Chancy counterfactuals, redux: Response to Dodd

J. Robert G. Williams
j.r.g.williams@leeds.ac.uk

January 19, 2010

Chancy counterfactuals are a headache. Dylan Dodd (2009) presents an interesting argument against a certain *general strategy* for accounting for them, instances of which are found in the appendices to Lewis (1979) and in Williams (2008). I will argue (i) that Dodd's understates the counterintuitiveness of the conclusions he can reach; (ii) that the counterintuitiveness can be thought of as an instance of more *general* oddities arising when we treat vagueness and indeterminacy in a classical setting; and (iii) the underlying source of discontent which animates Dodd's complains is to be found in a certain general *constraint* one might impose on conditionals—what I'll call the counterfactual Ramsey bound. Unfortunately, the counterfactual Ramsey bound is just as problematic as its famous indicative cousin. The moral is that there's *no* comfortable resting place in this area; for violations of the counterfactual Ramsey bound are *going* to lead to *prima facie* surprising results.[1]

1 Background

I will present Dodd's core argument in the context of the account of counterfactuals and chance developed in Williams (2008) (see that paper for a critical evaluation of the objections to Lewis's version of the proposal developed in Hawthorne (2005)). The backdrop is a Lewisian semantics for counterfactuals:[2]

- '$A \mathrel{\Box\!\!\rightarrow} B$' is true at w iff all the closest A worlds to w are B worlds.

[1]Thanks to all with whom I've discussed this material. Special thanks to Daniel Elstein, Andrew McGonigal and Dylan Dodd for discussion of this material. The work in this paper was funded by a British Academy Research Development Award (BARDA: 53286); and was supported by the workshops and materials provided by Spanish Government grant FFI2008-06153 (MICINN).

[2]For simplicity of formulation I make the 'limit assumption' that Lewis himself would deny; nothing hinges on this.

See Williams (forthcoming) for arguments against the Lewisian version and for the Stalnakerian version. We could redo the arguments below in this setting, though it is not *immediate* that the dialectic remains the same. Dodd states that one assumption of the argument is the duality of might and would counterfactuals, guaranteed by the above dual definitions. I'm not convinced we need this, but I won't try to go through the various reformulations here.

- '$A \diamond\!\!\rightarrow B$' is true at w iff some closest A world to w is a B world.

Given formal constraints on the triadic 'closer than' relation, this fixes a logic for the counterfactual conditional. One might stop here, leaving 'closer than' primitive (or analyzing it *in terms of* counterfactuals, rather than the other way round). But Lewis (1979) goes further, giving what is intended to be an informative explanation of the 'closer than' relation in terms of non-conditional facts. The account he gives was tailored to the case where laws are deterministic; and the upshot is that to assess the closeness of worlds to *w*, we should look at (i) whether the laws of *w* are, or are not, violated at the worlds, and whether any violations there are 'local' or 'diverse and widespread'; (ii) the extent to which the worlds *exactly* match *w* across wide stretches of space-time. Lewis gives instructions for weighing up the respects of similarity against one another, and argues they match our intuitive verdicts about 'non-back-tracking' counterfactuals well enough.

The account cannot be sensibly applied to cases where the laws of *w* are chancy, as Lewis was well aware. Weird things can happen *with no violation of chancy laws*—weird enough to make true paradigmatically objectionable counterfactuals such as 'If Nixon had pressed the nuclear button in the 70's, the world would now be as it actually is'. The account needs tweaking.

The idea in Williams (2008) is to keep the underlying form of the informative treatment of closeness the same. We respond to the worries simply by figuring out a more general replacement for (i). The idea is that 'fitting with laws' in the chancy case is not simply a matter of *not violating* those laws (as it may be in the deterministic case). Worlds which are *atypical* by the lights of chancy laws *fit less well* with those laws. For example, a world where a fair coin lands heads on each one of the billions of times it is flipped is arguably *atypical* by the lights of a law that ascribes that coin a 50/50 chance of landing heads. How to analyze 'typicality' is then a main issue (it can't be just a matter of 'containing very low probability' events). This, however, is no new cost. If we like the Lewis-Ramsey treatment of laws of nature, it seems we *already* need this kind of understanding of 'fit with laws' (see (Elga, 2004)).

Let's see this in action. Suppose our laws of nature ascribe a vanishingly small, but non-zero, chance to a cup quantum-tunnelling through the floor conditionally on being dropped. In the unmodified Lewisian closeness ordering, a drop+quantum tunnelling world will involve *no great violation of actual law*. It seems it will be one among the closest set of worlds where the cup is dropped. Such worlds occupy only a tiny corner of the class of all the closest drop-worlds, but on the Lewisian truth-conditions for counterfactuals, just one such world is enough to render the ordinary counterfactual 'If the cup were dropped, it'd fall to the floor' false.[3] Nicely, the generalization of Lewis's closeness analysis can give the intuitive verdict here. It need only make the case that quantum-tunnelling events are *atypical*. Given this, they will be held to fit less well with the chancy laws than more ordinary situations; and ipso facto will be pushed further 'out' in the closeness ordering. Counterfactual error theory avoided.

The account treats structurally similar cases analogously (cases that are the principle focus of Hawthorne (2005)). For example, consider the fair coin that may or may not be flipped a billion times. According to Williams (2008) an 'all heads' result will count as atypical by the light of the chancy laws, and therefore the counterfactual 'If the coin were flipped a billion

[3] See Hájek (MS.) for arguments that this is the right verdict. I don't believe him.

times, the result would be a sequence other than all-heads'. Some might find this intuitive ('come off it, it wouldn't land *that way*!) but others are uncomfortable. Williams (2008) discusses various worries one might have about this proposal, and seeks to soothe them. Dodd now mounts a new challenge.

2 Vagueness in what is typical

It's important to realize that the notion of typicality in play in Williams' paper (and Elga's) is not given a rigourous mathematical analysis. To be sure, the accounts take inspiration from the Gaifman & Snir (1982) rigourous work on 'non-random worlds'. But that account is suitable only for sorting between *infinitely long* typical and atypical sequences.[4] The literature on randomness contains other candidates that do have application (drawing on the resources of complexity theory, for example). But what Williams urges is that these should convince us that typicality is an *real, objective phenomenon*—it should discredit knee-jerk reactions that 'typicality' is as response-dependent as Lewis's 'remarkableness' seems to be. Typicality is officially a working primitive, for Williams. That may make the account less attractive than it would be if we could take a full reductive account of the notion 'off the shelf'; but pending further elaboration, it's what we have to live with.

Given all this, we can't assume that the notion is *non-vague*. Indeed, we have every reason to think it will be vague. How long a sequence of fair coin flips is long enough for an 'all heads' result to count as atypical, for example? Not a length 1 or length 2 sequence. But if the account is to work as advertised, a sequence of a billion or so flips better count as atypical. It looks like we have the resources for a sorites series here. If a sequence of n heads isn't atypical, how can one more suddenly make the difference? If we say it can't, we'll be able to argue that no finite sequence of all-heads results should count as atypical. And that would knock the proposed fixed up account of chancy counterfactuals out of the action.

On the other hand, just *because* it seems so similar to the sorites, there's a great temptation to take the moral to be that 'typicality' is a vague notion. There are clear cases on either side, and a vague penumbra where it is *indeterminate* whether the sequence of heads is or is not atypical. Of course, using vague predicates in the analysis of closeness is no novelty—Lewis (1973, 1979) is explicit that his accounts has this feature, and argues this is only appropriate, given the undeniable vagueness of many counterfactuals.

Treatments of vagueness vary. If one must reach for a non-classical logic at this point, then of course the standard semantics and logic for counterfactuals would be called into question. But it would have been called into question much earlier, with the original Lewis proposals. Lewis, and many others, favour a classical account. For 'semantic indecision' theorists, a vague term like 'typical' has a variety of *precisifications*—perfectly sharp interpretations of the predicate that fit with whatever meaning-fixing facts constrain the extension of the term. On the simplest 'supervaluationist' reading of the proposal, claims involving 'typicality' will be true if true on all such precisifications, false if false on all such; and otherwise are 'gappy'.[5]

[4] Or more generally, world-descriptions which feature infinitely many events.

[5] See Fine (1975).

On this way of proceeding, for a range of n, 'a sequence of n flips of a fair coin, all of which come up heads, is atypical' will be neither true nor false. But the supervaluationist do not on this account have to give up the core claims of classical logic—indeed, their logic will be *supraclassical* (every classical valid sequent will be supervaluationally valid). So we don't need to drastically revise our understanding of the logic of counterfactuals on this account.

Because of this supraclassicality, the conclusion of the sorites reasoning must be embraced: there will be a least n, such that an all-heads sequence of length n is not atypical, and an all-heads sequence of length $n+1$ is atypical. As a result, the Williams account of counterfactuals will endorse:

- There is some n, such that were the fair coin to be flipped n times, the result might be all-heads; but were the fair coin to be flipped $n+1$ times, the result would not be all heads

This is simply the reflection, in counterfactual form, of the 'unpalatable existential' we can expect to extract from any sorites series, under a classical or supervaluational treatment of vagueness.

Notice, though, that there need be no true-substitution instance for this. For each choice of N, the following will either be false or a truth-value gap:

- Were the fair coin to be flipped N times, the result might be all-heads; but were the fair coin to be flipped $N+1$ times, the result would not be all heads

Again, the endorsement of the existential, and the rejection of each substitution instance, rests on nothing peculiar to counterfactuals, but on the underlying treatment of vagueness and indeterminacy that is presupposed.

Here we find a connection to Dodd's main argument. He invites us to choose $2n$ "just long enough" that a sequence of $2n$ all-heads results will count as atypical. And he then invites us to consider what we will say about a sequence of n flips—he takes it that this will be a clear non-case of atypicality. His main argument is then based on a situation where we have the following:

- (for the given choice of n) If the fair coin had been flipped n times, the result might be all-heads; but if the fair coin had been flipped $2n$ times, the result would not be all heads

But this is ambiguous between the existentially generalized reading, on which it must surely count as true, and some particular *instance* thereof. The existential generalization is true—indeed, it's truth follows, under minimal assumptions, from the truth of the $n/n+1$ case given above.

I argued that in the $n/n+1$ case, no instance of the existential generalization would be true. What of *Dodd's* case? This entirely depends on how extensive the penumbra of vagueness is.

To fix ideas, suppose that the last clear case of non-atypicality is a billion flips long, but the first clear case of atypicality is a million flips long. Then (despite the truth of the existential) there will be no instance of the above that will be true—every single instance of the above scheme will be either false or a truth-value gap. And so Dodd's argument will lapse—if it requires the truth of an instance of the scheme he presents. Of course, there will be *some* m, such that we will be able to find true instances for the above with mn and n replacing $2n$ and n respectively—but if m is too large, the intuitive force of Dodd's complaint will be gone.[6]

I think the penumbra of vagueness for typicality will be quite extensive—indeed, that is one interesting moral to take from Dodd's discussion. Because of this, complaints based around *instances* of the above schema will rest on untrue premises.

What remains, of course, is to work directly with the existential generalization. Here I reinforce Dodd's case, rather than rejecting it. Indeed, I invite him to strengthen his case—-what could be more surprising than the case with n and $n+1$ above! But, as one might anticipate, the appropriate response at this point is to locate the counterintuitiveness as a special instance of the counterintuitiveness of classical treatments of vagueness—with no particular lessons forthcoming for our treatment of chancy counterfactuals.

3 Dodd's argument

The above might seem over-hasty—I haven't even considered Dodd's argument yet! And indeed, Dodd presents it in a striking form (which can, if we like, make even more dramatic). Yet the source of the worry, and the appropriate complaint, are exactly the same.

Supposing we have such a $n/2n$ pair as above, Dodd asks us to consider the following might-counterfactual:

- If the fair coin had been flipped $2n$ times, and came up heads on the first n flips, it might have come up heads on the remaining flips.

If, among the closest $2n$ flip worlds there are any where the *first half* all come up heads (and why shouldn't there be? Ex hypothesi, there's nothing atypical in such worlds) then the truth of this last sentence is inconsistent with the earlier counterfactuals. But, says Dodd, *surely* this might-counterfactual shouldn't be denied!

Let us strengthen Dodd's case, by moving to the $n/n+1$ case (which, I've argued, involves essentially the same phenomena). Then the analogue of the above is, for the relevant choice of n:

- If the fair coin had been flipped $n+1$ times, and came up heads on the first n flips, it might have come up heads on the final flip.

[6] Dylan Dodd tells me that the reading he intended does require this particular instance, rather than the existential generalization.

Even if one were to balk at Dodd's original case, *this* verdict looks crazy. How on earth could a run of success counterfactually exclude one last outcome of the same kind on the final flip?

My reply reiterates the morals of the previous section. No instance of Dodd's argument works—not in the $n/n+1$ case, and not in his $n/2n$ case either (granting that there is a wide penumbra of vagueness in what counts as typical). To be sure, one can existentially generalize, and say there's *some n* such that, were the coin flipped that many times with a string of successes, the final flip would be tails. But again, these are exactly unpalatable existentials springing from the vagueness of the counterfactuals and the particular treatment of indeterminacy in play, not from independent faults in the account.[7]

The 'unpalatability' of this existential is pretty extreme, so we better try to be sure that we're laying the blame in the right place. Can we somehow test whether our disquiet is towards the treatment of indeterminacy in this case, or to the treatment of counterfactuals? One obvious strategy is to vary the treatment of indeterminacy, and see if the problematic case varies with it.

So suppose we took a truth-functional truth-value gap treatment of vagueness rather than a supervaluational one. Each instance of Dodd's scheme would be either false or at least untrue, just as in the supervaluational setting. But we would not even be committed to the truth of the existential statement—since for the truth-functional theorist, no existential can be true unless some instance is. So were we to change the treatment of indeterminacy, and we would get rid of exactly the residual, quantified shadow of Dodd's worry.[8]

4 An underlying worry

The preceding discussion participates in philosophical trench warfare. Lewis proposed a treatment of counterfactuals, and a patch for the chancy case. Hawthorne raised objections.

[7]Because I do not endorse any instance as determinately true, I evade the arguments in the latter part of Dodd's paper intended to embarrass those who *do* accept such things as true. I will say something about these arguments, however, since they raise independently interesting issues.

Dodd's main thought is that Lewis-Williams counterfactuals won't have the sort of *cognitive role* we typically associate with them. In particular, when we *fully believe* a counterfactual, and come to fully believe its antecedent (while retaining belief in the conditional), we are disposed to come to fully believe its consequent. But, says Dodd, Lewis-Williams counterfactuals wouldn't work this way. I disagree.

All parties agree that the counterfactuals are false when the antecedent is true and consequent false, in cases where we *learn that the antecedent is satisfied* (this comes from weak centering constraint on closeness orderings, and is required for modus ponens to be valid). This means that our confidence in the truth of a Lewis-Williams counterfactual may markedly decrease when we learn the antecedent is true, since our confidence in *these* false-making worlds has been raised. Indeed, if we assume strong centering, then the truth value of a counterfactual in antecedent-worlds exactly matches the truth-value in that world of its consequent. Hence, in the limit where we become certain of the antecedent, our credence in the counterfactual is fixed independently of anything else we say about closeness (so issues of typicality, quasi-miracles etc become irrelevant), which only influences the truth-value of the counterfactual in worlds where the antecedent is false. So it looks unlikely we can get any argument against their view *based on situations where we come to learn with full confidence that the antecedent holds* that wouldn't generalize to any other account meeting the same formal constraints. (In Dodd's case, for example, I would be inclined to say that upon learning the antecedent is fulfilled, one should not *fully believe* the counterfactual—it relevantly like a lottery proposition since its truth will exactly match a lottery-like process).

[8]That's not to say I recommend giving up on supraclassical treatments of indeterminacy! Choosing between revisionary logics and counterintuitive existentials is a debate for another day.

Williams proposed what he took to be a principled fix to the account. Dodd objects; and I have responded. Maybe what I say above will convince the reader; maybe Dodd will be able to make a compelling reply. But can we stand back and see what forces are shaping the debate?

It seems to me that the intuitive force of objections that Hawthorne and Dodd highlight can be rooted in dissatisfaction with one particular feature of the accounts that Lewis and Williams offer of chancy counterfactuals.[9] As is made very salient in the coin counterfactuals we have considered, there is a disconnect between the hypothesized *conditional chance* of the consequent on the antecedent, and the seeming confidence that the advocates of such accounts would have us invest in the counterfactuals. A true believer in Lewis or Williams' account could, in the limit where they are certain the antecedent does not obtain, be certain of the truth of:

> If I were to flip a fair coin a billion times, it wouldn't come up heads every time

At the same time, the chance of not-all-heads conditionally on a billion flips, while extremely high, is not 1.

Hawthorne's 'exclusion of the more probable' problem invited us to consider counterfactuals such as the following (where S is a particular, random, sequence of flips):

> If I were to flip the coin a billion times, and the result were to be either S, or all-heads or all-tails, then it would be S.

The point is that S is the only non-atypical (or non-remarkable) world consistent with the antecedent; so by the lights of Lewis or Williams, they should count this true. Likewise, there would be something truly repugnant about the $n/n+1$ case above (if we could only find an instance that made it true!):

> If I were to flip the coin a billion times, and the first billion-minus-one came up heads, then the last one would come up tails.

Notice that in Hawthorne's example, the counterfactual could be known, with certainty, to be true, even though the conditional chance of the consequent on the antecedent is only 1/3. And the Dodd example is similar—the conditional chance this time being 1/2.

What is the common theme here? It is that we are *permitted to be more certain of the counterfactual than of the corresponding conditional chance*. This is already present in

[9] I'm wary of speculating about the motivations of particular authors. Indeed, some of Dodd's paper suggests that he sees the primary problem as concerning the interaction of epistemic modals with counterfactuals. As already indicated, Williams (forthcoming) agrees on this particular issue (and Lewis (1979) agrees on a "narrow scope nomic modal" reading of 'might' counterfactuals). Compatibly with the *semantics of counterfactuals* that Lewis and Williams offer, one could dispute the Lewis duality analysis of might-counterfactuals. One could, for example, construe these as Bennett (2003) suggests, as concerning counterfactual (idealized) knowledge. So this aspect could be given up.

embryonic form in the original case—and Hawthorne and Dodd devise ways of 'blowing up' the disparity so that it stares us right in the face.

Whatever the results of the local fighting over this or that example, everyone should acknowledge the failure, on the Lewis-Williams approach, of a highly intuitive relationship constraining credences in counterfactual conditionals:

> *One's credence in a counterfactual $A \mathrel{\Box\!\!\rightarrow} B$ should be no higher than one's expectation of the conditional chance of B, on A.*

In symbols:

$$P(A \mathrel{\Box\!\!\rightarrow} B) \leq \sum_w P(w) \cdot Ch_w(B|A)$$

Call this the *Counterfactual Ramsey Bound*—since, like the constraint on the probabilities of conditionals that goes by a similar name, it places a constraint on credence in conditionals.

Not only is this equation intuitive, one can also see how useful it would be if true. By asserting a counterfactual, we indicate that our credence in it is high. If we could assume the bounds above, then we would thereby be communicating information about what we take the relevant chances to be. And our expectation of such chances are deeply integrated into our practical rationality—arguably they are the determinants of when, by our lights, it is rational to choose one option over another.[10] This looks bad for the Lewis-Williams approach—not as a matter of details, but in terms of its very conception and ambition.

However, one should recall at this point the fate of *other* intuitive, highly motivated constraints on the probability of conditionals—also constraints that would connect conditionals with (one conception of) practical rationality.[11] Equating the probability of an indicative conditional with the conditional probability of the consequent on the antecedent has been subjected to scathing criticism over the last thirty years.[12] So one should pause before imposing the chance-counterfactual link as a desiderata, and consider whether it is even a satisfiable demand.

I argue elsewhere that the Counterfactual Ramsey Bound above is just as problematic as the notorious probabilistic constraints on indicative conditionals; it turns out that Lewis's core triviality argument can be adapted to reduce this to absurdity.[13] So it is rather heroic to impose the Counterfactual Ramsey Bound—absent fancy footwork, arguments that counterfactuals have *no truth conditions at all* loom.[14]

To summarize. I contend that one *underlying motivation* for objectors to the Lewis-Williams treatment of counterfactuals and chance is unhappiness with the disconnect between credence

[10] I am thinking of chance-formulations of causal decision theory, on which—in the usual string of cases—expectation of chances are the 'weights' by which we average the utilities of individual worlds to determine the utility of a given option.

[11] This time, Jeffrey's evidential decision theory.

[12] Lewis (1976) kicked this off. For a review of the state of the art, see (Hájek & Hall, 1994).

[13] Can other impossibility arguments be replicated? Briggs (manuscript) argues that they can.

[14] Compare Skyrms (1994) and Edgington (1997).

in counterfactuals and their known conditional chance. And the failure to meet this constraint *does* seem to me a worry, and one that no doubt can be made dramatic in a variety of ways (whether or not the way that Dodd tries to bring it out in his paper is ultimately successful). But if the strong tendency in the literature on the indicative case is a good guide, imposing the constraint will make problematic even the idea that counterfactuals *even have truth-conditions*. Perhaps, faced with the alternative, we might be kinder to accounts that fail to live up to this ultra-demanding ideal.

5 Conclusion

Dodd's paper raises an interesting new challenge for extant accounts of chancy counterfactuals. It does highlight counterintuitive consequences of the overall theory—but those consequences, I contend, should be diagnosed as sourced in the (implicit) treatment of *vagueness and indeterminacy*, rather than problems with the treatment of counterfactual closeness per se. In the final section, I've drawn attention to one general argumentative move in discussing counterfactuals and chance—the intuitiveness of a certain kind of constraint imposed by expected chance on our credence in counterfactuals. But my advice is to be wary of any objection drawn directly from this debate—for as is familiar from the literature on indicative conditionals, intuitive probabilistic constraints, applied without due care and attention, may undermine the very project of giving truth-conditions for conditionals.

References

BENNETT, JONATHAN. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

BRIGGS, RACHAEL. manuscript. 'Two interpretations of the Ramsey Test'.

DODD, DYLAN. 2009. 'Quasi-miracles, typicality and closeness'. *Synthese*.

EDGINGTON, DOROTHY. 1997. 'Truth, Objectivity, Counterfactuals and Gibbard'. *Mind*, **106**, 107–116.

ELGA, ADAM. 2004. 'Infinitesimal Chances and Laws of Nature'. *Australasian Journal of Philosophy*, **82**, 67–76.

FINE, KIT. 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.

GAIFMAN, HAIF, & SNIR, MARC. 1982. 'Probabilities over rich languages, testing and randomness'. *The Journal of Symbolic Logic*, **47**(3), 495–548.

HÁJEK, ALAN. MS.. *'Most counterfactuals are false'*.

HÁJEK, ALAN, & HALL, NED. 1994. 'The hypothesis of the conditional construel of conditional probability'. *In:* EELS, E., & SKYRMS, B. (eds), *Probability and Conditionals*. Cambridge: CUP.

HAWTHORNE, JOHN. 2005. 'Chance and counterfactuals'. *Philosophical and Phenomenological Research*, 396–405.

LEWIS, DAVID K. 1973. *Counterfactuals*. Oxford: Blackwell.

LEWIS, DAVID K. 1976. 'Probabilities of Conditionals and Conditional Probabilities'. *Philosophical Review*, **85**, 297–315. Reprinted with postscripts in Lewis, *Philosophical Papers vol. II* (Oxford University Press, 1986) 133-52. Also reprinted in Jackson (ed) *Conditionals* (Oxford University Press, 1991) 76-101.

LEWIS, DAVID 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.

SKYRMS, BRIAN. 1994. 'Adams conditionals'. *Pages 13–26 of:* EELLS, E., & SKYRMS, B. (eds), *Probability and conditionals*. Cambridge: Cambridge University Press.

WILLIAMS, J. ROBERT G. 2008. 'Chances, counterfactuals and similarity'. *Philosophy and Phenomenological Research*, **77**(2), 385–420.

WILLIAMS, J. ROBERT G. forthcoming. 'Defending conditional excluded middle'. *Nous*. Penultimate draft available at: `http://www.personal.leeds.ac.uk/~phljrgw/wip/mightCEM.pdf`.