

CONDITIONALS, CAUSATION, AND DECISION

DOROTHY EDGINGTON

University of London

Introduction

This paper is about the interface between the theory of conditionals and the theory of rational decision. Its focus is on conditionals of the form “If I do x , such-and-such will happen,” when these might provide reasons for or against doing x . Needless to say, conditional judgments of this form are typically uncertain judgments. Decision theory concerns how these typically uncertain judgments, and the values, positive or negative, that we attach to the various possible consequences of an action interact to determine which action we should perform.

I am in favor of using conditional probability as a measure of one’s degree of uncertainty in conditional judgments. Richard Jeffrey has a classic development of decision theory which employs conditional probabilities. But there are apparent counterexamples to Jeffrey’s decision theory, and as a consequence, rival theories use instead the probability of a counterfactual conditional of the Stalnaker–Lewis kind. The two sides to the resultant debate go under the names “Evidential Decision Theory” (Jeffrey’s) and “Causal Decision Theory.” I here defend a sort of cross-bench position: I am in favor of causal decision theory but against its elaboration in terms of the probability of counterfactual conditionals. Like Jeffrey, I favor the use of conditional probabilities in decision theory.

Richard Jeffrey’s Decision Theory

We get a handle on the motivation for Jeffrey’s theory by considering it as a reaction to a principle held by Leonard Savage, which Savage called the “Sure Thing Principle.” The principle also goes under the name “dominance.”

Here is Savage’s example:

A businessman contemplates buying a certain property. He considers the outcome of the next presidential election relevant to the attractiveness of the purchase. So, to clarify the matter to himself, he asks whether he would buy if he knew that the Republican candidate were going to win, and decides that he would do so. Similarly, he considers whether he would buy if he knew the Democratic candidate were going to win, and again finds that he would do so. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event . . . will obtain. . . . It is all too seldom that a decision can be arrived at on the basis of the principle used

by this businessman, but . . . I know of no other extralogical principle that finds such ready acceptance.¹

That is, roughly, if you would prefer A to C assuming that B is true, and you would prefer A to C assuming that B is false, you prefer A to C .

This principle might be all right under special conditions which are satisfied in this example, but it is not generally valid. Here is one of Jeffrey's counter-examples, which he calls "the Marlboro man." The Marlboro man says:

Sure, smoking makes it more likely for me to die before my time, but smokers can live to a ripe old age. Now if I'm to live to a ripe old age I'd rather do it as a smoker, and if not, I would still enjoy the consolation of smoking. *In either case*, I'd rather smoke than quit. So I'll smoke.²

The difference between the two examples is that in the first, it is assumed that whether the businessman buys the property has no effect on the probability of the result of the presidential election; but in the second, whether he smokes does affect the probability of living to a ripe old age. Here is how Jeffrey would analyze the second example. (I made up the figures. Obviously, we have to make precise "ripe old age"—say over 70).

$$p(\text{Long life given Smoke}) = 0.4; p(\text{Not long life given Smoke}) = 0.6$$

$$p(\text{Long life given Quit}) = 0.8; p(\text{Not long life given Quit}) = 0.2$$

$$V(\text{Long life and Smoke}) = 11; V(\text{Not long life and Smoke}) = 1$$

$$V(\text{Long life and Quit}) = 10; V(\text{Not long life and quit}) = 0.$$

The expected value of smoking is the weighted average of the values of smoking and long life, and smoking and not long life, the weights being the conditional probabilities of the different outcomes, given smoking, that is,

$$(0.4 \times 11) + (0.6 \times 1) = 5.$$

The expected value of quitting:

$$(0.8 \times 10) + (0.2 \times 0) = 8.$$

With these numbers, the expected value of quitting exceeds the expected value of smoking. Despite the fact that smoking is better than quitting in both scenarios, the fact that the highly valued Long Life has a much higher probability in the second, quitting, scenario more than compensates for that. If the probabilities of the relevant uncertain outcomes are the same in both rows, the Sure Thing Principle holds. If not, there is no guarantee that it holds.

Another of Jeffrey's examples is nuclear deterrence. The argument for unilateral disarmament is as follows: either there will be war or peace. If there's

1. L. Savage, *The Foundations of Statistics*, 2nd ed. (Dover, 1972), p. 21.

2. R. Jeffrey, *Subjective Probability: The Real Thing* (CUP, 2004), p. 100.

war and we have no arms, we'll be living under an abhorrent regime but that's better than the (near) extinction of human life. If there's peace, and we have no arms, things are better than peace with arms. Either way, it's better to disarm. The believer in deterrence replies that there is much more likely to be war if we disarm than if we don't, and the expected value of keeping the weapons is higher than the expected value of disarming. The details are like that of the Marlboro man above.

So in general, for Jeffrey, where $O_1, \dots, O_n$ are the alternative possible outcomes of some possible action, A ,

$$V(A) = p(O_1|A)V(O_1 \& A) + \dots + p(O_n|A)V(O_n \& A).$$

Note that the theory uses the conditional probabilities of outcome given action, in calculating the expected value of an action.

Counterexamples to Jeffrey's Decision Theory

Newcomb's far-fetched example (see Nozick³) will be discussed, in a revised version, later, but in honor of its originality, some less far-fetched examples have come to be known as "medical Newcomb cases." Here is one. The statistician Sir Ronald Fisher put forward the hypothesis that the correlation between smoking and cancer was not to be explained by the fact that smoking causes cancer, but that a certain genetic condition causes both cancer and the sort of nervous disposition which inclines one toward smoking. That is, smoking and cancer are effects of a common cause, like barometer readings and rain, or spots and fever, but one doesn't cause the other. Suppose that were true. The conditional probability of getting cancer, given that you smoke, may still be considerably higher than the conditional probability of getting cancer, given that you don't smoke (because smoking is a sign that you have the bad gene). But this is no longer a reason not to smoke. You either have the gene or you don't, and refraining from smoking isn't going to reduce your chances of getting cancer. Fisher's hypothesis was, alas, false, but far from unthinkable, and its structure is clear. Another more plausible example: Eating chocolate is correlated, in susceptible people, with getting migraine headaches, but it might well not be a cause. Rather, a craving for chocolate might be a side effect of pre-migraine tension—again an effect of a common cause. So the probability of getting a migraine, given a craving for chocolate, is high—but on the "common cause" hypothesis this is *not* a reason to resist the chocolate: Nothing you can do now can change your chances of getting migraine. In no way does it depend causally on your eating chocolate.

Jeffrey's decision theory gives the wrong answers. Jeffrey's theory works fine when the conditional probabilities are grounded in beliefs about appropriate causal connections, but not when they are grounded in beliefs about joint effects of a common cause.

3. R. Nozick, "Newcomb's Problem and Two Principles of Choice," in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Reidel, 1970), 114–5.

Counterfactual Decision Theory

The most common response to the counterexamples has been to switch to what is called “causal” decision theory.⁴ Its original and most popular version is due to Stalnaker⁵ and Gibbard and Harper,⁶ and it is broadly endorsed by David Lewis.⁷ The change that needs to be made to Jeffrey’s decision theory is very simple, according to this view. Instead of the weights being conditional probabilities, the weights are probabilities of counterfactual or subjunctive conditional propositions. (Note: I shall stick to the label “counterfactual,” as most participants in the debate do, because the issue is not really one of grammar but one of function. But, not to be misled, you have to realize that these are “counterfactuals” which do *not* presuppose the falsity of the antecedent. It is just a convenient label for a type of conditional, a conditional which in English has a “would” in the consequent, which includes those that do presuppose the falsity of the antecedent.)⁸ I think Stalnaker stated this first:

$$V(A) = p(A > O_1)V(A \& O_1) + \dots + p(A > O_n)V(A \& O_n),$$

where $O_1 \dots O_n$ are the alternative possible outcomes of doing A , and “ $A > O$ ” stands for the counterfactual conditional, “If it were the case that A , it would be the case that O .”

Note that the only difference between this theory and Jeffrey’s is that we have replaced conditional probabilities with probabilities of counterfactual conditionals.

Take the Fisher case. The idea is that while the conditional probability of cancer given smoking is greater than the conditional probability of cancer given nonsmoking (because smoking is a sign of the presence of the genetic condition), there is also this thought: “Were I to smoke, I would get cancer” is no more likely than “were I not to smoke, I would get cancer,” for the genetic condition is present or it is not, and smoking isn’t going to make a difference. So in *this* case, the Marlboro man’s thinking is right. We get the same probabilities of cancer for smoking and nonsmoking, and you might as well smoke if you want to. Stalnaker writes about this case: “It seems clear that the probability of the conditional (If I were to smoke, I would get cancer), and not the conditional probability is what is relevant.”⁹

4. The scare-quotes round “causal” are there because, for many at least, causation does not figure directly in the theory—rather, the theory runs in terms of counterfactual conditionals.

5. R. Stalnaker, *Letter to David Lewis*, reprinted in W. Harper, R. Stalnaker, and G. Pearce (eds.), *Ifs* (Reidel, 1972), pp. 151–2.

6. A. Gibbard and W. Harper, *Counterfactuals and Two Kinds of Expected Utility*, reprinted in W. Harper, R. Stalnaker, and G. Pearce (eds.), *Ifs* (Reidel, 1978), pp. 153–90.

7. D. Lewis, “Causal Decision Theory,” *Australasian Journal of Philosophy*, 59 (1981), pp. 5–30. Reprinted in his *Philosophical Papers* vol. II. Oxford University Press, 1986, pp. 305–339. Page references to reprint.

8. Note also: It is assumed by those defending this view that counterfactuals have truth conditions along the lines proposed by Stalnaker and Lewis.

9. Stalnaker (1972): 152. Some philosophers, such as A. Gibbard (*Two Recent Theories of Conditionals*, in W. Harper, R. Stalnaker, and G. Pearce [eds.], *Ifs* [Reidel, 1981], pp. 211–47) and J. Bennett (*A Philosophical Guide to Conditionals* [OUP, 2003]), take indicative conditionals to be “epistemic”

David Lewis¹⁰ first states his theory in terms of “hypotheses of causal dependence”; but he endorses the Stalnaker–Gibbard–Harper account in terms of counterfactual conditionals: “If we want to express a dependency hypothesis in ordinary language, it is hard to avoid the use of counterfactual conditionals saying what would happen if the agent were to realise his various alternative options” (p. 325). He agrees that it is the probabilities of counterfactual conditionals that matter for decision theory—but with a qualification; for, as Lewis realizes, counterfactuals are highly context sensitive, and, in context, it is permissible to use counterfactuals in the backtracking way, supported via a common cause. “If Jack were to ask Alice to marry him, she would say yes—because he wouldn’t have asked unless he were really sure that she would accept.” Or: Again suppose that Fisher’s hypothesis is true. The doctor examines a patient who is complaining of a pain in the chest. “Do you smoke?” asks the doctor. “No,” says the patient. Later, a student who was observing asks the doctor why that question was asked. “Because, were it the case that she smoked, that would be a sign that she had the genetic condition that causes cancer,” answers the doctor. “If she had smoked, it would have been more likely to be cancer.”

Lewis says

It is clear that not just any counterfactual could be part of a pattern expressing causal dependence. . . . We must exclude “back-tracking counterfactuals” based on reasoning from different supposed effects, back to different causes and forward again to differences in other effects. Suppose I am convinced that stroking [the cat] has no influence over purring, but that I wouldn’t stroke Bruce unless I were in a mood that gets him to purr softly by emotional telepathy. Then I give credence to

If I were to stroke Bruce, he would purr softly

taken in the back-tracking sense, but not taken in the sense it must have to be part of a [reason to stroke him].

Lewis continues:

Some will doubt that causal counterfactuals [the ones we need for decision theory] can be distinguished from others except in causal terms. I disagree, and think it possible to delimit the causal counterfactuals in other terms and thus provide noncircular analyses of causal dependence and causation itself. But that is a topic for other papers.¹¹

So, according to Lewis, and he presumes that the other counterfactual-causal decision theorists would agree, it is only a subset of counterfactuals that

and assessed by conditional probability, and counterfactual conditionals to be “causal,” and to be assessed in the Lewis/Stalnaker way. So the debate, for them, is whether it is indicative or counterfactual conditionals which we need in deliberating about what to do. I shall return to this question.

10. Lewis (1981).

11. *Ibid.*: 326–7.

are fit to be used in decision theory. But now the case for using counterfactuals at all is somewhat weakened. **If we find that we need to restrict our theory to a particular subset of counterfactuals, then why couldn't we have restricted Jeffrey's theory to that subset of conditional probabilities which track causal relations, excluding the unwanted ones?** If it suffices to exclude backtracking counterfactuals, then restrict the conditional probabilities that figure in a theory of decision to those that are not grounded in backtracking considerations. (We will see later that backtracking is not the root of the problem, but never mind that now.) If we need to be more up-front in our account of specifically causal counterfactuals, we can equally be up-front about those conditional probabilities that figure in a theory of decision: Those that could be more explicitly expressed as the probability, on the assumption that I do *A*, that my doing so will bring it about (or help bring it about) that *C* as a result. Mimicking Lewis, we can say: "It is clear that not just any judgement of conditional probability could be part of a pattern expressing causal dependence. . . . We must exclude 'back-trackers' based on reasoning from effects back to causes and forward again to different effects. . . ."

If this point is correct, we have, at this stage, a draw, a standoff. I argue later that we do need conditional probabilities, suitably restricted, not the probability of counterfactual conditional propositions, suitably restricted.

Counterexamples to Counterfactual Decision Theory

Andy Egan¹² has a powerful paper entitled "Some Counterexamples to Causal Decision Theory" (2007).¹³ I admire this paper. But I don't like its title. It seems to me to have excellent counterexamples to the claim that it is the probability of counterfactuals that feed into our decisions, and no counterexamples at all to the claim that causal considerations are crucial in decision theory. On the contrary, it provides further evidence for that claim.

Here are Egan's flagship examples. First:

Mary is debating whether to shoot her rival, Alfred. If she shoots and hits, things will be very good for her. If she shoots and misses, things will be very bad. (Alfred always finds out about unsuccessful assassination attempts, and he is sensitive about such things.) If she doesn't shoot, things will go on in the usual okay-but-not-great kind of way. Though Mary is fairly confident that she will not actually shoot, she has, just to keep her options open, been preparing for this moment by honing her skills at the shooting range. Her rifle is accurate and well maintained. In view of this, she thinks it is very likely that, if she were to shoot, then she would hit. So far, so good. But Mary also

12. A. Egan, "Some Counterexamples to Causal Decision Theory," *Philosophical Review*, 116 (2007), pp. 93–114.

13. Egan was convinced, in light of the original counterexamples to Jeffrey, of the need for causal decision theory. He puts forward his counterexamples to the latter regretfully, and says he does not know what to offer in its place.

knows that there is a certain kind of brain lesion that tends to cause both murder attempts and bad aim at the critical moment. . . . Happily for most us (but not so happily for Mary) most shooters have this lesion, so most shooters miss. Should Mary shoot?¹⁴

No. Here she *should* backtrack, and think: “suppose I do actually shoot. Then it’s quite likely that I have the lesion, which will make my hand wobble, and make me miss.”

Second example¹⁵: Paul is debating whether to press the “kill all psychopaths” button. It would, he thinks, be much better to live in a world with no psychopaths. He’s discussing this with a friend, who says “Only a psychopath would press such a button.” Paul comes to think that might well be true. Paul strongly prefers living in a world with psychopaths to dying. Should Paul press the button?

No. He thinks to himself: suppose I do press the button; then that is evidence that I am a psychopath, in which case I will thereby kill myself.

Recall that backtracking counterfactuals had to be explicitly excluded, in order to avoid the counterexamples to Jeffrey. We are not allowed to use counterfactuals like “If Mary were to shoot, that would mean that she probably had the lesion, and so she would probably miss.”

Egan has made a slight change to the causal structure of the common cause case. We do argue back from the action *A*, to its likely cause *C*, which will also *cause*, together with *A*, *E*. There is no failure of causal connection any more between *A* and *E*. In Figure 1 below, you have the structure of the ordinary common-cause case on the left, and the Egan structure on the right.

Egan shows¹⁶ that he has a recipe for converting any counterexample to Jeffrey into a counterexample to counterfactual decision theory. Start with a counterexample to Jeffrey in which some condition is (believed to be) a common cause both of some action *A* and of some undesirable outcome *O*.

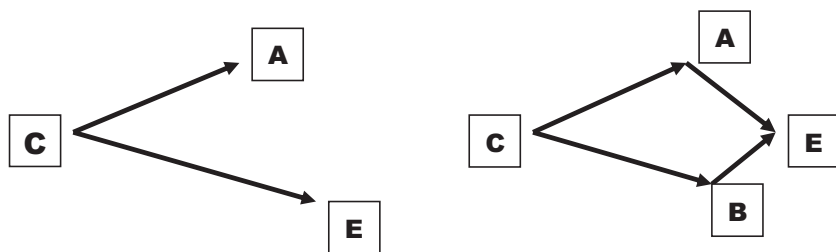


Figure 1: Common causes, simple and Egan-style.

14. Egan (2007): 97.

15. Ibid.

16. Ibid.: 103.

Change the case so that, rather than directly causing *O*, the condition puts in place an enabling condition which allows *A* to cause *O*. Point out that our intuitions switch about what one ought to do, while the recommendations of the counterfactual decision theory remain the same.

Finally, he has a nice twist on the Newcomb problem itself. The original problem: Box B contains \$1,000. Box A contains either \$1,000,000 or nothing. You can take just box A, or both boxes. An uncannily accurate predictor has put \$1,000,000 in box A if he predicted that you will take just box A, nothing in A if he predicted that you take both boxes. There is no backward causation: It is now fixed whether or not he has put the money in box A. There's a strong argument for two-boxing. You can now have no effect on the distribution of the money so you have nothing to lose by taking both boxes.

Now for Egan's twist. There are two boxes before you. Box A definitely contains \$1,000,000. Box B definitely contains \$1,000. You have two choices: take only box A (call this one-boxing), or take both boxes (call this two-boxing). You will signal your choice by pressing one of two appropriately labeled buttons. There is, as usual, an uncannily reliable predictor on the scene. If the predictor has predicted that you will two-box, he has planted an incendiary bomb in box A, wired to the *two-box* button, so that pressing the two-box button will cause the bomb to detonate, burning up the million dollars. If the predictor has predicted that you will one-box, no bomb has been planted—nothing untoward will happen, whichever button you press. The predictor, again, is uncannily accurate.¹⁷

Now in this version, it is rational to one-box: If you press the two-box button, there is a large risk that you will *cause* the incineration of your \$1,000,000, which is certainly sitting there just waiting for you. The fact that your choice will *cause* the incineration is the difference between that and the original Newcomb problem.

Go back to Mary's plans to murder Alfred. *S* is the proposition that she shoots. *X* is the proposition that if she were to shoot, she would hit. Most of the $\neg S$ -worlds are worlds in which, if she were to shoot, she would hit. Few of the *S*-worlds are worlds in which, if she were to shoot, she would hit. Overall (and especially because she thinks she won't actually shoot, there's a high probability that if she were to shoot, she would hit. See Figure 2. (This is how Egan defends the claim that the probability of the counterfactual conditional is high. See note 18, pp. 98–9).

Suppose Mary is presented with the above argument that there is a high probability for the proposition that if she were to shoot, she would hit. Her correct reply would be: "Very well, it's probable that if I were to shoot I would hit, but this gets most of its probability from the worlds in which I don't shoot! I'm interested in what happens *on the assumption that I do actually shoot!* And in the worlds in which I shoot, it's likely that I have the lesion, so unlikely that I'll hit if I shoot."

17. Ibid.: 109.

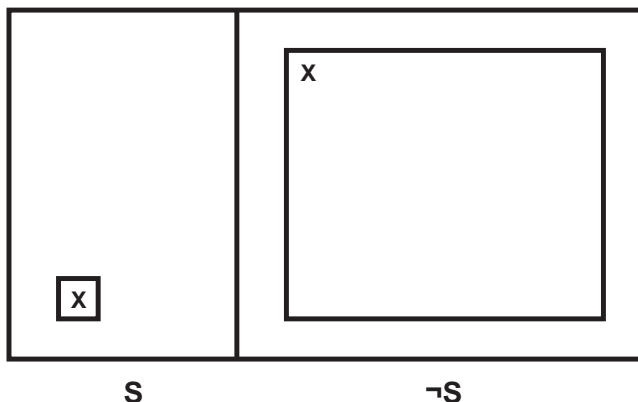


Figure 2: Lack of probabilistic independence.

Egan's examples exploit a very general problem for the claim that what matters in decision taking is the probability of a conditional proposition. **No two contingent, logically independent propositions are probabilistically independent in all probability distributions**—or in all belief distributions. There are always possible states of information in which one bears positively or negatively on the other. Maybe conditional propositions are normally independent of their antecedents, in which case it doesn't matter whether you consider $p(A \rightarrow C)$ or $p((A \rightarrow C) | A)$. But where they do come apart, as in Egan's examples, it is the conditional probability that matters: the probability of the conditional *on the assumption that you do the action*. Also, on (almost) any understanding of the conditional proposition, $p((A \rightarrow C) | A) = p(C | A)$.¹⁸ So: (provided we are focusing on the restricted category based on causal relations) it is sometimes essential to use conditional probabilities, and it never does any harm to use the conditional probabilities.

The Lessons

Egan's examples show that doing decision theory in terms of non-backtracking counterfactuals is no good. Indeed, the ban on backtracking is a bad idea. We want all the evidence we can get about what the causal situation will be, *on the assumption that I do A*. In the counterexamples to Jeffrey, what the backtracking evidence reveals is that there is no causal connection between smoking and

18. The equality holds provided $A \& B$ entails $A \rightarrow B$. This is so on standard accounts of counterfactuals. Good reasons for the claim are found in L. Walters, "Morgenbesser's Coin and Counterfactuals with True Components," *Proceedings of the Aristotelian Society*, 109 (2009), pp. 365–79.

getting cancer, and that is why the probability of getting cancer, given that you smoke, is irrelevant to the decision whether to smoke. In the counterexamples to Counterfactual Decision Theory, the backtracking evidence reveals that there is a causal connection. All the examples reinforce the idea that causation is central, ignored at our peril, in a theory of decision. It is what is likely to be *brought about* by my action that matters.

A properly causal decision theory should be up-front about causation (like causal theories of other things). We need conditional judgments like “If I do x , that will bring it about that C (or cause, or produce as a result, or have as an outcome, that C). More cautiously: “If I do x , that will contribute to the bringing about that C .”

We have also seen a reason to go for conditional probabilities rather than probabilities of some sort of conditional proposition. The latter might get most of its probability from the case in which the antecedent is false, and that gives us no reason to act on it.

We need to assess the probability, on the assumption that I do x , that C will be a causal consequence.

Let’s try to see how this idea works out in the three kinds of case we have considered. Return to the smoking example used to illustrate Jeffrey’s case against the Marlboro man (see pp. 2–3). Again, let the values be:

$$V(\text{Long Life and Smoke}) = 11; V(\text{Short life and Smoke}) = 1.$$

$$V(\text{Long life and Quit}) = 10; V(\text{Short life and Quit}) = 0.$$

Long life = making it to your 70th birthday; short life = not making it to your 70th birthday. Cases (1), (2), and (3) are: the ordinary causal case, the Fisher case, and the Egan case.

I suggest that the weights should be something like:

	(1)	(2)	(3)
(a) $p(\text{Short life as a result of smoking} \text{Smoke})$	0.6	0	0.6
(b) $p(\text{Not (short life as a result of smoking)} \text{Smoke})$	0.4	1	0.4
(c) $p(\text{Long life as a result of quitting} \text{Quit})$	0.8	0	0.8
(d) $p(\text{Not (long life as a result of quitting)} \text{Quit})$	0.2	1	0.2

First, in the ordinary case in which smoking tends to shorten life, let the probabilities be as they were above: (a) 0.6 and (b) 0.4; (c) 0.8 and (d) 0.2, and we get the same answer as before: It is rational to quit.

In the Fisher case—the pure common cause case (a) is close to 0 (c) is close to 1; (c) is close to 0 (d) is close to 1; for simplicity, pretend that they are 0 and 1. (a) = (c) (b) = (d). Then as $V(\text{Long life and Smoke})$ exceeds $V(\text{Long life and quit})$, it is rational to smoke.

In an Egan-style smoking case,¹⁹ the agent believes that a desire to smoke is produced by a genetic condition which also causes one’s lungs to be vulnerable to cigarette smoke, such that smoking causes cancer in those with the genetic

19. The example occurs in Egan (2007): 103.

condition, but not in those without. In this case, the probabilities of (a)–(d) revert to those in the first case, and it is rational not to smoke.²⁰

So far so good. (It would be rash to be very confident that there are not problematic cases in the vicinity yet to be considered!)

Are the Conditionals of Practical Deliberations Indicative Conditionals or Counterfactual Conditionals?

What of the plain conditionals that we accept and assert as bases for actions? “If you offer him \$500, he’ll accept”, etc.? (Typically, of course, these are conditional judgments about which we are less than certain.) Some philosophers, for instance, Gibbard²¹ and Bennett,²² hold that indicative conditionals are assessed according to the conditional probability of consequent given antecedent, whereas counterfactual conditionals are to be treated in the Stalnaker–Lewis fashion. With that perspective on the two kinds of conditional, the argument of this paper supports the view that it is indicative conditionals, not counterfactual conditionals, which matter for decision theory.

I also hold that indicative conditionals are assessed as conditional probabilities (see, e.g., Edgington).²³ But I hold that counterfactual conditionals are also assessed as conditional probabilities. Nevertheless, I think that the conditionals of practical deliberation are best thought of as indicative conditionals.

Keith DeRose,²⁴ in a recent paper in *Mind* (2010), “The Conditionals of Deliberation,” also argues that the conditionals of practical deliberation are indicative conditionals. He takes this to be a heretical view. He doesn’t go into decision theory, and doesn’t even get seriously into the semantics of conditionals, but takes paradigm cases of each sort, notes some of their features, and argues that the conditionals of deliberation share the features of indicatives, not those of counterfactuals. He ends the paper by noting that if he is right, indicative conditionals are more important than they have often been thought to be. He also notes that some indicative conditionals are “deliberationally useless,” when the causation does not go the right way. For instance: Your

20. An alternative, more explicit, and more symmetric way of setting out the possibilities in this example:

	(1)	(2)	(3)
(a) $p(\text{Short life as a result of smoking} \text{Smoke})$	~0.6	~0	~0.6
(b) $p(\text{Long life as a result of smoking} \text{Smoke})$	~0	~0	~0
(c) $p(\text{Smoking does not affect length of life} \text{Smoke})$	~0.4	~1	~0.4
(a') $p(\text{Short life as a result of quitting} \text{Quit})$	~0	~0	~0
(b') $p(\text{Long life as a result of quitting} \text{Quit})$	~0.8	~0	~0.8
(c') $p(\text{Quitting does not affect length of life} \text{Quit})$	~0.2	~1	~0.2

21. Gibbard (1981).

22. Bennett (2003).

23. D. Edgington, “On Conditionals,” *Mind*, 104 (1995), pp. 235–329.

24. K. DeRose, “The Conditionals of Deliberation,” *Mind*, 119 (2010), pp. 1–42.

friend Jack is thinking about asking his girlfriend, Alice, to marry him. But Jack is both careful and astute: He is not going to risk asking unless he is very sure that she will say yes. You know all this. So you are pretty sure that if he asks, she will accept (because you know that he won't ask unless he has very strong evidence of this). You can go around telling people that if Jack asks Alice to marry him, she'll accept. You can take bets on it, etc. But the one thing you can't do is to tell Jack as a piece of practical advice: "Jack, I know that if you ask her, she'll accept." Your grounds for the conditional are not such that you can advise his acting on the conditional.

Similarly, in one of DeRose's examples, a card game, Pete knows that if he plays he'll win, because he knows that his henchman has slipped him a piece of paper telling him his opponent's cards, although he has not yet looked at it. He tells his worried wife "If I play I'll win." But this knowledge doesn't give him a reason to play.

DeRose claims that what he calls the "were"-d-up version—not "If I do x . . ." but "If I were to do x . . ."—excludes the deliberationally useless, forces the deliberationally useful, causal reading. I think this is overoptimistic. Recall Lewis's example: If I were to stroke the cat, he would purr, because I wouldn't stroke him unless I were in a mood to make him purr through emotional telepathy.

I have argued against the claim that it is counterfactual propositions construed in the Lewis–Stalnaker way that figure in decision making. But I do not accept that view of counterfactuals anyway. I think that counterfactuals, as well as indicatives, are assessed by conditional probability.²⁵ But they are, typically, not the conditional probabilities grounded in your present belief state: They involve some distancing, some shift from your present belief state. For instance, you take yourself to know that Oswald killed Kennedy, and consider how likely it was that Kennedy would have survived if Oswald hadn't killed Kennedy. Similarly for those "counterfactuals" which do not presuppose the falsity of the antecedent: The doctor in the A&E ward says "If the patient had taken arsenic he would have just these symptoms"—those that he actually has. In assessing this conditional, she distances herself from the fact that she knows the patient does have these symptoms. It is sometimes said that counterfactual thinking is "off-line" thinking. Now it's pretty clear that when deliberating about what to do, you need "on-line" thinking—thinking grounded in your present belief state. You add the assumption that you do *A*, and consider what the causal consequences of so doing are likely to be.

Finally: Although "causation" is in my title, you haven't learned anything about causation in this talk. I don't apologize for that. You don't learn anything about causation in Grice's "The Causal Theory of Perception,"²⁶ Davidson's "Actions, Reasons and Causes,"²⁷ Goldman's "A Causal Theory of

25. D. Edgington, "Counterfactuals," *Proceedings of the Aristotelian Society*, 108 (2008), pp. 1–22.

26. H.P. Grice, "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, Supplementary Volume, 35 (1961), pp. 121–68.

27. D. Davidson, "Actions, Reasons and Causes," *Journal of Philosophy*, 60 (1963).

Knowing,”²⁸ or Kripke’s causal theory of reference.²⁹ This is part of a tradition of using causation in the explanation of other phenomena, which has become quite respectable in philosophy. Those opposed to causal decision theory either belong to a now rather unfashionable camp of thinking that causation is so problematic it is not to be used in explaining anything else (Jeffrey might have belonged to that camp; so does Arif Ahmed³⁰; or they have a theory of causation which is based on a theory of decision and action (Huw Price³¹ comes to mind). I take causation to be a very basic concept, not reducible to anything else (see Edgington³²). So it is quite proper to appeal to it in a theory of decision.

Lewis³³ is open to the possibility that decision theory needs to be up-front about causation. But Lewis has only two kinds of conditional available to him: the truth-functional conditional and his counterfactual conditional. The probability of the truth-functional conditional is a nonstarter for feeding into a theory of decision. I have argued that the probability of a Lewisian counterfactual doesn’t work either. What we need instead are conditional probabilities, of a causal sort.

28. A. Goldman, “A Causal Theory of Knowing,” *Journal of Philosophy*, 64 (1967), pp. 355–72.

29. S. Kripke, *Naming and Necessity* (Basil Blackwell, 1980).

30. A. Ahmed, “Causation and Decision,” *Proceedings of the Aristotelian Society*, 110 (2010), pp. 111–31.

31. H. Price, “Agency and Probabilistic Causality,” *British Journal for the Philosophy of Science*, 42 (1991), pp. 157–76.

32. D. Edgington, “Causation First: Why Causation is prior to Counterfactuals,” in C. Hoerl, T. McCormack, and S. Beck (eds.), *Understanding Counterfactuals / Understanding Causation* (Oxford University Press, 2011).

33. Lewis (1981).

R. Jeffrey, *The Logic of Decision*, 2nd ed. (McGraw-Hill, 1985).