Logique & Analyse **242** (2018), 119-124

MIGHT/WOULD DUALITY AND THE PROBABILITIES OF COUNTERFACTUALS

Michael J. Shaffer

Abstract

In this paper it is argued that Lewis' account of might counterfactuals and his account of the probabilities of counterfactuals lead to a result that is at odds with the way in which might counterfactuals operate in ordinary language.

Keywords: might, would, counterfactuals, conditional, probabilities, David Lewis

Lewis (1973a and 1973b) famously defended the following analysis of the might conditional (i.e. "Might/Would Duality"):

(MWD) $\neg(p \mathrel{\Box\!\!\rightarrow} \neg q) \equiv p \mathrel{\Diamond\!\!\rightarrow} q$.

MWD is both an interesting and prima facie plausible principle about might and would counterfactuals. Following Adams (1965 and 1975), many logicians have also entertained the interesting and prima facie plausible view that the probability of a conditional is a conditional probability (see, for example, Bennett 2003, Hájek 1994, Hájek and Hall 1994 and Arló-Costa 2014):

(CCCP) $P(p > q) = P(q \mid p)$ for all p, q in the domain of P such that $P(p) > 0$.[1]

Here > is being used to represent a generic conditional that includes indicatives. So understood CCCP does not necessarily have an obvious application to counterfactuals. More importantly, Lewis (1976) explicitly rejected CCCP with respect to many types of conditionals and suggested that the probabilities of these conditionals should rather be understood as policies for *feigned* minimal belief revision. On this view, the probability of such a conditional should be understood to be the probability of the consequent, given the minimal revision of $P(\cdot)$ that makes the probability

[1] One reason that this topic is important is because the probabilities of conditionals can be used as a criterion for the assertability of conditionals. On such views a conditional proposition is assertable to the degree that it has a high probability. Lewis himself (1976) endorses this view.

doi: 10.2143/LEA.242.0.3284746.

of the antecedent of the conditional equal to 1. Formally, Lewis (1981) understands imaging as follows:

(IMAGE) $P(p > q) = P'(q)$, if p is possible.

Here $P'(\cdot)$ is the minimally revised probability function that makes $P(p) = 1$. Lewis tells us that we are to understand this expression along the following lines. $P(\cdot)$ is to be understood as a function defined over a finite set of possible worlds, with each world having a probability P(w). Furthermore, the probabilities defined on these worlds sum to 1, and the probability of a sentence, p for example, is the sum of the probabilities of the worlds where it is true. In this context the image on p of a given probability function is obtained by 'moving' the probability of each world over to the p-world closest to w. Finally, the revision in question is supposed to be the minimal revision that makes p certain. In other words, the revision is to involve only those alterations necessary for making $P(p) = 1$. What is interesting is that Lewis (1976, 308-312) explicitly believes that IMAGE correctly applies to Stalnaker conditionals and that Stalnaker's account of conditionals is basically correct for a wide range of counterfactuals. This very strongly implies that Lewis is committed to a version of IMAGE as an account of the probabilities of at least some *counterfactuals*, specifically those that conform to Stalnaker's account of the logic of conditionals. This is because Stalnaker's theory is actually a special more-restricted case of Lewis' theory of counterfactuals. Lewis' theory involves a well-ordering of all possible worlds while Stalnaker's theory involves only a weak total ordering of possible worlds. This then gives rise to the crucial point where the theories differ. Stalnaker's theory assumes the limit and uniqueness assumptions. The details of the limit assumption are not important here, but the uniqueness assumption can be stated as follows:

(uniqueness) for every world i and proposition A there is at most one A-world minimally different from i.

The uniqueness assumption is what effectively rules out ties in the similarity of worlds. There cannot be two worlds that are equally similar to a given possible world. Stalnaker (1981) admits that this is an idealization that he has made with respect to the semantics of counterfactuals, specifically with respect to the selection function.[2]

[2] The argument introduced here raises some interesting possibilities with respect to Lewis' theory of counterfactuals. Specifically, if uniqueness is denied, then there are many closest worlds to any given world and it is not entirely clear how IMAGE could be applied to Lewis' own more general account of counterfactuals. One might simply apply IMAGE to all counterfactuals and interpret the image on p of a given probability function as 'moving' the probability of each world over to the total set of p-worlds closest to w such that the revision in question is the minimal revision that makes p certain. In a sense then one could deploy IMAGE generally to all counterfactuals while ignoring the slight differences among

However, it turns out that jointly adopting MWD and IMAGE is deeply problematic and one or both must go for Stalnaker conditionals. In order to see this let us look at what these two claims imply about the probabilities of might counterfactuals. First, following Howson and Urbach 1993, the probability calculus tells us that:

(PR) $P(\neg p) = 1 - P(p)$.

By MWD $P(p \Diamond\!\!\rightarrow q)$ is logically equivalent to $P(\neg(p \,\Box\!\!\rightarrow \neg q))$. By PR $P(\neg(p \,\Box\!\!\rightarrow \neg q))$ is equal to $1 - P(p \,\Box\!\!\rightarrow \neg q)$. Finally, applying IMAGE, $1 - P(p \,\Box\!\!\rightarrow \neg q)$ is equivalent to $P'(q)$. Thus we derive the following crucial theorem:

(PMC) $P(p \Diamond\!\!\rightarrow q) = P'(q)$.

This all looks very straightforward, but PMC seems to be deeply problematic when we take a closer look at the usage of might counterfactuals in English.

We can see this quite clearly by introducing a basic urn model as follows. In urn_1 there are 99 white balls and 1 black ball and this is the sample space for our chance set up. All draws from all urn_1 are replaced. Let D_i represent the proposition that a draw is made from urn_i, let W_i represent the proposition that a white ball is drawn from urn_i and let B_i be the proposition that a black ball is drawn from urn_i. Now consider the following claims:

(c1) If I were to draw a ball from urn_1, then it might be a white ball.
(c2) If I were to draw a ball from urn_1, then it might be a black ball.
(c3) If I were to draw a ball from urn_1, then it would be a white ball.
(c4) If I were to draw a ball from urn_1, then it would be a black ball.

c1 and c2 can be regimented as follows:

(c1) $D_1 \Diamond\!\!\rightarrow W_1$.
(c2) $D_1 \Diamond\!\!\rightarrow B_1$.

According to PMC and the description of urn_1 the probabilities of c1 and c2 are supposed to be as follows:

(Pc1) $P(D_1 \Diamond\!\!\rightarrow W_1) = 1 - P'(\neg W_1) = .99$.
(Pc2) $P(D_1 \Diamond\!\!\rightarrow B_1) = 1 - P'(\neg B_1) = .01$.

These values can be determined because both $P'(\neg B_1)$ and $P'(\neg W_1)$ are fully fixed by the constitution of urn_1. Importantly, in this perfectly ordinary chance set up the probability calculus commits us to $P(W_i \vee \neg W_i) = 1$ and

the members of the set of worlds closest to a given world. Alternatively, one might just replace IMAGE with a more general related principle that applies to the full class of counterfactuals independent of the limit and uniqueness assumptions. It would then be interesting to see if such a generalized version of IMAGE raises similar problems for Lewis' system.

since in this particular set up our sample space fixes it that $(\neg W_i \equiv B_i)$ we get $P(W_i \vee B_i) = 1$. Note too that these values for the probabilities associated with the draws are not just equal to the probabilities of the consequents independent of the antecedents and so the conditionality involved is important and makes a difference. If no ball were drawn then there would be no chance it would be white and no chance it would be black (i.e. we would have probability 0 in both cases). So, by IMAGE, if we feign a revision of belief such that $P(D_1) = 1$ (i.e. we feign that we are certain that a ball is drawn from urn$_1$), it is clear both that $P'(\neg B_1) = .99$ and that $P'(\neg W_1) = .01$. Moreover, according to IMAGE and given the urn$_1$ model the probabilities of c3 and c4 are also as follows:

(Pc3) $P(D_1 \mathbin{\Box\!\!\rightarrow} W_1) = P'(W_1) = .99$.
(Pc4) $P(D_1 \mathbin{\Box\!\!\rightarrow} B_1) = P'(B_1) = .01$.[3]

Despite the seemingly odd result that c1 and c3 have the same probability and that c2 and c4 have the same probability (thus collapsing that distinction in probabilistic contexts), this all looks to be quite straightforward and is simply a consequence of jointly endorsing MWD and IMAGE.

However, on careful inspection, the probabilities of c1 and c2 so determined are at odds with the ordinary usage of counterfactual conditionals in English. The relevant English correlates of Pc3 and Pc4 are as follows:

(Pc3) The probability that if I were to draw a ball from urn$_1$, then it would be a white ball is .99.
(Pc4) The probability that if I were to draw a ball from urn$_1$, then it would be a black ball is .01.

These two expressions and their associated probabilities seem reasonable. But, this does not seem to be true in the case of the relevant English versions of Pc1 and Pc2:

(Pc1) The probability that if I were to draw a ball from urn$_1$, then it might be a white ball is .99.
(Pc2) The probability that if I were to draw a ball from urn$_1$, then it might be a black ball is .01.

[3] Notice here that if one takes the probabilities of conditionals to be directly related to assertability conditions for conditional propositions—as suggested in fn. 1, then it is exceptionally odd that Pc1 and Pc3 and Pc2 and Pc4, respectively, have the same probabilistic assertabilities. It should be much easier to assert Pc1 than Pc3 and it should be much easier to assert Pc2 than Pc4. This is because it is less evidentially demanding to assert a might conditional than it is to assert a corresponding would conditional. The results here show that Lewis principles fail do discriminate might and would conditionals in this regard and that Pc1 and Pc2 are incorrect.

Unlike Pc3 and Pc4, these sentences that specify the probabilities of might counterfactuals do not seem to be correct in terms of ordinary English usage.[4] The probabilities of c1 and c2 should not be .99 and .01. The probability that I *might* draw a white ball upon drawing a ball from urn$_1$ should be 1. This is because it is possible that I might do so. The same thing applies in the case of the probability that I *might* draw a black ball upon drawing a ball from urn$_1$. So, something is wrong with the principles that have been used to derive PMC. Notice that this does not depend on the constitution of the urns in terms of the proportion of white to black balls where both are contained in an urn. Upon any draw from any urn in which there are both white and black balls, the probability *that I might draw a ball of a given color* is 1. This is because it is true that these outcomes might happen in those chance set-ups. So no matter the proportion of white to black balls in any such urn it is certain that both possible outcomes, drawing a white ball and drawing a black ball, are possible outcomes. So, it appears to be the case that the English usage of might in these conditional contexts is sensitive *only* to modal factors and that in English might counterfactual 'might' is not sensitive to probabilistic considerations whereas 'would' appears to be sensitive to such factors. But this is not reflected in systems that incorporate both MWD and IMAGE. Thus, from the perspective of ordinary English usage MWD and IMAGE are incompatible and one or both of them must go as principles for Stalnaker conditionals.[5]

Stalnaker (1984, 143-145) himself suggests that MWD is false and that might counterfactuals ought to be alternatively analyzed as follows. Where $\Diamond_e$ is a suitable notion of epistemic possibility the following equivalence holds (i.e. "Stalnaker Might/Would Duality"):

(SMWD) $\Diamond_e(p \mathrel{\Box\!\!\rightarrow} q) \equiv p \mathrel{\Diamond\!\!\rightarrow} q$.

[4] These translations of the relevant conditionals into English are based on the assumption that we can move from the object level to the meta-level in order to apply the relevant meta-level principles to those English claims. I see no reason why it is not permissible to do this here, especially as the English translations are perfectly coherent.

[5] In the postscript to Lewis 1979 from Lewis 1986b — for rather different reasons — Lewis himself entertains the possibility that MWD is false and that there may be an alternate reading of some might counterfactuals. Thus he contrasts what he calls the "not-would-not" (i.e. nwn) analysis of might counterfactuals with what he calls the "would-be-possible" (wbp) analysis (lewis 1986c, 64). On this alternate analysis c1 would be analyzed as follows: If I were to draw a ball from urn$_1$, then it would be possible that it is a white ball. Similarly, c2 would be analyzed as follows: If I were to draw a ball from urn$_1$, then it would be possible that it is a black ball. Our intuitions are more consonant with this alternative reading, but Lewis officially endorses the nwn analysis captured by MWD in various works despite the apparent correctness of the wbp analysis. Moreover, he notes his reservations about this alternative (see 1986c, 63). In any case, independent of Lewis' own official commitments, there are deeply interesting questions about the acceptability of MWD that are highlighted in looking at how it might interact with a more general form of IMAGE.

This is that claim that the might counterfactual expresses the epistemic possibility that if p were then case, then q would be the case. SMWD treats might as an epistemic possibility operator on would counterfactuals. This broadly accords with the understanding of the English usage of might in probabilistic contexts suggested here. The interesting point then to be made is that SMWD gets c1 and c2 correct, unlike MWD. If one knows the constitution of the urns, then *that I might draw a ball of a given color* on any draw from any urn in which there are both white and black balls is epistemically possible and has a probability of 1. So, the considerations raised here favor SMWD over MWD as the proper analysis of at least some might counterfactuals.

References

[1] Adams, E. (1965), "The Logic of Conditionals," *Inquiry* 8, 166-197.

[2] Adams, E. (1975), *The Logic of Conditionals*. Dordrecht: Reidel.

[3] Arló-Costa, H. (2014), "The Logic of Conditionals", *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/logic-conditionals/>.

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

[5] Eells, E. and B. Skyrms, eds., (1994), *Probability and Conditionals: Belief Revision and Rational Decision*, Cambridge: Cambridge University Press.

[6] Hájek, A. (1994), "Triviality on the Cheap?" In E. Eells and B. Skyrms (1994), 113-141.

[7] Hájek, A. and N. Hall (1994), "The Hypothesis of Conditional Construal of Conditional Probability", in E. Eells and B. Skyrms (1994), 75-113.

[8] Howson, C. and P. Urbach (1993), *Scientific Reasoning: The Bayesian Approach*, 2nd ed. Open Court, Chicago.

[9] Lewis, D. (1973a), *Counterfactuals*. Harvard University Press, Cambridge.

[10] Lewis, D. (1973b), "Counterfactuals and Comparative Possibility," *Journal of Philosophical Logic* 4: 418-446.

[11] Lewis, D. (1976), "Probabilities of Conditionals and Conditional Probabilities," *Philosophical Review* 85, 297-315.

[12] Lewis, D. (1979), "Counterfactuals and Time's Arrow," *Nous* 13: 455-476.

[13] Lewis, D. (1986), "Probabilities of Conditionals and Conditional Probabilities II," *Philosophical Review* 95, 581-589.

[14] Stalnaker, R. (1981), "A Defense of Conditional Excluded Middle," in *Ifs*, eds. W. Harper, R. Stalnaker and G. Pearce. Dordrecht: D. Reidel, 87-104.

[15] Stalnaker, R. (1984), *Inquiry*. Cambridge: MIT Press.

Michael J. Shaffer
Department of Philosophy (365N)
St Cloud State University
720 4th Ave. S.
St Cloud, MN 56301