens* the admissibility of precisifications. Suppose the function of discourse is *exploratory*, as in the following example:

> If I had practiced the guitar a lot, I would have had a record deal with Sony. In fact, I might have been rich and famous enough to do without the support of a record label if I had practiced the guitar a lot.

In this discourse type, the would-counterfactual doesn't seem to falsify the might-counterfactual. 'I practice the guitar a lot' is semantically underdetermined in the might-counterfactual above such that antecedent-worlds – which are more remote than the maximally similar antecedent-world in which the speaker's gets a record deal with Sony – are relevant to the evaluation of the might-counterfactual. The function of exploratory discourse renders said precisification admissible (N.B. This is, evidently, not to say that any of the antecedent-worlds selected by such admissible precisifications of the antecedent are worlds in which the speaker is rich and famous enough to do without a record label. That is the *next*, and final, step in the truth conditional evaluation of a might-counterfactual).

In other contexts, the aim of counterfactual discourse is, along Stalnaker's lines, to find out what the actual world would have been like if the antecedent had been true. Now consider counterfactual discourse with this, *truth-aiming* purpose. The familiar might-counterfactual

> Had the pirates not threatened the operation, we might have found the vessel

can be falsified by the would-counterfactual

> If the pirates had not threatened the operation, we would not have found the vessel.

In truth-aiming counterfactual deliberation, remote counterfactual possibilities are rendered false by claims that approximate counterfactual truth. In these contexts, the range of admissible precisifications of the antecedent is *constrained* and this has the consequence of falsifying the might-counterfactual under consideration.

Like many elements of conversational score, precisification-admissibility cannot vary wildly: this would render might-counterfactuals unintelligible. Consider again the conjunction 'If Messi had played for Real Madrid this season, Real Madrid might have won La Liga but Real Madrid might also not have won La Liga.' Let's say this conjunction is true in virtue of it being the case that

> if Messi had played for Real Madrid [and Iniesta were frequently injured] this season, then Real Madrid *would* have won la Liga; &
> if Messi had played for Real Madrid this season [only throughout the second leg], then Real Madrid *would not* have la Liga.

There is a stable criterion – let's suppose that in this conversation it's about the ordinary, significantly likely possibilities throughout a football season – that regulates precisification-admissibility in *both* conjuncts. It would be very odd and implausible if the above conjunction were true in virtue of it being the case that

> if Messi had played for Real Madrid [and Iniesta were frequently injured] this season, then Real Madrid *would* have won la Liga; &
> if Messi had played for Real Madrid this season [and Guardiola traveled back in time to sign Pele onto Barça], then Real Madrid *would not* have won la Liga;

where, initially, only significantly likely football happenings are revelant but then, by the second conjunct, the possibilities that time travel offers are all of a sudden relevant. Thus, like conversational score, the criterion for precisification-admissibility in counterfactual discourse typically cannot vary in a wild manner.

This is a mere sketch of how the semantic underdetermination in might-counterfactuals is regulated by the conversational context. A lot more could be said, but I hope to have outlined the main claims (and thereby given a truth conditions for might-counterfactuals):

counterfactual discourse can either be exploratory or truth-aiming and this discourse-type affects the way in which the admissibility of precisifications of φ is regulated by something akin to Lewis's 1979a *conversational score*, *i.e.*, whether it is slackened or constrained.[11]

6.1. Back to CXM

Throughout previous sections, I have developed an ontic account of might-counterfactuals that – I hope to have demonstrated – meets DeRose's challenge. However, the overarching purpose of this paper, which I will now succinctly take on, was to show that this ontic account (unlike DT) is compatible with Stalnaker's CXM-preserving semantics.

I have claimed that $\varphi \Diamond\!\!\rightarrow \psi$ is true iff it is the case that ψ in at least one of the class of φ-worlds selected by the range of admissible precisifications of φ. To be sure, the precisification of φ that figures in $f(\varphi, i)$ is among the admissible precisifications of φ. Therefore,

$$\varphi \,\Box\!\!\rightarrow \psi \supset \varphi \Diamond\!\!\rightarrow \psi$$
$$\varphi \,\Box\!\!\rightarrow {\sim}\psi \supset \varphi \Diamond\!\!\rightarrow {\sim}\psi$$

And:

$${\sim}(\varphi \Diamond\!\!\rightarrow \psi) \supset (\varphi \,\Box\!\!\rightarrow {\sim}\psi)$$
$${\sim}(\varphi \Diamond\!\!\rightarrow {\sim}\psi) \supset (\varphi \,\Box\!\!\rightarrow \psi)$$

Notice in the bottom pair of validities that the bi-directional entailment, which held in DT, doesn't hold. This is a desired feature of my account which both provides an escape from inescapable clashes and renders the 'might' argument against CXM, outlined in Section 2.2, a non-starter.

7. Conclusion

Owing to the notorious problem of inescapable clashes, idealized epistemic accounts of might-counterfactuals, such as DeRose's 1999,

[11] Lewis's examples 2 and 6 on *Permissibility* and *Relative Modalities* respectively may be particularly relevant to the dynamics of counterfactual discourse function and precisification-admissibility.

have recently gained popularity over ontic accounts. In a different vein, the might argument against conditional excluded middle has rendered CXM a contentious principle to incorporate into a logic for conditionals. The aim of this paper has been to rescue both ontic might-counterfactuals and conditional excluded middle from these disparate debates and show how they are indeed compatible.

According to the proposed account of might-counterfactuals, the antecedent of a might-counterfactual is semantically underdetermined with respect to the antecedent-worlds it selects for evaluation. This explains (1a) how might-counterfactuals are able to select multiple antecedent-worlds as they apparently do and (1b) why the utterance of a might-counterfactual confers a weaker alethic commitment on the speaker than does the utterance of a would-counterfactual, as well as (2) provides an ontic solution to the problem of inescapable clashes. I have also briefly sketched how the semantic underdetermination, and consequently the truth conditions, of semantically underdetermined might-counterfactuals are regulated by the conversational context. Namely, a conversational score keeps track of the stringency of precisification-admissibility and thereby determines the truth conditions of any might-counterfactuals under evaluation.

The proposed account should be favored by those who share my intuition that there are counterfactually possible states of affairs which might-counterfactuals serve to describe. Additionally, unlike with epistemic theories which succumb to weak counterfactual skepticism, the proposed ontic theory is able to account for the concurrent knowability of would- and corresponding might-counterfactuals. Alternately, if the assumption that counterfactuals claims are knowable *at all* turned out problematic, this would undermine the central spirit of epistemic theories while leaving the ontic accounts practically intact.

Ivar Hannikainen
Department of Philosophy
University of Sheffield
45 Victoria Street
Sheffield, S3 7QB, UK
ivar.hannikainen@gmail.com

References

DeRose, Keith. 1991. Epistemic possibilities. *The Philosophical Review* 100, 581-605.

DeRose, Keith. 1999. Can it be that it would have been even though it might not have been? *Philosophical Perspectives* 13, 385-413.

Eagle, Antony. 'Might' counterfactuals. Unpublished. Retrieved through the author's website on September 26, 2010 from dl.dropbox.com/u/6362052/might-cfacts.pdf.

Fine, Kit. 1975. Vagueness, truth and logic. *Synthese* 30, 265-300.

Hacking, Ian. 1967. Possibility. *The Philosophical Review* 76, 143-168.

Hawthorne, Jon. 2005. Chance and counterfactuals. *Philosophy and Phenomenological Research* 70, 396-405.

Heller, Mark. 1995. Might-counterfactuals and gratuitous differences. *Australasian Journal of Philosophy* 73, 91-101.

Lewis, David. 1973. Counterfactuals. Oxford: Basil Blackwell.

Lewis, David. 1979a. Scorekeeping in a language game. *Journal of Philosophical Logic* 8, 339-359.

Lewis, David. 1979b. Counterfactual dependence and time's arrow. *Noûs* 13, 455-476.

Lewis, David. 1986. Postscript to 'Counterfactual Dependence and Time's Arrow'. In *Philosophical Papers Vol. II*. Oxford: Oxford University Press.

Stalnaker, Robert. 1968. A theory of conditionals. *Studies in Logical Theory: American Philosophical Quarterly Monograph Series* 2, 98–122.

Stalnaker, Robert. 1981. A defense of conditional excluded middle. In *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Edited by W. Harper, R. Stalnaker and G. Pearce. Dordrecht: D. Reidel.

van Fraasen, Bas. 1966. Singular terms, truth-value gaps, and free logic. *Journal of Philosophy* 63, 481-495.

Williams, J. Robert G. 2008. Chances, counterfactuals and similarity. *Philosophy and Phenomenological Research* 77, 385-420.

Williams, J. Robert G. 2010. Defending conditional excluded middle. *Noûs* 44, 650-668.