Defending Conditional Excluded Middle

J. Robert G. Williams*

(June 24, 2008)

Lewis (1973) gave a short argument against conditional excluded middle, based on his treatment of 'might' counterfactuals. Bennett (2003), with much of the recent literature, gives an alternative take on 'might' counterfactuals. But Bennett claims the might-argument against CEM still goes through. This turns on a specific claim I call Bennett's Hypothesis. I argue that independently of issues to do with the proper analysis of might-counterfactuals, Bennett's Hypothesis is inconsistent with CEM. But Bennett's Hypothesis is independently objectionable, so we should resolve this tension by dropping the Hypothesis, not by dropping CEM.

Orthodoxy has it that Lewis (1973) gave decisive considerations against Conditional Excluded Middle: the principle that $A \mathrel{\Box\!\!\rightarrow} B \vee A \mathrel{\Box\!\!\rightarrow} \neg B$, for arbitrary A, B. Consider Bizet and Verdi—if they were compatriots, would they be Italian? If they were compatriots, would they be non-Italian? Lewis argues both counterfactuals are false. Stalnaker (1980) urged that instead that both these conditionals are *indeterminate*, but that their disjunction is true.[1]

This paper begins by providing a positive general argument for Conditional Excluded Middle. The burden of proof is therefore shifted to those who would argue against the principle.

The main focus of the paper is on the merits of an influential argument that Lewis (1973) gives against Conditional Excluded Middle—an argument that argues the the principal undermines the best account of might-counterfactuals. The argument, however, turns on appeal to a contentious 'duality' analysis of might-counterfactuals. I argue that the duality result is problematic, and we should look instead to analyze might-counterfactuals as combinations of independently motivated theories of counterfactuals and epistemic modals.

What becomes of the might-argument against Conditional Excluded Middle in this setting? One might think it falls apart. But Bennett (2003) has argued that we can derive the needed duality premise as a *consequence* of an independently motivated treatment of might-counterfactuals. I argue that Bennett's contention turns on tacitly assuming a certain kind of epistemic constraint on true counterfactuals—Bennett's Hypothesis. Bennett's Hypothesis alone (independently of

*Thanks are owing to many people, including Daniel Nolan, Daniel Elstein, Carrie Jenkins, Ross Cameron, Andy McGonigal, Elizabeth Barnes, Rich Woodward, Ant Eagle, and the Rutgers philosophy of language discussion group.

[1]Stalnaker appeals to a supervaluational conception of indeterminacy, of the kind elaborated in, for example, Fine (1975).

the question of how to analyze might-counterfactuals) gives us an argument against Conditional Excluded Middle.

In the final section of the paper I argue that Bennett's Hypothesis, again independently of consideration of the analysis of epistemic modals, threatens an error theory of ordinary counterfactuals. We thus have reason to reject it. But without it, there is no reason to think that duality will hold good for any independently motivated theory of might-counterfactuals; and without this, there is no might argument against Conditional Excluded Middle.

1 Opposite and negated conditionals

How can one express the negation of a counterfactual conditional? Consider the following:

(a) If I were to jump on the ice, I would fall in

(b) It is not the case that: if I were to jump on the ice, I would fall in

(c) If I were to jump on the ice, I wouldn't fall in.

To express the negation of (a), one might use clumsy formulations such as (b). But far more natural is to express one's disagreement with the so-called 'opposite conditional' (c).

However, if an opposite conditional is to express the negation of a counterfactual then the relevant instance of the following must hold:[2]

Equiv: $(A \mathrel{\Box\!\!\rightarrow} \neg C) \Leftrightarrow \neg(A \mathrel{\Box\!\!\rightarrow} C)$

For conditionals with possibly true antecedents, this is equivalent to the corresponding instance of the following:

CEM: $(A \mathrel{\Box\!\!\rightarrow} C) \vee (A \mathrel{\Box\!\!\rightarrow} \neg C)$

This is Conditional Excluded Middle for counterfactual conditionals.[3]

[2] $\Box\!\!\rightarrow$ will be used for the counterfactual conditional; and $\Diamond\!\!\rightarrow$ for 'might' counterfactuals.

[3] CEM is equivalent to (unrestricted) right-to-left direction of the equivalence: $(A \mathrel{\Box\!\!\rightarrow} \neg C) \Leftarrow \neg(A \mathrel{\Box\!\!\rightarrow} C)$. (a) Suppose $\neg(A \mathrel{\Box\!\!\rightarrow} C)$. CEM $(A \mathrel{\Box\!\!\rightarrow} C) \vee (A \mathrel{\Box\!\!\rightarrow} \neg C)$ and disjunctive syllogism allows us to conclude $A \mathrel{\Box\!\!\rightarrow} \neg C$, as required. (b) Consider the following instance of the law of excluded middle: $(A \mathrel{\Box\!\!\rightarrow} C) \vee \neg(A \mathrel{\Box\!\!\rightarrow} C)$. The right-to-left branch of the equivalence allows us to replace the second disjunct with $A \mathrel{\Box\!\!\rightarrow} \neg C$. But this is then just CEM.

The left-to-right direction of the equivalence *for possibly true antecedents* can be derived simply by appeal to a standard feature of conditional logics: so-called *conditional non-contradiction* (CNC):$\neg(A \mathrel{\Box\!\!\rightarrow} C \vee A \mathrel{\Box\!\!\rightarrow} \neg C))$. (a) Suppose we have $A \mathrel{\Box\!\!\rightarrow} \neg C$. Then if $A \mathrel{\Box\!\!\rightarrow} C$ holds we have $(A \mathrel{\Box\!\!\rightarrow} \neg C) \wedge (A \mathrel{\Box\!\!\rightarrow} C)$. But for contingent A this conflicts with the relevant instance of *conditional non-contradiction* (CNC). (b) Consider the following instance of the law of non-contradiction: $\neg((A \mathrel{\Box\!\!\rightarrow} C) \wedge \neg(A \mathrel{\Box\!\!\rightarrow} C))$. Given the left-to-right direction of the equivalence, elementary logical moves give us CNC.

For a general discussion of CEM, CNC and the equivalence, see Pizzi and Williamson (2005). This includes more formal versions of the proofs just sketched, and discussion of some modifications to the standard Lewis-Stalnaker framework that can give us the equivalence in full generality.

One might consider explaining away the intuitions backing **Equiv** (Williamson, 1988). There are, however, reasons for thinking that its role runs deeper than merely the intuitive appeal of this principle. von Fintel and Iatridou (2002) connect CEM with the behaviour of conditionals under extensional quantifiers. Their work suggests the argument that follows:

Premise 1: The following are equivalent:

A. No student would have passed if they had goofed off

B. Every student would have failed to pass if they had goofed off.

Premise 2:[4] (A) and (B) can be regimented respectively as follows:

A*. [No x: student x](x goofs off $\Box\!\!\rightarrow$ x passes)

B*. [Every x: student x](x goofs off $\Box\!\!\rightarrow$ $\neg$ x passes)

Premise 3:[5] For any *F*, "[No x: Fx]Gx" is equivalent to "[Every x: Fx]¬Gx"

From these three premises, we argue as follows. By an instance of premise 3, A* is equivalent to:

C*. [Every x: student x] not(x goofs off $\Box\!\!\rightarrow$ x passes)

There is then a chain of equivalences running from (C*) back to (B*). (C*) is equivalent to (A*), which is equivalent to (A) (premise 2) which is equivalent to (B) (premise 1) which is equivalent to (B*) (premise 2). So (C*) is equivalent to (B*).

It seems to me that we can generalize each of 1-3. *Quite generally*, 'No *F* would *G* if they *H*' and 'Every *F* would fail to *G* if they *H*' seem equivalent. And by moves paralleling the above we can argue that the following pair are equivalent:

[4] Kratzer (1986) suggests that all natural language if-clauses express quantifier-restrictions rather than dyadic connectives. This view of the 'if' clauses above would motivate rejecting premise 2. von Fintel and Iatridou (2002) argue that the truth-conditions of (A) and (B) come apart from the truth-conditions predicted by viewing the embedded 'if' as a restrictor.

[5] Given classical interdefinability of quantifiers, premise 3 looks irresistible–"no x" is equivalent to "not: some x" which is in turn equivalent to "every x not:". Matters get more complex in settings such as intuitionistic logic where existential and universal quantifiers are not automatically duals under negation. Yet even in that setting $\neg\exists x\phi$ and $\forall x\neg\phi$ are always equivalent, so it's not clear whether a retreat to non-classicism would help evade the argument—even if the retreat could be motivated.

Connectedly, suppose that *a* is the only candidate for being *P*, and suppose it is indeterminate whether *a* is *P*. Suppose that we have a conception of indeterminacy under which this means that it is neither true that *a* is *P*, nor false that *a* is *P*. There is some temptation to think that 'no *x* is *P*' is true (since there's no *x* such that it is true that *x* is *P*') but 'every *x* is not *P*' is false (since there's an *x* such that it is not true *x* is *P*).

Notice, however, that the rationale given in parentheses involve the quantifiers scoping over a truth-operator, whereas no truth-operators appear in the statements whose equivalence is at issue. And in the "truth-value-gap" setting we are considering, the truth operator is non-redundant—'there is no *x* such that is true that *x* is *P*' does not express the same content as 'there is no *x* which is *P*'. On the most popular, supervaluational articulation of these ideas, the former statement may well be true, but the latter will be indeterminate. Mutatis mutandis for the argument that 'every *x* is not *P*' is false—in the supervaluational setting, it will in fact be indeterminate. So I do not see that truth-value gap indeterminacy poses a challenge to the equivalence invoked here.

[Every x: Fx]$\neg(Hx \boxright Gx)$

[Every x: Fx]$(Hx \boxright \neg Gx)$

But if these are equivalent for all choices of F, then it follows straightforwardly that $\neg(Hx \boxright Gx)$ and $Hx \boxright \neg Gx$ are coextensional.[6]

So, from the intuitive equivalence of the embedded counterfactuals in premise (1), we derive instances of the equivalence of negated and opposite conditionals. The corresponding instances of CEM follow classically in a couple of steps.

The upshot is that denying CEM comes to seem rather heroic: one needs to explain away, not only the intuitive appeal of (Equiv), but also the apparent equivalences between quantified conditional statements. This, I think, creates a strong prima facie case for CEM.

2 The might-argument against CEM

The two originators of the 'Lewis-Stalnaker' theory of conditionals differ over the standing of the equivalence and CEM: for Stalnaker (1968), CEM is sustained; Lewis (1973) argues vigorously against it. Some of these considerations depend on delicate issues in the semantics of counterfactuals.[7] But the concern in this paper is with a more direct argument that Lewis offers against CEM.

Lewis argued that CEM leads to absurd conclusions about 'might' counterfactuals (conditionals of the form: "if it were the case that *A*, it might be that *C*"). Lewis argues that such locutions, symbolized '$\diamondright$', should be analyzed as follows:

Duality: $A \diamondright C := \neg(A \boxright \neg C)$

Given this analysis of $\diamondright$, Lewis offers a short argument showing that, given CEM, 'if it were that *A*, it would be that *B*' is *equivalent* to 'if it were that *A*, it might be that *B*'. But this is absurd—if I were to go shopping tomorrow, I might go clothes shopping and I might not. It does not follow that, were I to go shopping, I would go clothes shopping and I also would not go clothes shopping.

The short argument is as follows:[8]

[6]Suppose otherwise. Then there must be some x satisfying '$\neg(Hx \boxright Gx)$' but not '$Hx \boxright \neg Gx$'. Call that object x. Then substitute '$x = a$' for F and we will have a counterinstance to the equivalence to the equivalence established above. Thus, given that the equivalence does hold, the two counterfactuals must be coextensional.

[7]The semantic standing of CEM turns on the 'uniqueness assumption' that Stalnaker includes in his semantics: that for any proposition *P*, there is a closest worlds where *P* is true (though it might be indeterminate which worlds this is.) See Bennett (2003, §73) for an overview of the debate. I believe that neither side has a decisive case in this regard, so that we must look to other considerations to determine the standing of CEM.

[8]See Lewis (1973, p.80-81), Bennett (2003, p.189).

1. $(A \mathrel{\Box\!\!\rightarrow} \neg C) \vee (A \mathrel{\Box\!\!\rightarrow} C)$ CEM.
2. $\neg(A \mathrel{\Box\!\!\rightarrow} \neg C) \supset (A \mathrel{\Box\!\!\rightarrow} C)$ Follows from 1, by definition of the material conditional.
3. $A \mathrel{\Diamond\!\!\rightarrow} C \supset A \mathrel{\Box\!\!\rightarrow} C$ From 2 and Duality
4. $A \mathrel{\Box\!\!\rightarrow} C \supset A \mathrel{\Diamond\!\!\rightarrow} C$ Duality
5. $A \mathrel{\Box\!\!\rightarrow} C \equiv A \mathrel{\Diamond\!\!\rightarrow} C$ From 3,4.

If you have Duality, you have a might-argument against CEM.

That something is wrong with Duality, at least as a general theory of 'might' counterfactuals, is suggested by the an observation from (Stalnaker, 1980). The following statement seems perfectly fine:

> If I'd left the house only five minutes before the train, I might have missed it; but I believe that if I'd left then, I'd have made it

On Lewis' theory this is straightforwardly equivalent to a Moorean paradox, and so should sound terrible. But it doesn't.

Another problem with Duality arises under certain assumptions about counterfactual logic. On Tuesday I have a fair coin. Were I to flip it, it might land heads, and it might land tails. That seems as clear as anything could be. But now suppose that we are in a world where I do flip it on Wednesday, and it in fact does land heads. Now a widespread (though not uncontroversial) principle of conditional logic is that when $A \wedge B$ is true, $A \mathrel{\Box\!\!\rightarrow} B$ is true too (this 'And-to-If' inference is, for example, endorsed by Lewis (1973)). Assuming our logic has this feature, in the setting described an utterance on Tuesday of 'Were I to flip the coin, it'd land heads' will in fact be true, even though I couldn't have known that this was so. That result is easy to learn to live with—after all, even if true, it will be common ground that the counterfactual is not assertible. But given duality, it entails the *falsity* of the Tuesday-utterance 'were I to flip this coin, it might land tails'.[9] And that seems plain wrong—this might-counterfactual seems perfectly correct, no matter whether I flip the coin, and no matter which way it lands. The believer in Duality would be well-advised to give up And-to-If.

A third problem with duality (emphasized recently by Hájek (MS) and Eagle (MS)) is that it can be used as a premise in an argument that most ordinary counterfactuals are false. The crucial premise here is that 'if I were to drop a cup, it might not hit the floor' should be construed as true (the grounds for this will be that if I were to drop the cup, according to popular interpretations of contemporary physics, there's a non-zero objective chance of it quantum-tunnelling through the floor, rather than hitting it). If such might-counterfactuals are true, then given duality, 'if I were to drop the cup, it would fall to the floor' must be false. Given the ubiquity of such tiny chances of strange things happening, it appears the same argument can be given for almost any ordinary counterfactual. So we appear to be landed in an error-theory for ordinary counterfactual judgement.

As it stands, the argument is rather weak. For why argue from the truth of the might-counterfactual, to the falsity of the ordinary counterfactual, rather than vice versa? This, I think, would have been Lewis's reaction to such arguments.[10] And in the present setting, it does no good to cite the fact that in the counterfactual setting, it would be in various senses *possible* that

[9] By duality, 'flip$\mathrel{\Diamond\!\!\rightarrow}$ tails' is equivalent to '$\neg$ (flip $\mathrel{\Box\!\!\rightarrow} \neg$ tails)'. But the embedded conditional is true, since 'flip$\mathrel{\Box\!\!\rightarrow}$ heads' is. So its negation, and thus the might-counterfactual, is false.

[10] See especially the appendices to Lewis (1979) for Lewis's discussion of the case.

the cup not hit the floor—for might-counterfactuals do not, on the duality theory, express such counterfactual possibilities. The reasoning just sketched, however, is important, for we shall see later that a rival to the duality analysis of might-counterfactuals—one designed to reinstate the might-argument against CEM—is, unlike Lewis, in no position to resist error-theory by biting bullets.

Duality seems on reflection a surprising proposal. After all, 'might' has a separate life, commonly appearing outside conditionals in phrases like 'somebody might be in the cellar'. What could be more natural than the suggestion that might-counterfactuals are the result of the combination of the independently characterized counterfactuals and mights? A natural thought is to look for a story about might-counterfactuals by searching for the right understanding of 'might', and the right understanding of how 'might's and counterfactuals combine.

The major theme in the literature, starting from Stalnaker (1980), is to respond to the might-argument in exactly this way. We can sort such responses along two dimensions (a) whether they scope the might over the conditional ('wide-scope might') or vice versa ('narrow-scope might'); (b) what meaning, or range of meanings, they assign to 'might'. Writing the modal expressed by 'might' as '$\diamondsuit$', the widescope proposal is:

Wide:[11] $A \mathbin{\Diamond\!\!\rightarrow} B := \diamondsuit(A \mathbin{\Box\!\!\rightarrow} B)$

The narrow-scope proposal is:

Narrow:[12] $A \mathbin{\Diamond\!\!\rightarrow} B := (A \mathbin{\Box\!\!\rightarrow} \diamondsuit B)$

These proposals are compatible with a variety of views on what modal 'might' expresses (or can express). The dominant view is that 'might' expresses some sort of epistemic possibility, though there is much discussion about exactly how this works. A natural proposal would be that 'might p' is true (as uttered by x) iff p is compatible with what x knows. But most think that this does not fit the data, and must be replaced by something more sophisticated.[13]

Stalnaker (1980) defends the view that certain uses of 'might' express something other than a straightforward epistemic modal of the kind just indicated. He thinks for this usage, 'might p' will only be true when complete knowledge of the subvening facts still won't tell us whether or not p. Consider, for example, a borderline red colour patch. Even full knowledge of its exact wavelength won't allow us to figure out if it is red or non-red. Bennett (2003) also claims that 'might' expresses an 'idealized epistemic' possibility: the epistemic possibilities for someone who has access to all the facts up to a certain point in history.[14] At another remove from the

[11]This is defended by Stalnaker (1980), DeRose (cite) and others.

[12]In various forms this is defended by Bennett (2003), Eagle (MS), Gillies (2007) and, surprisingly, Lewis (1979) (the latter treats it as one disambiguation of might-counterfactuals; with the other being given by not-would-not).

[13]DeRose (1991) argues that 'might p' is only true when we do not know *nor could in relevant ways find out* anything incompatible with p. See Gillies and von Fintel (2007) for a recent representative of this tradition of liberalized epistemic readings of 'might'. (Such a proposal could be combined with wide or narrow scope readings: in fact, DeRose favours the wide-scope analysis, and Gillies the narrow-scope analysis.)

[14]Bennett (2003, pp.190-191). Bennett links this to Stalnaker's appeal to a 'quasi-epistemic' modality. I'm not so sure that this is Stalnaker's reading: what Stalnaker calls an quasi-epistemic 'might' is really the dual of the supervaluationist's *Determinacy* operator. Unlike Bennett's story, there is no need for Stalnaker to restrict the idealized knowledge to facts obtaining before a certain point in time.

standard epistemic theories, Lewis (1979, appendices) argues for a narrow-scope reading of 'might' conditionals where 'might' is read as expressing a nomic modality. Eagle (MS) argues that there is a disambiguation of might on which it expresses the ontic modality standardly expressed by 'can'.

Our purpose here is not to decide between these views. They face tricky challenges but have been ably defended.[15] However, I'll be assuming that the shared idea is correct—might-counterfactuals should be analyzed as the upshot of independent theories of modals and conditionals. If duality is right at all, it needs to be argued for as a consequence of a some such story, not put forward as an analysis of might-counterfactuals as an idiom.

3 Bennett's might-argument

Lewis didn't posit duality on a whim: 'Were I to run for the train, I might miss it, but were I to run for it, I would get it' sounds terrible. A natural and straightforward thought is that the conjunction sounds bad because the conjuncts are inconsistent with one another. And if '$P \boxright Q$' and '$P \diamondright \neg Q$' are in general inconsistent, duality holds.

The data is real, but we should try to explain it without positing semantic inconsistency between the conjuncts. A well-trodden path for explaining such data is to look for *pragmatic* rather than *semantic* explanations.

Consider, for example, 'it might not be that I will catch the train, but I will catch the train'. This sounds just as bad as the clash between might and would counterfactuals to which the friend of duality points. But in this unconditional case, few would want to maintain that $\diamond\neg P$ and P are inconsistent—i.e to think that Q follows from $\diamond Q$—and so we could quickly derive a contradiction from the seemingly innocuous statement 'I might miss the train and I might not'. The overwhelmingly plausible option in this latter case is that the badness of the conjunction derives from some sort of pragmatic badness in *asserting* 'I catch the train' straight after one asserts 'I might not catch the train'.

Asserting $P \wedge \diamond\neg P$ sounds just as bad as asserting '$P \boxright Q \wedge P \diamondright \neg Q$. The temptation is to give a uniform explanation of these two might-involving unassertible conjunctions. But explaining the badness in terms of inconsistency is simply not an option in the case of the unconditional clashes. Uniformity warns us against positing semantic inconsistency in the conditional case too.

[15]See Gillies (2007) for a series of puzzles to both wide and narrow-scope epistemic readings of might-counterfactuals.

One crucial bit of kit for defenders of narrow-scope epistemic 'mights' is the account that Gillies defends, whereby epistemic modalities in the consequents of conditionals are all restricted to possibilities compatible with the information presented in the antecedent of the conditional. Without this, such accounts would be vulnerable to a certain recipe for constructing counterexamples. To see this in the crudest case, assume that 'might p' is true iff p is compatible with what the speaker knows. Consider

> If $\neg p$ and I don't know that $\neg p$, then it might be that p

This always sounds wrong. But by construction,

> If $\neg p$ and I don't know that $\neg p$, then it's compatible with what I know that $\neg p$

is clearly right. The pattern is easily generalized.

The friend of duality might complain that the parallelism between conditional and unconditional 'might' clashes is merely superficial—not strong enough to motivate a uniform treatment. But parallels continue. For example, we earlier noted Stalnaker's observation that when we insert 'I am pretty confident that' into the middle of our bad-sounding conjunction of 'might' and 'would' conditionals, we get something that sounds ok: 'If I were to run, I might miss the train, but I'm pretty confident that If I were to run I'd catch it'. Duality seems to predict, wrongly, that such statements are of a Moore-paradoxical form, and thus unassertible. Inserting 'confident' into the unconditional version seems to pattern the same way: 'I might miss the train, but I'm confident that I'll catch it' seems perfectly ok.[16] The data just presented is hardly decisive, but it does make a uniform account of conditional and unconditional clashes seem attractive.

We shouldn't be overhasty, however. First, the most we get from Stalnaker's data is the non-contradictoriness of the relevant 'might' and 'would' counterfactuals *on at least one reading of 'might'*. To remove the threat of the might-argument altogether, we'd need that there's *no* reading of 'might' that sustains the inference. And it's not clear that we get anything like that simply from considering the above (Lewis himself, recall, thought there was *one* reading of might-counterfactuals which made them consistent with the opposite would-counterfactual: the narrow-scope nomic reading). Second, one does not need the full force of Lewis's analysis to get the might-argument up and running. Some among the rival analyses of 'might' may, for all we have so far seen, support the inferential move from '$A \diamond\!\!\rightarrow \neg B$' to '$\neg(A \,\Box\!\!\rightarrow B)$'. So we need to look in further detail.

Bennett explicitly claims that his favoured analysis of might-counterfactuals sustain the argument against CEM. Indeed, he holds something much stronger than this—that his account will always agree with Lewis's on the truth-values of 'might'-counterfactuals:

> Happily, although [the revised account of 'might' counterfactuals] drops Lewis's account of what $A \diamond\!\!\rightarrow C$ means, it does not deprive him of his 'might'-using argument against CEM. The argument still stands, because on any reasonable understanding of subjunctive conditionals $A \diamond\!\!\rightarrow C$ will be true in exactly the cases where $\neg(A \,\Box\!\!\rightarrow \neg C)$ is true. Thus, my definition and Lewis's are equivalent...
> (Bennett, 2003, p.192)[17]

Bennett does not explain why 'any reasonable' theory should have this result, but the idea is a suggestive one. Unlike Lewis's might argument, it does not treat might-counterfactuals as an idiom, but as involving narrow-scope modals. And, if combined with the view that this is just *one* disambiguation of the epistemic 'might' in the consequent, then Bennett needn't be troubled by Stalnaker's observation that we don't get the Moorean paradoxes predicted by duality. For he may reply that '$A \diamond\!\!\rightarrow \neg B \wedge$ I believe that $A \,\Box\!\!\rightarrow B$' is consistent *when the might-counterfactual is read with some less-than-fully-idealized epistemic possibility*.[18]

It is striking that Bennett's claim has even the *chance* of being true—none of the other options that we have considered makes anything like this remotely plausible (so far as I can tell).

[16] A tempting diagnosis in the unconditional case is that one should only assert p when one knows that p; but so that when one asserts 'p and might($\neg p$)' one inter alia represents that $K(p \wedge M\neg p)$ which entails $K(p) \wedge M(\neg p)$. But plausibly $M(\neg p)$ entails $\neg K(p)$. So in asserting the above, one is representing both that $K(p)$ and that $\neg K(p)$. Inserting 'confident that' blocks this inference: one represents that one is *confident* that one knows that p; but that's not the same as representing directly that one knows. See Williamson (2000, ch.7).

[17] I have altered the symbols to accord with those used in this paper.

[18] I'm not sure whether Bennett himself would accept the ambiguity of 'might' in might-counterfactuals.

Consider wide-scope readings. The might-counterfactuals claim that a certain counterfactual—$A \boxright \neg B$—is *possibly* true in some relevant sense. But what reading of the modal involved would allow us to infer from this to the *actual* truth of $\neg(A \boxright B)$? Epistemic possibility, idealized epistemic possibility, non-definite-falseness, alethic possibility: none of these make the inference tempting.

Something similar goes for many narrow-scope analysis of mights, which seem if anything even more easily compatible with the opposite would-counterfactual (so easy, in fact, that a major challenge facing such accounts is to explain why we feel the clash between the two).[19] However, Bennett's interesting suggestion is that a narrow-scope epistemic 'might' *if it's sufficiently idealized* will do the job. And there's some intuitive appeal to this idea. Suppose that, were I to drop the plate, not even an ideal knower, with access to all the facts up until the plate-dropping, would know that I was about to get injured. Given this, it's certainly not obvious that the counterfactual—were I to drop the plate, I'd get injured—can be true. So Bennett's claims deserve careful attention from friends of CEM. For to repeat: if even one disambiguation of ordinary 'might'-counterfactuals is incompatible with the corresponding 'would', we can use it to run the might-argument against CEM.

4 Bennett's Hypothesis

Bennett's idea can be broken into two parts. The first is a suggested reading of 'might' counterfactuals, and the second a certain claimed equivalence. Let '$\Diamond$' stand for his idealized epistemic reading of 'might':

Bennett's Analysis
$(A \Diamond\!\!\rightarrow B) \Leftrightarrow (A \boxright \Diamond B)$

Bennett's Hypothesis
$(A \boxright \Diamond B) \Leftrightarrow \neg(A \boxright \neg B)$

However, without some restrictions, Bennett's Hypothesis seems refutable. For let A be the conjunction, $\neg p \wedge \Diamond p$. The following seem undeniable:

- $A \boxright \neg p$
- $A \boxright \Diamond p$

But the latter, by Bennett's Hypothesis, entails:

- $\neg(A \boxright \neg p)$

[19] See Eagle (MS) for a review of some options.

But now we have a contradiction.[20]

Granted that Bennett's Hypothesis cannot hold unrestrictedly, one might wonder if it can hold in a restricted fashion—say where the antecedents don't themselves involve tricky modals. To investigate this, consider what would seem to be pair of paradigm cases. Suppose, for example, the p is 'I flip a fair coin', and q is 'the coin lands heads'. Two relevant instance of Bennett's Hypothesis tell us that:

$$(p \boxright \Diamond q) \Leftrightarrow \neg(p \boxright \neg q)$$

$$(p \boxright \Diamond\neg q) \Leftrightarrow \neg(p \boxright q)$$

We could still get an argument against CEM on the basis of this pair. Appealing to Bennett's Analysis, from the above we get two instances of Duality. If we are then allowed to assume $p \boxright \neg q \vee p \boxright q$, we can run two instances of Lewis's schematic argument, obtaining as a result the pair:

$$p \diamondright q \Leftrightarrow p \boxright q$$

$$p \diamondright \neg q \Leftrightarrow p \boxright \neg q$$

But these are absurd. From the obviously true: 'were the fair coin flipped, it might land heads, and it might land tails' they would allow us to derive the obviously false: 'were the fair coin flipped, it would land heads and it would land tails'. We don't need the full power of Bennett's Hypothesis to get a might-argument against CEM—granted Bennett's Analysis, each instance gives an instance of Duality, and so even if Bennett's Hypothesis only holds in a small range of paradigmatic cases, we still have a 'might'-argument against CEM.

It is exactly such instances of Duality that we earlier found to be in tension with the And-to-If inference characteristic of several counterfactual logics, including Lewis's. The moral seemed there to be that the believer in Duality cannot afford to endorse And-to-If—and the same result carries over one who endorses Bennett's Analysis and Hypothesis.[21] In fact, when we look at the case, it seems that appealing to Bennett's Analysis is inessential to making this point. For even an ideal knower in a situation where a fair coin is about to be flipped couldn't know whether it was going to land heads or tails. So, forgetting about might-counterfactuals, we certainly seem to have the conjunction $p \boxright \Diamond q$ and $p \boxright \Diamond\neg q$. But now, by two instances of Bennett's Hypothesis, we get $\neg(p \boxright q)$ and $\neg(p \boxright \neg q)$. Now we can make the point against the And-to-If inference straightaway—for in a situation in which I flip the coin, and it turns up heads, we get $p \boxright q$, explicitly contradicting the above. Independently of consideration of epistemic modals, the believer in Bennett's Hypothesis must give up And-to-If.

[20]This observation is closely related to the pattern of objections to narrow-scope readings of 'might' earlier noted. The suggestion there, drawn from Gillies, was that the epistemic 'mights' in the consequent were contextually restricted, with the material given in the antecedent playing a role in setting the relevant context. Be this as it may, what the above observation shows is that no modality $\Diamond$ which sustains Bennett's hypothesis unrestrictedly—it doesn't matter if the counterexample that shows this isn't expressable by a natural language 'might' conditional.

[21]Bennett (2003, §93)—unlike Lewis—does in fact reject And-to-If.

The idea behind Bennett's Hypothesis—which we can roughly summarize as the thought that counterfactuals are epistemically constrained—taken in isolation already has strong consequences for the shape of counterfactual logic. This can be extended. One can argue that Bennett's Hypothesis *by itself* is inconsistent with conditional excluded middle. For take a case where even an ideal agent with exact knowledge of the facts up to an A event, wouldn't know whether or not B. Then (independently of any proposal to *analyze* might-counterfactuals in such terms) we have:

$$(A \mathrel{\Box\!\!\rightarrow} \Diamond B) \wedge (A \mathrel{\Box\!\!\rightarrow} \Diamond\neg B)$$

And from this and the relevant instance of CEM, we have:

$$((A \mathrel{\Box\!\!\rightarrow} \Diamond\neg B) \wedge (A \mathrel{\Box\!\!\rightarrow} B)) \vee ((A \mathrel{\Box\!\!\rightarrow} \Diamond B) \wedge (A \mathrel{\Box\!\!\rightarrow} \neg B))$$

But two instances of Bennett's Hypothesis make each disjunct inconsistent. So CEM is incompatible with Bennett's Hypothesis—at least if his analysis of 'might' is to be non-trivial.

What this suggests is that issues to do with how epistemic modals interact with counterfactuals are ultimately irrelevant to Bennett's argument against CEM. Without Bennett's Hypothesis, there is no might argument against CEM from Bennett's Analysis alone. With Bennett's Hypothesis, we already have an inconsistency with CEM whether or not we analyze might-counterfactuals in the way Bennett favours or in some completely different way.

While our observation does not give us reason independent of CEM to think Bennett's Hypothesis false, it does refocus our attention. Bennett's argument is not at heart a might-based argument against conditional excluded middle, but an argument against conditional excluded middle based on the view that true counterfactuals are epistemically constrained in a certain way—the way articulated by Bennett's Hypothesis.

Independent considerations against Bennett's Hypothesis

Let us consider the merits of Bennett's Hypothesis in its own right. Even for those who see no merit in the idea that actual truths must be possibly knowable, the idea that counterfactual truth can't outrun ideal knowability seems interesting to explore.

I will argue, in what remains of this paper, that the epistemic constraint on counterfactual truth imposed by Bennett's Hypothesis will commit one to an error-theory about ordinary counterfactual judgements. The price of accepting such an error theory, I urge, is so high that whatever intuitive appeal Bennett's Hypothesis may have vanishes.

Take any ordinary counterfactual, say, 'if I were to drop this plate, it'd fall to the floor'. First of all, step back and reflect how disastrous it would be if all such statements were false. Error-theories in general should be avoided where possible, I think; but an error-theory concerning counterfactuals would be especially bad. For counterfactuals are one of the main tools of constructive philosophy: we use them in defining up dispositional properties, epistemic states, causation etc. An error-theory of counterfactuals is no isolated cost: it bleeds throughout philosophy.[22]

[22] See Hájek (MS) for a defense of an error-theoretic reading.

However, contemporary physics suggests that there is a tiny but non-zero chance that the plate, when dropped, flies off sideways.[23] Unless we can counterfactually exclude highly atypical outcomes of chance processes, then almost all ordinary counterfactual judgements will be false.

There is considerable pressure, therefore, to develop a theory of counterfactuality which allows atypical (though not chance-0) outcomes to be excluded. Lewis (1979, postscripts) sketches one attempt to provide such an account. (Williams,forthcoming) presents an alternative account with the same feature.[24]

Such theories have a certain side-effect: they say that if I tossed a fair coin a billion times, it would not have landed heads each time.

(A) 10^9 flips $\Box\!\!\rightarrow \neg$ all-heads

One might quibble with the details of these theories of counterfactual similarity, but quite generally it is plausible that any theory that makes (A) false, will have a hard time avoiding making ordinary counterfactual judgements true. The relevant facts seems so similar—the antecedents counterfactually imply that the consequents are overwhelmingly probable, but not that it is absolutely certain that they will occur. Moreover, in each case, the outcomes which cause the trouble—the chance that a plate will quantum-tunnel through the floor, or that a coin will land heads billions of ties in a row—seem similarly structured. In each case, though the threatening outcome is not impossible, it involves a strikingly atypical conspiracy and apparently coordinated series of outcomes to occur. No wonder that the extant theories of such chancy counterfactuals treat the two cases in parallel.

I shall assume, therefore, that if we are to avoid an error theory of ordinary counterfactuals, counterfactuals such as (A) must also be true. One who wishes to resist at this point is invited to come up with a plausible theory of counterfactuals and chance that *doesn't* have this consequence. In any case, it should at least be conceded that the reasons here offered to think that (A) is true are independent of any of the considerations hitherto cited in this paper—CEM, the

[23] The GRW interpretation of quantum mechanics will give us this result, for example. But quantum theory is not obviously necessary: the chances invoked by statistical mechanics will have this result. Both these results are of course, disputable.

[24] The latter proposal that 'all-heads' proposals in the relevant sense do not *fit* with the chancy laws, because they are *highly non-random* by the lights of those laws, in roughly the sense described by Elga (2004).

Lewis would see the 'all-heads' outcome as a *quasi-miracle*—a remarkable, low-probability event—and for him this would make for dissimilarity between that world and the actual world, allowing us to counterfactually exclude it. Lewis's theory is problematic, though: see Hawthorne (2005) and Williams (ming).

Bennett (2003, §96) discusses the kind of cases we are considering. His discussion seems to me not to mesh well with his earlier material on might-counterfactuals. Rather than devising a closeness ordering that makes atypical worlds farther away, as Lewis does, he canvasses altering the truth-clauses for counterfactuals so that (at least when the antecedent is false) their truth requires only truth at most of the closest antecedent-worlds, rather than truth at all of them.

He also (p.250) endorses the consequence of this combined with his acceptance of duality—the corresponding 'might' counterfactuals are, surprisingly, false. But given his specific narrow-scope epistemic understanding of might-counterfactuals, this will turn into a specific thesis about what an ideal knower could or couldn't know—and I will argue below that, contra Bennett, we must count might-counterfactuals under this reading as true.

Bennett's positive 'near miss' proposal (and an associated assertibility-theoretic proposal which he also considers) both seem to me much more problematic than a (suitably patched) version of Lewis's proposal for handling the cases. But I will not pursue this here.

analysis of epistemic modals, or Bennett's Hypothesis. So it is a reasonable starting point to get independent traction on such principles.

To evaluate Bennett's Hypothesis, we can consider what it consequences it would allow us to draw from (A)'s truth. Now, Bennett's hypothesis tells us that the counterfactuals:

$$A \mathrel{\Box\!\!\rightarrow} B$$

$$A \mathrel{\Box\!\!\rightarrow} \Diamond(\neg B)$$

cannot be true together. So if (A) is true, Bennett's Hypothesis tells us that the following *cannot* be true:

(B) 10^9 flips $\mathrel{\Box\!\!\rightarrow} \Diamond$(all-heads)

To argue from the truth of (A) against Bennett's thesis, then, it suffices to argue for the truth of (B).

The case for (B) is three-fold. (i) it is the intuitively correct verdict. (ii) what (B) describes is a counterfactual version of a lottery predicament, where consensus has it that agents fail to have the relevant knowledge, supporting the truth of (B). (iii) Theoretical principles that support the standard line on lottery cases will also entail (B).

On the first point, I take it that (B) is overwhelmingly intuitive. The idea that an ideal knower (in the relevant sense) could epistemically rule out an 'all-heads' outcome seems just wrong. What it seems right to say is that the known chances provide the ideal knower with plenty of evidence that all-heads won't come about—but this evidence falls short of supporting knowledge.

On the second point, it is clear, I think, that the situation described is just another case of the epistemic predicament with lotteries. Where the judgement that one's ticket is a loser is based solely on the known improbability that it will win, the consensus is that the judgement doesn't count as knowledgable. In the current setting, a judgement that the coin will not land heads on every occasion is to be made on the basis simply of the known improbability and atypicality of an all-heads result. If we trust the consensus about lottery cases, then the application to the current case has it that (B) is true.[25]

Now, it might be that *all* lottery intuitions are ill-founded. Perhaps those who argue that admitting ignorance in paradigm lottery cases forces a more general scepticism are right. If so, then the anti-knowledge intuitions in the coin-flipping version just sketched would be in bad shape. But I think, as with other sceptical paradoxes, we are entitled to stick with the original verdicts until we are convinced that best theory requires us to do otherwise. The intuitive data is that *purely probabilistic* justifications cannot lead to knowledge; and this applies directly to the putative judgement that the result of the coin-flips won't be all-heads. Reconciling this with the data that *other* kinds of justification—perceptual, inductive etc—can lead to knowledge is an interesting task, but we should believe pro tem that the two can be reconciled.

[25]Our 'billion coin-flip' case is equivalent to the following lottery: we each have a ticket inscribed with a string of a billion 1's and 0's. A random-number generator will produce a string of a billion 1's and 0's, and you win iff your ticket matches the generated sequence.

On the third point. The case that even the ideal knower lacks knowledge that the result won't be all-heads can be given theoretical backing by appeal to *sensitivity* constraints on knowledge. Such constraints tell us that one can only know that p if, were p to be false, one wouldn't believe it. Since the probabilistic basis for a judgement that the coins won't all land heads is the same no matter what the result in fact is, then if the ideal knower judged that the coin won't land heads every time in one case, she will so judge in all cases—and in particular, she will so judge in the closest case where the coins do land all-heads. So the ideal knower's belief violates the sensitivity condition on knowledge.

It might be replied that other modal conditions on knowledge play differently. For example, the safety condition requires that throughout the close cases where the judger believes that p, that belief be true. And now it is relevant that the case where the coin lands all-heads is counterfactually further away than the cases where more typical results obtain (by the lights of the theories of closeness that makes (A) true). So even if the ideal knower's putative belief that the coin won't land all-heads is insensitive, it may yet be safe. On the one hand, if one were convinced that safety was the *only* relevant condition on knowledge, we might be able to deny (B). On the other hand, if the safety constraint on knowledge makes a distinction without a difference between paradigm lottery cases and the coin-cases here considered, so much the worse for it.[26]

None of the points just raised are knock-down—and neither was the case for (A)'s truth, which was the other part of the argument against Bennett's hypothesis.[27] But just as in that previous case, there is a standing challenge to one who wants to defend Bennett's Hypothesis, to develop an epistemology that makes the relevant discriminations between standard lottery cases and the case at hand.

And as with the argument given for (A) being true, the case for (B) is independent of any thought of CEM, Bennett's Hypothesis, or the analysis of epistemic modals. So both (A) and (B) can be defended independently of the main dialectic. As we have already seen, together they undermine Bennett's Hypothesis.

We can summarize the argument as follows. First, (B) holds—were a coin to be flipped a billion times, even the ideal agent wouldn't be able to exclude the possibility of it landing heads every times. Second, under the supposition that Bennett's Hypothesis holds, this entails that (A) is false: it is not the case that, if we flip the coin a billion times, it would land heads every time. But finally, counterfactuals like (A) just make obvious a structure that contemporary science

[26]For there to be a safety-based argument against knowledge in the first case, one must use a formulation of safety that requires one's belief to be true in *all* the closest possible worlds, not just *most* of them. See Pritchard (forthcoming, §6) for discussion, defence and references.

Note that the safety-based approach has been criticized in the past for its alleged inability to handle lottery intuitions correctly (Greco, 2002), and contrasted unfavourably with sensitivity-based accounts in this respect. Just so, the suggestion in the text is that it is a point of *criticism* of a pure safety account of knowledge, that it would return the result that the ideal knower could know the result was not going to be all-heads. This criticism is now effective even against the amended definition of safety suggested by (Pritchard, ming).

[27]One might be tempted to offer a direct argument in support of the idea that the ideal knower can know that the outcome will not be all-heads. We are granting that the counterfactual '10^9 flips $\Box\!\!\rightarrow \neg$ all-heads' is true. Shouldn't the ideal knower (in circumstances where the antecedent holds) know this fact? But then, since she will know the antecedent, she should be able to apply *modus ponens* and know that the consequent holds?

However, the counterfactual is certainly *false* in the case where the coin is flipped and in fact lands heads every time. If the ideal knower doesn't know he isn't in this situation, he doesn't know the counterfactual is true. The epistemic predicament is not improved by shifting attention from the consequent to the counterfactual itself.

seems to tell us underlies almost all ordinary counterfactuals—that the antecedent will make the consequent highly probable, but not certain. In the light of this, it is reasonable to believe that any *non*-error theoretic account of counterfactuals will make (A) true, conflicting with the predictions of Bennett's Hypothesis. The suspicion that (A) and ordinary counterfactual judgements will be treated in parallel is borne out by extant accounts (both error-theoretic and non-error-theoretic) of such counterfactuals. Overall, the defender of Bennett's Hypothesis has plenty of work to do if they are to stave off the threat of an error theory.[28]

5 Conclusion

Counterfactual excluded middle is sometimes treated as a curiosity, and folklore, I think, has treated it as the clear loser in the face of Lewis's 'might' arguments. But Lewis's argument against CEM relies on the Duality analysis of might-counterfactuals that looks far less attractive than competing accounts that analyze them as the interaction of independently motivated accounts of epistemic modals and counterfactuals. However, if Bennett's Analysis and Hypothesis were right, then Lewis' might-argument against CEM goes through.

Indeed, we have seen that Bennett's Hypothesis alone would do the job—not a might-argument against CEM this time, but an argument based on a general principle that counterfactual truth should be epistemically constrained.

But Conditional Excluded Middle is no mere curiosity—*prima facie* it has an important role to play within an account of the sort of natural language phenomenon highlighted at the beginning of this paper. And absent further explanation, Bennett's Hypothesis threatens an error-theory of ordinary counterfactual judgements.

The reasonable procedure at this point, I contend, is to accept conditional excluded middle, and deny Bennett's Hypothesis. With the denial of Bennett's Hypothesis goes the last best hope of a successful might-argument against conditional excluded middle.

[28]This result can be viewed as a generalization of an argument that Hájek (MS) uses to argue *for* an error theory. That argument works from (i) duality; and (ii) the truth of the relevant might-counterfactuals, e.g. if a coin had been flipped a billion times, it might have landed heads every time. We have replaced (i) with Bennett's Hypothesis, and (ii) with the truth of counterfactuals about the lack of knowledge of ideal agents. Against Hajek, one might be tempted to 'bite the bullet' and simply deny (ii). But as just emphasized, the analogous move here lands one is to take a contentious stance in general epistemology.

Of course, given Bennett's Analysis, the truth of (B) will secure the truth of

(B*) If the coin were flipped a billion times, it might have landed heads every time

I'm inclined to think that (B*) is true, and so I think this is a good feature of Bennett's Analysis. (Of course, without Bennett's Hypothesis, the Analysis alone can't serve to undermine CEM.)

Some theories cannot deliver (B*)'s truth, and this I take to be a strike against them. For example, Stalnaker's 'quasi-epistemic' wide-scope reading of the might-counterfactual will only be true when the corresponding counterfactual (A) is indeterminate (Stalnaker, 1980). And since (A) is determinately *true* by our lights, this is not a true reading.

What of epistemic account of 'might' more generally? On the wide-scoped reading, this is to ask: is (A) knowably true (whether by you or me, or by reasonable methods at our disposal; or by some idealized knower)? At least when we know that the antecedent is false, surely we're at least in a position to know the counterfactual is true. So it's hard to see a stable case for (B*) being true. So if, in general, narrow-scoped readings can make (B) true, then this constitutes an interesting advantage for such accounts over their wide-scoped epistemic rivals.

References

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford.

DeRose, K. (1991). 'Epistemic possibilities'. *The Philosophical Review*, **100**, 581–605.

Eagle, A. (MS). 'Might counterfactuals'.

Elga, A. (2004). 'Infinitesimal chances and laws of nature'. *Australasian Journal of Philosophy*, **82**, 67–76.

Fine, K. (1975). 'Vagueness, truth and logic'. *Synthese*, **30**, 265–300. Reprinted with corrections in Keefe and Smith (eds) *Vagueness: A reader* (MIT Press, Cambridge MA: 1997) pp.119-150.

Gillies, T. (ms. 2007). 'Iffiness: very drafty draft'.

Gillies, T. and von Fintel, K. (MS June 2007). 'Might made right'.

Greco, J. (2002). 'Virtue and luck, epistemic or otherwise'. *Metaphilosophy*, **34**, 353–66.

Hájek, A. (MS.). 'Most counterfactuals are false'.

Hawthorne, J. (2005). 'Chance and counterfactuals'. *Philosophical and Phenomenological Research*, pages 396–405.

Kratzer, A. (1986). 'Conditionals'. *CLS*, **22**(2). Reprinted in Arnim von Stechow and Dieter Wunderlich, eds., Semantics: An international handbook of contemporary research. Berlin: de Gruyter, 1991, 651-656.

Lewis, D. K. (1973). *Counterfactuals*. Blackwell, Oxford.

Lewis, D. K. (1979). 'Counterfactual dependence and time's arrow'. *Noûs*, **13**, 455–76. Reprinted with postscript in Lewis, *Philosophical Papers II* (Oxford University Press, 1986) 32–51. Also reprinted in Jackson (ed) *Conditionals* (Oxford University Press, 1991) 46-76.

Pizzi, C. and Williamson, T. (2005). 'Conditional excluded middle in systems of consequential implication'. *Journal of Philosophical Logic*, **34**(4), 333–362.

Pritchard, D. (forthcoming). Sensitivity, safety and anti-luck epistemology. In J. Greco, editor, *The Oxford Handbook of Scepticism*. Oxford University Press.

Stalnaker, R. (1968). 'A theory of conditionals'. In *Studies in Logical Theory*, pages 98–112. Blackwell, Oxford.

Stalnaker, R. (1980). 'A defense of conditional excluded middle'. In Harper, Stalnaker, and Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance and Time*. Kluwer Academic Publishers, The Netherlands.

von Fintel, K. and Iatridou, S. (MS 2002). 'If and when if-clauses can restrict quantifiers'. Available at `http://web.mit.edu/fintel/www/lpw.mich.pdf`.

Williams, J. R. G. (forthcoming). 'Chances, counterfactuals and similarity'. *Philosophy and Phenomenological Research*.

Williamson, T. (1988). 'Bivalence and subjunctive conditionals'. *Synthese*, **75**, 405–421.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press, Oxford.