Counterfactual Decision Theory

Brian Hedden

Abstract

I defend counterfactual decision theory, which says that you should evaluate an action in terms of which outcomes would likely obtain, were you to perform it. Counterfactual decision theory has traditionally been subsumed under causal decision theory as a particular formulation of the latter. This is a mistake. Counterfactual decision theory is importantly different from, and superior to, causal decision theory, properly so-called. Causation and counterfactuals come apart in three kinds of cases. In cases of overdetermination, an action can cause a good outcome without the latter counterfactually depending on the former. In cases of constitution, an action can constitute a good outcome rather than causing it. And in cases of determinism, either the laws or the past counterfactually depend on your action, even though your action cannot cause the laws or the past to be different. In each of these cases, it is counterfactual decision theory which gives the right verdict, and for the right reasons.

1 Correlation, Causation, and Counterfactuals

Correlation is not causation, as the saying goes. One thing can be correlated with another—or, more generally, can provide evidence for it—without the one causing the other. When correlation—or evidential support—comes apart from causation, which one (if either) matters when making decisions? Evidential decision theory says that you should evaluate your possible actions in terms of their evidential relevance. In a slogan: perform the action which (in expectation) provides the best evidence that good outcomes will obtain. Causal decision theory says that you should evaluate your possible actions in terms of their causal efficacy. In a slogan: perform the action which (in expectation) is most likely to cause good outcomes.

But evidential support and causal efficacy do not exhaust the range of possible relations between actions and outcomes that might matter for rational decision-making. Counterfactuals provide another. And while counterfactuals often go with both correlation and causation, they sometimes come apart from both. This opens the possibility of a third type of decision theory. Counterfactual decision theory says that you should evaluate your possible actions in terms of which counterfactuals, or subjunctive conditionals, are true. (The term 'counterfactual' is a slight misnomer, for they do not require the falsity of the antecedent.) In a slogan: perform the action which is such that, were you to perform it, good outcomes would (in expectation) most likely obtain. In a formula: choose the action with highest counterfactual expected utility, defined as $U_{\square\!\!\rightarrow}(A) = \sum_i P(A \mathrel{\square\!\!\rightarrow} O_i) \times u(O_i)$.[1]

Counterfactual decision theory was in fact developed before causal decision theory, properly so-called. It got its start with Stalnaker (1981b) and Gibbard and Harper (1981).[2] But historically, it has been subsumed under causal decision theory. In a unifying spirit,

[1] This definition of counterfactual expected utility relies on the truth of counterfactual excluded middle, which says that for all X and Y, $(X \mathrel{\square\!\!\rightarrow} Y) \vee (X \mathrel{\square\!\!\rightarrow} \neg Y)$. Assuming classical logic, this entails that either $X \mathrel{\square\!\!\rightarrow} Y$ is true or $X \mathrel{\square\!\!\rightarrow} \neg Y$ is true. I briefly discuss the possible falsity of counterfactual excluded middle, and its implications for the proper definition of counterfactual expected utility, in §6. But none of my key arguments in §3-5 turn on the truth of falsity of counterfactual excluded middle or the exact definition of counterfactual expected utility.

[2] Stalnaker's letter to Lewis was dated May 21, 1972, and the paper from Gibbard and Harper was first drafted in 1976 and published in 1978.

Lewis (1986f, 306) describes the counterfactual decision theories of Stalnaker, Gibbard, and Harper, a related theory of Sobel (1994), and the properly causal decision theories of himself and Skyrms (1980), as sharing a common core: 'we causal decision theorists share one common idea, and differ mainly on matters of emphasis and formulation.'[3]

I will argue that counterfactual decision theory is importantly different from, and superior to, causal decision theory. While causation and counterfactuals go together in Newcomb's Problem (§2), they come apart in cases of overdetermination, constitution, and determinism (§3-5). And in each of these three cases, it is counterfactual decision theory which gives the right verdict, and for the right reasons.

2 Newcomb's Problem

Start with Newcomb's Problem (Nozick 1969), a case where causal and counterfactual decision theory give the same verdict but disagree with evidential decision theory.

You enter a room with two boxes, one opaque and the other transparent. The transparent one contains $1,000. The opaque one contains either $0 or $1,000,000. You must choose whether to take only the opaque box (one-box) or both the opaque one and the transparent one (two-box). Whether there is $0 or $1,000,000 in the opaque box was determined by a highly accurate predictor. If the predictor predicted that you one-box, then it put $1,000,000 in the opaque box, and if it predicted that you two-box, then it put $0 in the opaque box. The prediction was made yesterday. What should you do?

Evidential decision theory recommends one-boxing.[4] One-boxing provides excellent evidence for a fantastic outcome, namely your getting $1,000,000. After all, it's very probable that if you one-box, then the opaque box contains $1,000,000. Two-boxing provides excellent evidence for a middling outcome, namely your getting $1,000. After all, it's very

[3] In a similar vein, Harper and Skyrms (1988, x) suggest that 'the various forms of causal decision theory are equivalent' and Joyce (1999, 171-2) concurs that 'there is no great difference between any of the approaches,' and so 'the causal decision theorist can adopt an attitude of benign indifference' about which one is correct.

[4] Here I ignore such complications as the 'tickle defense' (Eells 1981) and ratificationism (Jeffrey 1983).

probable that if you two-box, then the opaque box contains $0.

Causal decision theory recommends two-boxing. Since the prediction was made yesterday, you cannot cause the contents of the opaque box to be one way or another. If the opaque box is empty, then one-boxing will cause you to get $0 while two-boxing will cause you to get $1,000. And if the opaque box is full, then one-boxing will cause you to get $1,000,000 while two-boxing will cause you to get $1,001,000. Either way, two-boxing will cause you to get a better outcome than will one-boxing. So you ought to two-box.

Counterfactual decision theory agrees with this verdict, but for slightly different reasons. Here, the key thing is that the contents of the opaque box are *counterfactually* independent of your choice: they would be the same regardless of how many boxes you were to take. Suppose first that it is empty. Then, if you were to one-box, you would get $0, while if you were to two-box, you would get $1,000. Suppose next that it is full. Then, if you were to one-box, you would get $1,000,000, while if you were to two-box, you would get $1,001,000. Either way, you would be $1,000 richer if you were to two-box than if you were to one-box. So you ought to two-box.

Since both recommend two-boxing, causal and counterfactual decision theories are still on a par. Decision theorists have typically conflated the two or regarded the latter as a specific version of the former. But this seems not to have been on the basis of much argument. The most one finds is the observation that counterfactuals can be read in either a 'standard' way or instead in a 'backtracking' way, the observation that counterfactual decision theory coupled with backtracking counterfactuals would wrongly recommend one-boxing, and the claim that standard counterfactuals are 'causal.'[5]

Backtracking readings are made more salient by the addition of a 'have to have' phrase or other 'syntactic peculiarity' (Lewis 1986f). In Newcomb's Problem, the following backtracking counterfactuals are arguably true: First, supposing you actually two-box, if you were to have one-boxed, then the predictor would have to have predicted that (in which case you'd get $1,000,000). Second, supposing you actually one-box, if you were to have

[5]One example is Lewis (1986f, 326-8). Compare Skyrms (1980, 132-3).

two-boxed, then the predictor would have to have predicted *that* (in which case you'd get $1,000). Now, I have a hard time getting these backtracking readings, but I concede that others seem to get them more easily. In any case, counterfactual decision theory, coupled with the use of backtracking counterfactuals, would recommend one-boxing.

But why should we accept that standard counterfactuals are causal? Standard counterfactuals and causation certainly have something to do with each other. But so do causation and correlation, and so do counterfactuals and correlation. Perhaps the relationship between counterfactuals and causation is more intimate than these other two. Perhaps even some counterfactual analysis of causation is correct. But even defenders of such analyses, like Lewis (1986c), deny that the relationship is anywhere near as simple as the equivalence of 'A causes B' and 'if A were the case, B would be the case.' Even when backtracking counterfactuals are banished, counterfactuals and causation still come apart.

But even having banished backtrackers,[6] you might still worry that counterfactuals are too context-sensitive to bear the normative weight imposed upon them by counterfactual decision theory. If their truth value varies with the salience of certain possibilities or respects of similarity, how can we use them in deliberation? Perhaps the context-sensitivity of counterfactuals is a reason to favor causal (or evidential) decision theory.

Let me make three brief points in response.[7] First, it is not clear that such potential context-sensitivity is really a problem, for we already accept that modals in general, including deontic modals like 'should' and 'ought,' are context-sensitive (Kratzer 2012).

Second, insofar as context-sensitivity is a problem, it may be a problem for everyone. Causal claims may themselves be context-sensitive, especially insofar as causation is to

[6]Instead of banishing backtrackers, one might instead say that one-boxing and two-boxing are each permissible, depending on your context and on which readings of the counterfactuals are salient. See Horgan (1981) for discussion of backtracking counterfactuals and Newcomb's Problem.

[7]There is a fourth and more minor point to be made. It is linguistic expressions that can be context-sensitive. But counterfactual decision theory evaluates actions using your credences in counterfactual *propositions*, not counterfactual sentences, and the former are not themselves context-sensitive. Of course, one might then worry not about context-sensitivity, but about pluralism: there are many different counterfactual propositions that can be expressed by a given sentence (perhaps corresponding to different ways of weighting respects of similarity), and no way of privileging one over the others. I'll continue to discuss the problem in terms of context-sensitivity, however, as nothing much turns on whether it is framed in terms of context-sensitivity or in terms of pluralism.

be analyzed in terms of counterfactuals (Maslen 2004). And even claims about what credences you do or should have may be context-sensitive, for instance if it is a context-sensitive matter (i) how to weight charity against fit with behavior in ascribing beliefs and credences (Lewis 1974), (ii) what counts as your evidence (Greco 2017), or (iii) how to weight the various theoretical virtues (simplicity, fruitfulness, fit with the data, and so on) that determine the degree to which a body of evidence supports a hypothesis. In each of these cases, there may in fact be some privileged weighting (even if it is unknowable or indeterminate which one), or it may be that there is some range of admissible weightings, with agents or evaluators being permitted to arbitrarily choose some such weighting to work with (for permissivism about credences, see e.g., Schoenfield 2014).[8] But there is no special problem for counterfactual decision theory here.

Third, even if counterfactuals are somewhat context-sensitive, it is not a complete free-for-all. Lewis (1986a, 34) asserts the context-sensitivity of counterfactuals by appeal to the famous Quinean sentences 'If Caesar had been in command in Korea, he would have used the atom bomb' and 'If Caesar had been in command in Korea, he would have used catapults.' Lewis says that each is true in some contexts, but in no contexts are they true together. But these counterfactuals are *weird*. They involve imagining drastic departures from actuality which are hard to make sense of. Would Caesar have lived for two millennia? Or would he have been born later (essentiality of origins notwithstanding)? Or would he have been resurrected? Even if *these* counterfactuals exhibit a high degree of context-sensitivity, it is doubtful whether more ordinary counterfactuals, like those used in scientific and practical reasoning, are equally context-sensitive. And I'm not even convinced that each of the Quinean counterfactuals is (determinately) true in some context. I'm inclined to think that it is always indeterminate whether or not he would have used catapults, and also indeterminate whether or not he would have used the atom bomb.

[8]Compare Lewis (1986f, 330), discussing counterfactuals of indeterminate truth value in the context of Gibbard and Harper's theory: 'We can well afford to admit that the theory might fail occasionally to give a determinate answer. Indeed, I admit that already, for any version, on other grounds. I think there is sometimes an arbitrary element in the assignment of C and V functions to partly rational agents. No worries, so long as we can reasonably hope that the answers are mostly determinate.'

So far, then, we have no compelling grounds for favoring causal decision theory over counterfactual decision theory, or for regarding the latter as a special case of the former. The cases to follow, however, will give compelling grounds for regarding them as distinct and for favoring the latter over the former.[9]

3 Overdetermination

You want the window broken. You don't care how it happens; you just want it broken. You see that I've just thrown my rock at the window. If you do nothing, my rock will break the window. If you throw, your rock will hit mine and deflect it, and then *your* rock will hit the window, causing it to break. But it will also cause some unwanted energy expenditure.

Should you throw? Clearly not! This is what counterfactualism says. If you were to throw, the window would break and you would expend energy. If you were not to throw, the window would break and you wouldn't expend energy. So you shouldn't throw.

But the causalist guiding thought says you should throw. For if you throw, your action will cause the breaking of the window along with some energy expenditure. But if you don't throw, your action will cause neither.[10] And the goodness of the breaking outweighs the badness of the energy expenditure. (We'll see shortly that formalizations of causal decision theory correctly say that you shouldn't, but only by deviating from the causalist guiding thought.)

This is a case of overdetermination, and specifically of early pre-emption. The window

[9]Evidential decision theory also gives the correct verdict in the cases to follow. Evidential and counterfactual decision theory disagree, though, in Newcomb's Problem and various alleged counterexamples to causal decision theory in which causal and counterfactual decision theories give the same verdict.

[10]This could be questioned. Perhaps your non-throw—an omission—would be a cause of the breaking, for perhaps it would be a cause of my rock continuing on its course toward the window. There is much debate about whether there can be causation by omission. Lewis (2004) says 'yes,' but that this means causation isn't always a relation between events, since here one of the relata goes missing. Beebee (2004) says 'no' (and gives an error theory for why we might think otherwise), largely in order to preserve the claim that causation is a relation between events. I lean toward Beebee's view, but let me emphasize a 'meta' point: we know that you shouldn't throw even without settling issues in the metaphysics of causation like the possibility of causation by omission. This is itself a reason to favor counterfactual over causal decision theory, insofar as the latter holds rational decision-making hostage to practically irrelevant issues in the metaphysics of causation.

will break whether or not you throw. But if you throw, your throw will be an early pre-empting cause of the breaking. It will cut the causal chain that would have led from my throwing to the window's breaking, and then it will itself cause that event.

We could modify the case to make it one of late pre-emption or symmetrical overdetermination, and the same comments would apply. For late pre-emption, let the case be as before, except that if you throw your rock, it won't hit mine but will arrive at the window first, causing it to break. For symmetrical overdetermination, let it be that if you throw, your rock will arrive at the window at exactly the same time as mine. In these cases, counterfactual decision theory says not to throw, for the window would break regardless, and throwing would expend energy. But the causalist guiding thought says to throw, since if you throw, your throw will cause the window to break (along with some energy expenditure), while if you don't throw, your non-throw will cause neither. A caveat: there is no consensus about whether symmetrical overdetermining factors are causes. If they are, then the causalist guiding thought says to throw. The same goes if they count as causes-to-some-degree, assuming we would then discount the goodness of window's breaking by the degree to which your throwing causes it,[11] and assuming that the goodness of the window's breaking is sufficiently greater than the badness of the energy expenditure. If they don't count as causes at all, then the causalist guiding thought agrees with counterfactual decision theory here, but they still come apart in cases of early and late pre-emption.

Hitchcock (2013) considers overdetermination but argues that it is not a problem for causal decision theory. In his view, it shows that the relevant causal notion—the one that causal decision theorists should think matters for decision-making—is not actual causation, but rather *causal dependence* (see also Hall 2004, 268-9). Causal dependence is counterfactual dependence between disjoint events: e_2 causally depends on e_1 just in case (i) e_1 and e_2 are disjoint events, (ii) if e_1 were to occur, e_2 would occur and (iii) if e_1 were not to occur, e_2 would not occur. With overdetermination, an event such as your throwing might be an actual cause of an event such as the window breaking, even though the latter does not

[11] This is, in effect, the 'share-of-the-total' view criticized by Parfit (1984).

causally depend on the former.

Standard formalizations of causal decision theory agree with Hitchcock's emphasis on causal dependence over causation *per se*. I'll consider two such formalizations due to Joyce (1999) and Lewis (1986f). Joyce's formalization employs the notion of causal probability, where $P(X\backslash Y)$ is the causal probability of X given Y. Agents ought to perform the action with highest causal expected utility, defined as: $U(A) = \sum_i P(O_i\backslash A) \times u(O_i)$.

What are causal probabilities? Joyce (1999, 161) writes that 'the probability function $P(\cdot\backslash A)$ provides a measure of the agent's estimates of A's 'causal tendencies.'"[12] For this reason, '$P(X\backslash Y) > P(X\backslash Z)$ just in case the agent judges either that Y will causally promote X's occurrence more strongly than Z will or Y will causally inhibit X's occurrence less strongly than Z will' (162). Offhand, this suggests that where A's occurrence would be an overdetermining cause of O's occurrence, $P(O\backslash A)$ should be very high and $P(O\backslash \neg A)$ very low. For instance, your throwing will causally promote the window breaking more strongly than your not throwing. But then throwing your rock would have higher causal expected utility than not throwing, yielding the incorrect verdict that you ought to throw.

But Joyce distances himself from this intuitive understanding of causal probabilities. For he adds that '$P(X\backslash Y)$ will be high either when she thinks that Y will cause X or when she thinks that X is likely to hold whether or not Y does' (1999, 162).[13]

Lewis's (1986f, 313) formalization (which is similar to that of Skyrms 1980) employs the notion of a dependency hypothesis, which is a 'maximally specific proposition about how the things [the agent] cares about do and do not depend causally on his present actions.' Then, we can define causal expected utility as: $U(A) = \sum_j \sum_i P(K_j) \times P(O_i|A \wedge K_j) \times u(O_i)$.

So defined, not throwing has highest causal expected utility in our cases of early pre-emption, late pre-emption, and symmetrical overdetermination, as the reader can readily

[12]Compare Skyrms (1982, 696), who talks about replacing Jeffrey's conditional probabilities with 'a different probability function, CP, such that CP(C_i given A) represents the *causal propensity* for C_i given that action A is performed.' He goes on to analyze these causal propensity probabilities in terms of dependency hypotheses, along the lines of Lewis.

[13]Compare Gibbard and Harper (1981, 155), who write that 'To say $A \mathrel{\square\!\!\rightarrow} S$ is not to say that A's holding would bring about S's holding: $A \mathrel{\square\!\!\rightarrow} S$ is indeed true if A's holding would bring about S's holding, but $A \mathrel{\square\!\!\rightarrow} S$ is true also if S would hold regardless of whether A held.'

check. For in each of these cases, only one dependency hypothesis is needed: that I have thrown my rock, which is heading right for the window with enough speed to break it.

So these formalizations yield the correct verdicts in overdetermination cases, and they do so by concerning themselves, as Hitchcock suggests, with causal dependence rather than with causation *per se*. Let me make two points in response.[14]

First, in shifting from causation to causal dependence, these formalizations depart from the guiding thought that you should evaluate actions in terms of how likely they are to cause good results. You might think that I have foisted upon causal decision theory an unwanted guiding thought. But no. Causal decision theory's proponents and opponents alike highlight the same guiding thought. Start with proponents. Lewis (1986f, 305) begins his defense of causal decision theory by criticizing alternatives for neglecting the guiding thought that 'what's best to do is what the agent believes will most tend to cause good results.' Skyrms (1982, 695) writes that 'practical reason should evaluate actions in terms of their *causal consequences*.' Joyce (1999, 146) writes that 'Rational agents choose acts on the basis of their causal efficacy, not their auspiciousness; they act to bring about good results even when doing so might betoken bad news.' Hitchcock (2013, 138) writes that 'The guiding idea behind [causal decision theory] is that one should act in such a way that one's actions cause desirable outcomes.'[15] Now for opponents. Eells (1984, 177) describes causalism as the view that 'an act is evaluated according to the degree to which it can be expected to *cause*... good outcomes.' Egan (2007, 96) writes that the idea behind causal decision theory is that you should 'do what you expect will bring about the best results,' where *bringing about* is understood causally. And Ahmed (2014a, 8) says that the 'central

[14]There is a third point: causal dependence, for Lewis, is a relation between *events*, which may be individuated more finely than outcomes or actions. For instance, perhaps if you were to throw, it would result in a different window-breaking event than would occur if you were not to throw, for it would break at a slightly different time or manner. Or perhaps which window-breaking event would occur depends on exactly how you throw, with different possible throws constituting different events. But for purposes of decision theory, it doesn't matter whether events are individuated coarsely or finely. Indeed, events in this sense don't even show up in decision theory. All the more reason to take decision theory to be concerned with counterfactual dependence of outcomes on actions rather than with causal dependence between events.

[15]Similarly, Hitchcock (2016, 1161) says that 'Champions of (CDT) argue that we should evaluate our options by considering what our actions tend to *cause*' (emphasis in original).

idea of causalism' is that 'the practical value of an act is sensitive to its causal bearing upon outcomes of interest.'

Of course, we might say that all of these authors are just speaking loosely. Still, this departure is important. For a decision theory must be judged not only on whether its formula gives the right verdicts in various cases, but also on whether it is underpinned by a pretheoretically compelling idea about what matters in evaluating actions. The idea that you should perform the action most likely to cause good results is pretheoretically compelling, even though it turns out to be wrong. As for the alternative idea that you should perform the action such that good results are most likely to happen given that you perform that action and given how things causally independent of your action might be, weighted by your confidence that things are that way—well, it's a mouthful, and certainly less pretheoretically compelling than the original guiding thought. Moreover, while causation is a commonsense notion that is familiar to theorists and the folk alike, 'causal dependence' is a piece of technical jargon. I, for one, had never heard of causal dependence until I saw it introduced and defined by Lewis (1986c) in terms of counterfactual dependence.[16]

Second, and more importantly, shifting from causation to causal dependence will not help with the cases to follow. These cases involve counterfactual dependence without causal dependence, and so it is not merely causal decision theorists' slogan, or guiding thought, that is problematic; even their 'official theory' goes wrong.

4 Constitution

Sometimes, you want to do something *for its own sake*, and not for the sake of any knock-on effects. You might want to climb some unclimbed peak just for the sake of it, and not because doing so will cause thrills or fame. You desire the outcome of your climbing it. More down to earth, you might want to go for a walk just for the sake of it, and not because

[16]Google searches for 'cause' and 'causation' resulted in 1,430,000,000 and 21,400,000 hits, respectively while searches for 'causal dependence' and 'depends causally' resulted in 52,700 and 9,530 hits, respectively, with the main results being philosophy articles.

it will cause lower stress or better ideas. You desire the outcome of your going for a walk.

Supposing there are no downsides to climbing (or walking), should you do it? Clearly yes! Counterfactual decision theory agrees. If you were to climb the peak, then the outcome in which you climb it would obtain, and if not, not. But the causalist guiding thought says it's permissible not to. For your climbing the peak would not *cause* the occurrence of your peak-climbing—it would *constitute* it! And constitution is not causation.

Cases of constitution are neither rare nor unimportant. We very often 'just feel like' doing something or want to do it for its own sake. Constitution also matters for ethics. If you promised me not to watch the movie on your own before we watch it together, then your watching it would not *cause* the breaking of a promise. Rather, it would *constitute* it. And for non-consequentialists, it is thereby disvaluable (perhaps in an agent- and time-relative way) even if I never find out and it has no other negative causal effects.[17]

Consequentialists may also embrace goods which are constituted by actions. Perhaps such things as friendship, autonomy, and achievement are valuable (likely because they are components of well-being, which is itself valuable). And repaying a favor might constitute friendship, and not merely causally promote downstream friendship. Going to a protest might constitute autonomy, and not merely causally promote downstream autonomy. Climbing a mountain might constitute an achievement, and not merely causally promote downstream achievement.[18, 19]

[17] I have previously defended the thesis that a decision-maker's options consist only of mental acts of decision-making or intention-formation (Hedden 2012). But the problem of constitution will still arise, as it is plausible that such mental acts can also be valuable or disvaluable in their own right. For instance, it may be disvaluable to *intend* to cheat on one's partner, even if circumstances intervene and prevent one from following through.

[18] Pettit (2018) identifies three mistakes about doing good, one of which is thinking that the goods brought about by performing a given action must be causal effects of that action, thereby ignoring the possibility of goods that are constituted by it. Hurka (2014, 173) notes that many early twentieth century British moral philosophers, including ideal consequentialists like Rashdall, Moore, and Ewing, thought that acts could be intrinsically valuable, and, more generally, that there could be goods which are simultaneous with an act. Hurka (1993, 60) notes this implication of his own perfectionist consequentialism: acts 'sometimes embody perfection, and contribute to the good by instantiating it now rather than by allowing its occurrence later.'

[19] These cases of constitution involve events that are either identical or overlap completely. But similar problems will arise in cases where one event is part of another. Lewis (1986d, 259) gives the example of the event of writing 'Larry.' He says that the event of writing 'Larr' and the event of writing 'rry' are not distinct and that 'events that are not distinct cannot stand in causal dependence,' and so the former cannot be a cause of the latter. Suppose, then, that writing a book is valuable, and that you are deciding whether to write the

What do the formalizations of causal decision theory say about cases of constitution? Recall that for Joyce, the causal probability of X given Y, $P(X\backslash Y)$, has to do with how strongly the occurrence of Y would *causally* promote the occurrence of X. But then $P(Y\backslash Y)$ should be zero, for Y would causally promote Y's occurrence not at all! But Joyce also stipulates that $P(Y\backslash Y) = 1$. By what right? With this stipulation, along with the previous one to deal with overdetermination, we have moved further away from causal probabilities having to do with *causation*. Instead, they seem to be tracking probabilities of counterfactuals. (In fairness, Joyce ends up interpreting these causal probabilities in terms of counterfactual imaging, in which case my only objection is terminological; they should not be called 'causal' probabilities, and he should not self-identify as a 'causal' decision theorist.)

For Lewis, dependency hypotheses are maximally specific propositions about how things might be in ways that do not depend causally on your present actions. Offhand, this suggests that facts about which of your present actions is performed could be included in dependency hypotheses. After all, which of your present actions you perform does not *causally* depend on which of your present actions you perform! But then causal decision theory would crash, for each action A would be inconsistent with each dependency hypothesis K_j which entails $\neg A$. This would leave $P(O_i|A \wedge K_j)$, and hence the causal expected utility of A, undefined.[20]

Of course, we can stipulate that dependency hypotheses cannot include facts about which of your present actions you perform. And surely Lewis would agree. But then we have departed from the thought that dependency hypotheses specify how things might be in ways that do not depend *causally* on your actions. Instead, they are specifications of

first chapter. The latter action would not, by Lewis' lights, cause the desirable outcome of the writing of the book, nor would that outcome causally depend on that act. So causal decision theory may be unable to give the verdict that you ought to write the first chapter. By contrast, counterfactual decision theory faces no such trouble, for if you were to write the first chapter, you would likely write the book, and if you were not to write the first chapter, you definitely would not write the book. Thanks to an anonymous referee for pointing me to such cases.

[20]Compare Gallow (2020, 123), who sketches Lewis' version of causal decision theory and describes the dependency hypotheses as describing 'factors that are not causally downstream of your act.' Trivially, your action is itself not causally downstream of your action. See also Joyce (1999, 151), whose principle of 'Dominance with Causal Independence' prohibits acts which are dominated with respect to 'a partition of events that the decision maker regards as *causally independent* of her choice' (emphasis in original).

how things might be in ways that do not depend *counterfactually* on your actions. But this would mean a conversion to counterfactual decision theory.

5 Determinism

Consider the following case, adapted from Ahmed (2014a,b).[21] You must choose between two bets. Bet 1 pays \$1 if Q is true and costs you \$10 if Q is false. Bet 2 pays \$10 if Q is true and costs you \$1 if Q is false. You are certain that the laws of nature are deterministic (and would be no matter what you do).[22] Q is the proposition that the initial state of the universe and the laws entail that you take Bet 1. Because you are certain that the laws are deterministic and you will take one of the two bets, you regard $\neg Q$ as equivalent to the proposition that the initial state of the universe and the laws entail that you take Bet 2. Which bet should you take?[23]

Ahmed thinks you should take Bet 1. I agree. As a defender of evidential decision theory, Ahmed notes that it gives this intuitive verdict. For you regard yourself as certain

[21] Ahmed's case is briefly anticipated by Nozick (1969, 141): 'To get the mind to really boggle, consider the following.

	S_1	S_2
A:	10	4
B:	8	3

Suppose that you know that either S_1 or S_2 already obtains, but you do not know which, and you know that S_1 will cause you to do B, and S_2 will cause you to do A. Now choose! ('Choose?')'

[22] As we will see, some theorists hold that the laws of nature are counterfactually independent of your choice, and hence would be the same regardless of what you were to do. Lewis (1986a,e) thinks that the laws counterfactually depend on what you do, since if you were to act otherwise than you actually do, some small 'divergence miracle,' or violation of the actual laws, would occur, leading to your acting otherwise. Lewis does not say whether, supposing determinism is actually true, the laws of nature would still be deterministic if you were to act otherwise. Perhaps instead they would be indeterministic, such that whenever the actual laws say that one thing will happen if another does, the counterfactual laws merely say that the one thing has very high objective chance, condition on the other happening. But I think that Lewis' (1973) own best-systems analysis of laws suggests that the laws would still be deterministic; the best deductive system—the one which optimally balances simplicity against informational strength—will agree with the actual best deductive system for all regions of spacetime outside the region where the divergence miracle occurs while stipulating exactly what happens within that miraculous region. Then, the laws in that closest possible world where you act otherwise will be a deterministic variant of the actual laws, with a carve-out for the region in which the divergence miracle occurs. These deterministic laws would be much more informative than the aforementioned indeterministic ones, with only a slight decrease in simplicity.

[23] We assume the minimal compatibilist thesis that even if determinism is true, you are free to take Bet 1 or Bet 2 in whatever sense of 'freedom' is required for them to both count as options for you.

	Q	$\neg Q$
Bet 1	\$1	\$-10
Bet 2	\$10	\$-1

Table 1: Betting on History and Laws

to get \$1 if you take Bet 1 and certain to lose \$1 if you take Bet 2. We will see shortly that counterfactual decision theory likewise recommends Bet 1, albeit for different reasons.

What does causal decision theory say? Well, Q is a proposition about the initial state of the universe and the laws of nature. The initial state of the universe and the laws of nature are both beyond your causal control; you cannot cause either one to be different. Therefore, no matter how things beyond your causal control might be (i.e. no matter whether Q or $\neg Q$ is true), Bet 2 yields a strictly better outcome than Bet 1: if Q is true, Bet 1 yields \$1 while Bet 2 yields \$10; if $\neg Q$ is true, Bet 1 yields \$-10 while Bet 2 yields \$-1. Bet 2 thus causally dominates Bet 1, and so by causalist lights, you ought to take the former.

What about our formalizations of causal decision theory? I am actually not sure what Joyce's (1999) theory would say. It depends on what we say about the causal probability of an outcome given an action, where that outcome corresponds to you performing that action when circumstances beyond your causal control entail that you won't perform it. Joyce does not say what the causal probabilities are in such a case. Plausibly, they are undefined, in which case the causal expected utilities of Bet 1 and Bet 2 are undefined. The same is true of Lewis' formalization, as the causal expected utilities of Bet 1 and Bet 2 will involve conditional probabilities where the conditioned propositions—$Q \wedge \mathit{Bet}\ 2$ and $\neg Q \wedge \mathit{Bet}\ 1$—are metaphysically and epistemically impossible and hence have probability 0. Given the standard ratio analysis of conditional probability, these are undefined. And primitive conditional probabilities won't help, for proponents of primitive conditional probabilities (e.g., Hájek 2003) hold that probabilities conditional on metaphysical and epistemic impossibilities are still undefined (even if probabilities conditional on possible

but probability 0 propositions can still be defined).[24]

This is essentially the same problem raised for Lewis' formalization in the last section—something that doesn't causally depend on which of your present actions you perform can nonetheless entail which one you do. This means that if dependency hypotheses can specify anything that doesn't causally depend on your present action, then we'll have some dependency hypotheses which are inconsistent with some of your available actions, resulting in actions with undefined causal expected utility.

Lewis (1986f, 316) actually discusses the issue of a conjunction of an action hypothesis and a dependency hypothesis having probability 0. But he says that that case should never arise: 'Absolute certainty is tantamount to firm resolve never to change your mind no matter what, and that is objectionable. However much reason you may get to think that option A will not be realised if K holds, you will not if you are rational lower [$P(A \wedge K)$] quite to zero.' Clearly, he did not have in mind dependency hypotheses like ours, whose conjunctions with certain actions are impossible, and not merely wildly improbable.

Some philosophers have defended causal decision theory, or modifications thereof, in response to Ahmed. Joyce (2016, 226) argues that the case is not really a decision problem, for 'an agent who deliberates about a decision which is framed so that each state entails a single action (and outcome), is engaging in an epistemic exercise, not an agential one.'[25] Sandgren and Williamson (2021) defend *selective decision theory*, which involves first conditionalizing on the negation of all outcomes corresponding to an action-state combination where the state entails that the action won't be performed, and then applying

[24]Compare Skyrms (1980, 133), who also employs dependency hypotheses in his formalization of causal decision theory. He defines them as 'maximally specific specifications of the factors outside our influence at the time of decision.' Context makes clear that 'influence' is understood in causal terms in ways that will yield problems in deterministic cases. For he writes that 'in normal situations what has happened before the action is outside its influence' (ibid). But surely determinism would not by itself make for a situation being abnormal. And in deterministic situations, as we will see, what happened before the action is only causally, and not counterfactually, independent of the action. So as I read Skyrms, his formalization will also crash due to undefined conditional probabilities in Betting on History and Laws.

[25]This response to Ahmed seems to conflict with Joyce's own (1999) formalization of so-called 'causal probabilities' in terms of counterfactual imaging. Insofar as his formalization reflects a commitment to counterfactual, rather than causal, decision theory, Joyce should give the full-throated response to Ahmed that I develop below, rather than attempting to dissolve the problem on the grounds that it is not a genuine decision problem in the first place.

causal decision theory with respect to the resulting probability function. Solomon (2019) considers a variety of causalist responses, including one where agents apply causal decision theory relative to their credences conditional on their actions not being predetermined. While I won't discuss these arguments in detail here, let me make two quick points. First, we might agree with Joyce that there is something amiss with decision problems which are framed so that each state entails a single action. But the problem is precisely that causal decision theory allows propositions which entail a single action to count as states (dependency hypotheses). By contrast, counterfactual decision theory requires states to be counterfactually independent of actions; but states that entail single actions are not counterfactually independent of those actions.[26] So counterfactual decision theory will never put agents in the position that, for Joyce, means they are engaging in merely epistemic, and not agential, exercises. Second, while they deserve more consideration than I can give them here, the proposals from Sandgren and Williamson and from Solomon involve significant deviations from standard causal decision theory, and they are more complex than counterfactual decision theory.

Counterfactual decision theory gives the right verdict in this case: it says that you should take Bet 1. As briefly noted above, the crucial thing is that while Q and $\neg Q$ are *causally* independent of your choice, they are not *counterfactually* independent of it. More generally, supposing determinism is true, it follows that if anything—including your actions—were otherwise, then either the initial state of the universe would be different, or the laws of nature would be different, or both.[27] (Or an impossible world would obtain; I set this

[26]Strictly speaking, counterfactual decision theory, in the original form developed by Stalnaker and Gibbard and Harper, does not give a role to 'states' in its formalism. But the idea is that if we were to set up a decision matrix for purposes of applying dominance reasoning, counterfactual decision theory, even in that original form, would require that the propositions in the columns be counterfactually independent of the actions in order for such dominance reasoning to be legitimate. It would be more accurate to say that counterfactual decision theory will only rationalize dominance reasoning with respect to partitions such that each cell is counterfactually independent of the actions. Analogous comments apply to evidential decision theory. Even though Jeffrey's theory does not employ states in its formalism, it will only rationalize dominance reasoning with respect to partitions such that each cell is evidentially independent of the actions. See Gibbard and Harper (1981, §7-9) for further discussion.

[27]However, see Gallow (forthcoming) for an interesting attempt to avoid this conclusion by adopting a interventionist approach to causation. Very roughly speaking, we evaluate what he calls 'causal counter-

option aside, though see Nolan (1997) and Williamson and Sandgren (forthcoming).)

Philosophers disagree about whether it's the laws or the initial state that would differ depending on your action. Delving fully into this debate would go beyond the scope of this paper, but let me give a brief overview. Lewis (1986a) famously holds that if your action (or anything else) had been different, then the initial state would be exactly the same but the laws would be different. This is because were you to act differently, some small 'divergence miracle' or other—some violation of the *actual* laws of nature—would occur just before the time of your action, leading to your acting differently. The universe would then unfold in accordance with the actual laws. By contrast, philosophers like Bennett (1984, 2003), Loewer (2007), Goodman (2015), and Dorr (2016) hold that if your action had been different, then the laws would still have been the same, but the initial state would have been different.

Each side will attempt to make its seemingly counterintuitive verdict more palatable. Lewis combined his view of counterfactuals under determinism with a Humean, non-governing conception of laws of nature. Take Lewis' (1973) own best-system analysis of laws: the laws are the true generalizations that appear as axioms or theorems of the best deductive system for summarizing particular facts about the world, where the best deduc-

factuals' by looking at causal graphs and seeing what happens when we intervene on the variable which concerns whether the antecedent proposition A is true. In doing so, we hold fixed the values of any variables which are not causally downstream of the variable concerning whether A is true. Then, given a causal graph with a variable for 'whether a miracle occurs,' we hold fixed the actual value of this variable (namely, that no such miracle occurs) when intervening on the variable for whether A is true (changing it to 'true'). For this reason, the counterfactual 'if A were the case, then no miracle would occur' comes out true. Analogously, given a causal graph with a variable for 'what the initial state of the universe is,' we hold fixed the actual value of this variable when intervening on the variable for whether A is true. For this reason, the counterfactual 'if A were the case, then the initial state would have been as it actually was' comes out true.

I am not persuaded by Gallow's account, however. As he acknowledges, his framework cannot handle causal graphs with variables *both* for whether or not a miracle occurs *and* for what the initial state of the universe is, for the value of the variable for whether A is true cannot be fixed independently of the values for these other two variables. This is because certain values for these other two variables metaphysically entail what the value is for the variable concerning whether A is true. This means that his framework fails to assign a truth value to counterfactuals of the form 'if A were true, then it would have been the case that either there was a miracle or the initial state was different.' While Gallow accepts this result, it strikes me as deeply problematic. For one thing, such a counterfactual seems clearly true, assuming that A is actually false, for then the antecedent metaphysically entails the consequent. For another, it prevents his view (when coupled with an interventionist version of causal decision theory) from yielding any verdict about decision problems like Betting on History and Laws.

tive system is the one which strikes the optimal balance of simplicity and informational strength. On such a theory of laws, it's no big thing to say that were you to act differently, the laws would be different. For this is just to say that were you to act differently, then particular matters of fact would have be different in such a way as to be optimally summarized by a different deductive system. See Lewis (1986e), Mele and Beebee (2002), and Beebee (2000, 2003) for further discussion.

The initial state-varying view of Bennett, Loewer, Goodman, and Dorr can be made more palatable by holding that while the initial state would differ depending on your action, history would still be *almost exactly* the same up until the time of action. Lewis (1986a, 45) thought that this was a pipe dream, worrying that there may be no way of allowing the initial state to vary *at all* while keeping the whole past approximately similar to the actual past. But Dorr (2016) gives a persuasive rebuttal, emphasizing the continuous yet chaotic nature of the dynamics of our best deterministic physical theories. Their continuous nature makes it plausible that for any finite interval of time, we can vary the starting point of a system in a sufficiently small way that it will stay arbitrarily close (and hence differ only microscopically) to how it actually is throughout that interval. And their chaotic nature makes it plausible that there will be some such small variation which then leads to the counterfactual's antecedent being true. Each side will also emphasize that even though the (distant or recent) past would have differed depending on your action, this does not amount to backwards causation (Lewis 1986e, Dorr 2016).[28]

[28]In Lewis' case, this is because while some divergence miracle or other would have occurred in the recent past, had you acted otherwise, there is no *particular* divergence miracle that would have occurred. Rather, different divergence miracles would have occurred in different of the closest possible worlds where you act otherwise. Thus, by his definition of causal dependence, no particular past event causally depends on your action, and hence your action does not cause any past event. Dorr is concerned to make a *tu quoque* response to an objection that his account commits him to backwards causation: insofar as Lewis can avoid commitment to backwards causation, so can he, and so backwards causation is a problem for everyone or no one. And it is easy to see that Dorr could avail himself of a response similar to that of Lewis. While some different initial state 'event' or other would have occurred, had you acted otherwise, there is no *particular* initial state event that would have then occurred, and hence no initial state event that causally depends on your action, and hence your action does not cause any past event. (Note that this response relies on rejecting counterfactual excluded middle; embracing it would require some other way of avoiding commitment to backwards causation. See §6 for discussion of counterfactual excluded middle.)

What about backtracking counterfactuals? In §2, I proposed to banish them for purposes of counterfactual decision theory. But do they not reappear if we adopt the approach to counterfactuals under determinism

There is, of course, much more to be said. I will not take a stand on which view of counterfactuals under determinism is correct. For our purposes, the important thing is that it's a forced choice. You can't hold fixed both the initial state and the laws; something has to give. And moreover, it doesn't matter which view is correct. For either way, while Bet 2 *causally* dominates the intuitively preferable Bet 1, Bet 1 *counterfactually* dominates Bet 2.

Why is that? Let's first say more about the proposition Q. It says that the initial state of the universe and the (deterministic) laws entail that you take Bet 1. So Q can be seen as a disjunction of conjunctive claims about the initial state and the laws. It is implausible that the initial state of the universe entails that you take one option over the other, regardless of the laws. And it is implausible that the laws entail that you take one option over the other, regardless of the initial state. (At least, if they do, it's not determinism that is at issue.) Since there are two options, modeling this situation using Lewis' approach to counterfactuals would require two possible sets of laws, L_1 and L_2, and modeling it using the approach of Bennett, Loewer, Goodman, and Dorr would require two possible initial states, H_1 and H_2. For concision, I'll use a single model with the two possible sets of laws and the two possible initial states, which is suitable for either approach to counterfactuals. Suppose that both $H_1 \wedge L_1$ and $H_2 \wedge L_2$ entail that you take Bet 1, while both $H_1 \wedge L_2$ and $H_2 \wedge L_1$ entail that you take Bet 2 (Table 2).

We can then understand Q as the disjunctive proposition $(H_1 \wedge L_1) \vee (H_2 \wedge L_2)$ and $\neg Q$

advocated by Bennett, Loewer, Goodman, and Dorr? I think not. Again, there is a *tu quoque* response. We cannot just say that a counterfactual is a backtracker if it involves varying the past. After all, Lewis' counterfactuals also involve varying the past. The only ways to avoid varying the past are to embrace impossible worlds or to embrace what Dorr calls a 'big miracle' view, on which the past would have been exactly the same, but a big miracle would have occurred right at, and not before, the time of the antecedent. But, as he points out, the big miracle view has implausible implications.

How, then, can they distinguish between backtracking and standard counterfactuals? I offer no fully-developed view, though see Khoo (2017). But roughly, Lewis could say that while both backtracking and standard counterfactuals involve varying the past, standard ones hold the past exactly fixed for longer than do backtrackers. While standard ones hold the past exactly fixed until just prior to the antecedent time, backtrackers hold it exactly fixed only up until some earlier past time, perhaps one made salient earlier in the context, as on Khoo's account. And Bennett, Loewer, Goodman, and Dorr could say that while both backtracking and standard counterfactuals involve varying the entirety of the past in microphysical respects, standard ones hold the past approximately, or macrophysically, fixed for longer than do backtrackers. While standard ones hold the past approximately fixed until just prior to the antecedent time, backtrackers hold it approximately fixed only up until some earlier past time, perhaps one made salient earlier in the context.

	L_1	L_2
H_1	Bet 1	Bet 2
H_2	Bet 2	Bet 1

Table 2: The Deterministic Entailments

as the disjunctive proposition $(H_1 \wedge L_2) \vee (H_2 \wedge L_1)$. Now let's go through the case, first for the Lewis view and then for the Bennett/Loewer/Goodman/Dorr view.

For Lewis, the initial state of the universe is counterfactually independent of your choice, but the laws aren't (Table 3). Suppose first that H_1 is actual. In that case, if you were to take Bet 1, then L_1 would be true (along with H_1), and so Q would be true, and so you would win \$1. If you were to take Bet 2, then L_2 would be true (along with H_1), and so $\neg Q$ would be true, and so you would lose \$1. Suppose next that H_2 is actual. In that case, if you were to take Bet 1, then L_2 would be true (along with H_2), and so Q would be true, and so you would win \$1. If you were to take Bet 2, then L_1 would be true (along with H_2), and so $\neg Q$ would be true, and so you would lose \$1. So no matter how things are with the initial state of the universe, if you were to take Bet 1, you would win \$1, while if you were to take Bet 2, you would lose \$1. So you should take Bet 1.

	H_1	H_2
Bet 1	\$1	\$1
Bet 2	\$-1	\$-1

Table 3: Betting on History and Laws (Lewis Way)

For Bennett, Loewer, Goodman, and Dorr, the laws are counterfactually independent of your choice, but the initial state isn't (Table 4). Suppose first that L_1 is actual. In that case, if you were to take Bet 1, the H_1 would be true (along with L_1), and so Q would be true, and so you would win \$1. If you were to take Bet 2, then H_2 would be true (along with L_1), and so $\neg Q$ would be true, and so you would lose \$1. Suppose next that L_2 is actual. In that case, if you were to take Bet 1, then H_2 would be true (along with L_2), and so Q would be true, and so you would win \$1. If you were to take Bet 2, then H_1 would be true (along

with L_2), and so $\neg Q$ would be true, and so you would lose \$1. So no matter how things are with the laws, if you were to take Bet 1, you would win \$1, while if you were to take Bet 2, you would lose \$1. So you should take Bet 1.

	L_1	L_2
Bet 1	\$1	\$1
Bet 2	\$-1	\$-1

Table 4: Betting on History and Laws (Bennett/Loewer/Goodman/Dorr Way)

Summing up: While the (deterministic) laws and the initial state of the universe are both *causally* independent of your choice, at least one is not *counterfactually* independent of your choice. And whereas Bet 2 *causally* dominates Bet 1, Bet 1 *counterfactually* dominates Bet 2. So counterfactual decision theory recommends taking Bet 1, as desired.

This is not to say that it's all smooth sailing. We might consider bets on the initial state and the laws taken separately. First, consider Betting on the Laws (Table 5; Ahmed 2013, 2014a). Suppose that you're very confident in L_1. Ahmed thinks you should take Bet 3. After all, you're more confident in L_1 than L_2, and you have no control over the laws.

	L_1	L_2
Bet 3	\$1	\$0
Bet 4	\$0	\$1

Table 5: Betting on the Laws

Second, consider Betting on History (Table 6).[29] Suppose that you're very confident in H_1. Which bet should you take? It's tempting to think that you should take Bet 5. After all, you're more confident in H_1 than H_2, and you have no control over the initial state of the universe.

What does counterfactual decision theory say about these two cases? That depends on your views about the correct theory of counterfactuals under determinism. If you're

[29]See Williamson and Sandgren (forthcoming). This case is different from the one (Ahmed (2014a) calls 'Betting on the Past,' which is closer to the one I have called 'Betting on History and Laws.'

	H_1	H_2
Bet 5	\$1	\$0
Bet 6	\$0	\$1

Table 6: Betting on History

certain that Lewis is right, it says that it's permissible to take either Bet 3 or Bet 4, but that you should take Bet 5 over Bet 6. If you're certain that Bennett, Loewer, Goodman, and Dorr are right, then it says that you should take Bet 3 over Bet 4, but that it's permissible to take either Bet 5 or Bet 6.

To see this, let's fill in some details. Suppose that $H_1 \wedge L_1$ and $H_2 \wedge L_2$ both entail that you will take Bets 3 and 5, while $H_1 \wedge L_2$ and $H_2 \wedge L_1$ both entail that you will take Bets 4 and 6. Start with the case where you're certain that Lewis is right and that the initial state is counterfactually independent of your choice. Suppose that H_1 is true. Then, if you were to take Bet 3, L_1 would be true, winning you \$1, while if you were to take Bet 4, L_2 would be true, likewise winning you \$1. Now suppose that H_2 is true. Then, if you were to take Bet 3, L_2 would be true, winning you \$0, while if you were to take Bet 4, L_1 would be true, likewise winning you \$0. So regardless of how things are that are counterfactually independent of your choice—i.e. regardless of whether H_1 or H_2 is true—both bets would yield the same payoff. So it's permissible to take either. As for Bets 5 and 6, it's clear that you should choose Bet 5 over Bet 6. After all, you're certain that whether H_1 or H_2 is true is counterfactually independent of your choice, and since you're more confident in the former than the latter, you should bet on the former.

Now consider the case where you're certain that Bennett, Loewer, Goodman, and Dorr are right and that the laws are counterfactually independent of your choice. Clearly, then, counterfactual decision theory would recommend Bet 3 over Bet 4. After all, you're certain that whether L_1 or L_2 is true is counterfactually independent of your choice, and since you're more confident in the former than the latter, you should bet on the former. What about Bets 5 and 6? Suppose that L_1 is true. Then, if you were to take Bet 5, H_1 would be

true, winning you \$1, while if you were to take Bet 6, H_2 would be true, likewise winning you \$1. Suppose instead that L_2 is true. Then, if you were to take Bet 5, H_2 would be true, winning you \$0, while if you were to take Bet 6, H_1 would be true, likewise winning you \$0. So you're certain that regardless of how things are that are counterfactually independent of your choice—i.e. regardless of whether L_1 or L_2 is true—both bets would yield the same payoff, and so it's permissible to take either.

So if you're certain of either view of counterfactuals, then counterfactual decision theory gives a *prima facie* problematic verdict about one of these two cases. If you're certain of the Lewis view, then it says that it's permissible to take either Bet 3 or Bet 4 in Betting on the Laws, despite your high confidence in L_1. If you're certain of the Bennett/Loewer/Goodman/Dorr view, then it says that it's permissible to take either Bet 5 or Bet 6 in Betting on History, despite your high confidence in H_1.

One might worry that this constitutes a significant problem for counterfactual decision theory. This reaction would be mistaken, however. First, and most importantly, if you really are certain that the laws would vary depending on how you act, then it does seem permissible to bet against the laws you think are most likely to actually obtain. For you are certain that for each proposition about the laws, it would be true if you were to bet on it, and so you would win the same amount of money either way. Similarly, *mutatis mutandis*, if you are certain that the initial state would vary depending on how you act.

Second, and more tenuously, perhaps you ought to be uncertain about whether it is the past, or instead the laws, that would be different depending on how you were to act. If so, then counterfactual decision theory tells you to bet on the laws and the initial state that you think most likely obtain, i.e. to take Bet 3 over Bet 4 and Bet 5 over Bet 6. After all, you're plausibly certain that whether it is the laws or the initial state that would vary is itself counterfactually independent of your choice. And for each pair of bets, we have a situation where conditional on it being the one thing (the laws, or the initial state) that would vary, they're equally good, while conditional on it being the other thing that would

vary, the first is better than the second. Therefore, if you have some positive credence in each possibility about what would vary, Bet 3 and Bet 5 have higher counterfactual expected utility overall than Bet 4 and Bet 6, respectively. This gives the desired verdict that it's rationally impermissible to take Bet 4 over Bet 3 or Bet 6 over Bet 5.

How compelling is this response? One might worry that it involves a kind of semantic uncertainty that is irrelevant to ordinary decisions (i.e. decisions not concerning things like what to assert about semantics). If so, then I fall back on my first response. But it may be that it is (metaphysically) indeterminate whether it is the past that would vary or instead the laws, and this indeterminacy may warrant uncertainty that is relevant even to ordinary decisions. Or, as Dorr (2016, 267-70) speculates, it may be contingent whether it is the past or instead the laws that would vary. He suggests that in worlds where the laws are simpler (and hence easier to know and describe) than the initial state, it is the initial state that would vary, but that in worlds where the initial state is simpler than the laws, it is the laws that would vary. Then, ordinary, non-semantic, uncertainty about the relative simplicity of the laws and the initial state yields uncertainty about whether it is the laws or the initial state that would vary, in such a way as to yield the result that Bets 3 and 5 have higher counterfactual expected utility than Bets 4 and 6, respectively. Again, though, if you are unconvinced by this second response, you may still be convinced by the first.

Let me close this section by turning back to Ahmed. He considers and rejects the possibility of a decision theory based on counterfactuals which hold the laws fixed and allow the initial state of the universe to vary. (I think that his criticisms would apply equally to a decision theory based on Lewisian law-varying counterfactuals.) To begin with, he writes that such a decision theory advocates 'making choices as though causally unreachable aspects of the past turned upon one's present choice' and that 'If that doesn't stop it from counting as a Causal Decision Theory then I don't know what stops anything from counting as a Causal Decision Theory' (2014a, 139). Fair enough. It's not a causal decision theory; it's a counterfactual decision theory!

Ahmed also says that such a counterfactual decision theory is 'unmotivated' and 'arbitrary' (2014a, 139). Focusing on Loewer's own statistical mechanics-based, past-varying account of counterfactuals (referred to as '$>_{SM}$'), he writes,

> One can at least see the point of a decision theory that evaluates options on the basis of their causal effects. Similarly one can see the point of a theory that evaluates them on the basis of their news value. But to evaluate them on the basis of neither of these but rather on the basis of $>_{SM}$ is arbitrary. (2014a, 140)

Let me make four points in response. First, it prejudices the issue to contrast decision theories based on causal effects or news value with decision theory based on $>_{SM}$. The real contrast is with decision theory based on counterfactuals. It just turns out that if Loewer is right, $>_{SM}$ is the correct account of counterfactuals. An analogy: suppose that pain is identical with the firing of C-fibers. We might wonder why we should care about avoiding the firing of C-fibers. The obvious answer is that we should care about avoiding pain, and it just turns out that the firing of C-fibers is pain.

Second, as Dorr (2016, 276) notes in his discussion of Ahmed, and as we saw in §2, it is just as natural to invoke counterfactuals as to invoke causation in justifying two-boxing in Newcomb's Problem. And in response to the 'Why Ain'cha Rich?' objection, it's natural to say to the wealthy one-boxer that had she two-boxed instead, she would've been even richer, whereas if you, the poor two-boxer, had one-boxed instead, you would've been even poorer (Lewis 1981).

Third, as Gibbard and Harper (1981, 153) emphasize, it is plausible that 'rational decision-making involves conditional propositions.' DeRose (2011) talks about the 'conditionals of deliberation.' The candidate conditionals are indicatives and counterfactuals. Evidentialist reasoning can be expressed using indicatives,[30] and counterfactualist reasoning can, unsurprisingly, be expressed using counterfactuals. But there is no third type

[30]There is a complication stemming from Lewis's (1986b) famous 'triviality result' (really, one of four such results!), which shows that the probability of an indicative conditional cannot always equal the conditional probability of the consequent given the antecedent. If this is right, then evidential decision theory, which makes use of conditional probabilities, cannot be reformulated using probabilities of indicative conditionals.

of conditional that can be used to express causalist reasoning, in cases where causalist reasoning diverges from counterfactualist reasoning. (Of course, there are counterfactuals that employ causal terminology, but there are also indicatives that do so.)

Fourth, we have seen independent grounds for thinking that the causalist guiding thought is mistaken: rational agents should not evaluate actions by how likely they are to cause good outcomes. For that guiding thought would give incorrect verdicts in cases of overdetermination and constitution. Here, facts about what outcomes your actions would cause come apart from facts about what outcomes would obtain were you to perform them. And when these facts come apart, it is the latter that should guide your choices.

6 Counterfactual Excluded Middle

These arguments concerning cases of overdetermination, constitution, and determinism are largely independent of the details of how to formalize counterfactual decision theory, and that is why the discussion so far has been relatively symbol-free. In this section, I briefly discuss an important remaining issue for counterfactual decision theory and how to formalize it. At the start of this paper, I followed Stalnaker and Gibbard and Harper in defining counterfactual expected utility, the quantity to be maximized according to counterfactual decision theory, as $U_{\square\!\!\rightarrow}(A) = \sum_i P(A \mathrel{\square\!\!\rightarrow} O_i) \times u(O_i)$. This is the most straightforward and intuitive version of counterfactual decision theory. Moreover, as Bacon (forthcoming) notes, it has the merit of meshing well with a straightforward and intuitive gloss on the notion of actual, or objective, value/utility. Plausibly, we care about decision theory largely because following its dictates is in some sense an optimal means to getting what we ultimately care about; maximizing expected utility is an optimal means to maximizing actual utility. And it is natural to define actual utility in counterfactual terms. The actual utility of an action, even one that I don't in fact perform, is the utility of the world that *would* have obtained, had I performed it. Then, the counterfactual expected utility of an act, as defined above, is the expectation of its actual utility—a pleasing harmony.

But this definition of counterfactual expected utility, and this definition of actual utility, are suitable only if counterfactual excluded middle (CEM) holds. That is, they are suitable only if for all X and Y: $(X \mathrel{\square\!\!\rightarrow} Y) \vee (X \mathrel{\square\!\!\rightarrow} \neg Y)$. If CEM fails to hold, then (assuming classical logic) sometimes $X \mathrel{\square\!\!\rightarrow} Y$ and $X \mathrel{\square\!\!\rightarrow} \neg Y$ will both be false, and then the probability terms in the above definition of counterfactual expected utility will not sum to 1. Disaster.

CEM is controversial. While conceding that it is intuitively compelling and reflected in our ordinary way of speaking, Lewis (1973, 79-83) argued against it on grounds of arbitrariness and on grounds of a conflict with so-called *might*-counterfactuals. Suppose I have a fair coin but don't flip it. If I were to have flipped it, would it have landed heads? Or would it have not landed heads (i.e. landed tails)? First, to say that it would have landed heads rather than tails (or vice versa) would be arbitrary. What could make it the case that the one answer is true rather than the other? Worlds where the coin lands heads seem no more similar to the actual world than worlds where it lands tails, and vice versa. Second, it is plausible both that if I had flipped it, it might have landed heads, and that if I had flipped it, it might have landed tails. But Lewis thinks that the *might*-counterfactual (written as: $\diamond\!\!\rightarrow$) is the dual of the ordinary *would*-counterfactual: $X \mathrel{\diamond\!\!\rightarrow} Y =_{df} \neg(X \mathrel{\square\!\!\rightarrow} \neg Y)$. In that case, if the coin might have landed heads, had I flipped it, then it is false that it would have landed tails. And if the coin might have landed tails, had I flipped it, then it is false that it would have landed heads. But this would constitute a counterexample to CEM.

If we reject CEM, all is not lost, for there are various alternative definitions of counterfactual expected utility. Let me start off, though, with a non-starter. We might employ counterfactuals with chancy consequents: $U_{\square\!\!\rightarrow}(A) = \sum_x \sum_i xP(A \mathrel{\square\!\!\rightarrow} ch(O_i) = x)u(O_i)$. But insofar as we are inclined to doubt CEM, we'll also think that sometimes there are no precise chance facts that would obtain were you to perform a given action, for instance in cases of 'antecedent underspecification.' Here there is seemingly no exact manner in which you would perform that action, with different chance facts resulting from different manners of performing it. If you had walked home rather than biked, there is no precise amount of

time this would have taken you, for there is no precise manner in which you would have walked home. You might have walked a little more faster or a little more slowly, or taken a given corner more tightly or less tightly. But for that very reason, there is seemingly no precise objective chance function, yielding a precise chance of getting home in some exact amount of time, that would have obtained had you walked home.

But there are two alternative definitions that could fit the bill. First, we might employ counterfactual imaging.[31] Let P^A be the image of P on A, which is gotten from P by taking all the probability it assigns to any given $\neg A$ world and redistributing it evenly over the A-worlds which are 'closest,' in the sense relevant to counterfactual reasoning, to that $\neg A$ world.[32] Imaging could then be used to define counterfactual expected utility as: $U_{\Box\!\!\rightarrow}(A) = \sum_i P^A(O_i) \times u(O_i)$. Second, we might employ a K-partition along the lines of Lewis (1986e) and define counterfactual expected utility as: $U_{\Box\!\!\rightarrow}(A) = \sum_j \sum_i P(K_j) \times P(O_i|A \wedge K_j) \times u(O_i)$. But whereas Lewis defines the K's, which he calls 'dependency hypotheses,' in terms of causation (yielding problems raised earlier), a counterfactual decision theorist could employ dependency hypotheses defined in terms of counterfactuals, e.g., as maximally specific ways that things might be in ways counterfactually independent of your choice.[33]

Again, my arguments in §3-5 are largely independent of the status of CEM and the exact definition of counterfactual expected utility. But I prefer the original one from Stalnaker and Gibbard and Harper. It is simpler, it meshes neatly with an attractive definition of actual utility,[34] and the case against CEM is less compelling than it first appears.

[31] For discussion, see Lewis (1986e), Sobel (1994), Joyce (1999). As noted earlier, Joyce's use of counterfactual imaging is in tension with his claim to be a 'causal' decision theorist.

[32] If we deny the limit assumption, so that there are no closest A-worlds, then we would need to say e.g., that probability is restributed to A-worlds in proportion to their closeness to that $\neg A$ world.

[33] More exactly, they are conjuctions of a claim about which things are counterfactually independent of your choice and a claim about how those things are. Note that if most counterfactuals are false, as Hájek (ms) argues, then these dependency hypotheses will be 'thinner,' as there will be fewer things that would have been a given way regardless of which of your options you were to take.

[34] Bacon (forthcoming) shows that the definition of counterfactual expected utility based on imaging will also enable a credence-independent definition of actual utility such that the expected utility of an action is the expectation of its actual utility. However, the requisite definition of actual utility will be somewhat less natural than the definition of actual utility which meshes with the original definition of counterfactual expected utility from Stalnaker and Gibbard and Harper. Whether a definition of expected utility based on dependency hypotheses permits this mesh between expected utility and actual utility will depend on exactly

First, arbitrariness may be indicative of indeterminacy (Stalnaker 1981a). Take someone who is 180 meters tall. It may be arbitrary to say that he is tall rather than non-tall, or vice versa. But rather than rejecting excluded middle, we should say that it is determinate that he is either tall or non-tall but indeterminate which it is. Similarly, we might say that it is determinate that either if I had flipped the coin, it would have landed heads or if I had flipped the coin, it would have landed tails, but it is indeterminate which it is. Provided that you ought to respond to indeterminacy with uncertainty, this indeterminacy-based version of CEM would still enable us to use the original formula from Stalnaker and Gibbard and Harper.[35] (Alternatively, perhaps this indeterminacy yields indeterminacy in what you ought to do.)

Second, Stalnaker and many others reject Lewis' definition of the *might*-counterfactual as the dual of the *would*-counterfactual. Instead, they think that the meaning of the former is derived compositionally from the ordinary modal 'might' (often interpreted epistemically) and the *would*-counterfactual. 'If I had flipped the coin, it might have landed heads' means 'It might be that if I had flipped the coin, it would have landed heads,' which is perfectly consistent with 'If I had flipped the coin, it would have landed tails.'[36] Stalnaker (1981a) gives various data supporting this view.[37]

---

how dependency hypotheses are defined, but there are grounds for pessimism. And he shows there is no credence-independent definition of actual utility such that the evidential expected utility of an action is the expectation of its actual utility. Overall, then, if we care about having a natural gloss on credence-independent actual utility such that the expected utility of an action is the expectation of its actual utility, then we have strong reason to prefer the original Stalnakerian formula or one based on imaging over various rivals, and a smaller reason, rooted in simplicity, for preferring the Stalnakerian formula over one based on imaging.

[35] As a referee pointed out to me, there is an analogy with future contingents. On one view, future-tensed propositions about whether some fair coin will land heads may be indeterminate in truth value, but this does not prevent us from rationally assigning probabilities to them in the obvious way.

[36] Of course, defenders of CEM must also consider cases where 'might' takes narrow scope with respect to the counterfactual conditional and cases were it is interpreted non-epistemically, e.g., as having to do with nomological possibility or non-zero objective chance. But there is still no obvious conflict between *might*-counterfactuals and corresponding opposite *would*-counterfactuals. First, there is the case where 'might' is understood epistemically and takes narrow scope. Here, 'If A were true, it would be compatible with my knowledge that B' is logically consistent with 'If A were true, B would be false.' Second, there is the case where 'might' is understood in terms of nomological possibility or non-zero objective chance and takes narrow scope. Here, since $\Diamond P$ is logically consistent with $\neg P$ even on a nomological or chance-based interpretation of $\Diamond$, $A \mathrel{\Box\!\!\rightarrow} \Diamond B$ is consistent with $A \mathrel{\Box\!\!\rightarrow} \neg B$. Third, there is the case where 'might' is understood in terms of nomological possibility or chance and takes wide scope. Again because $\Diamond P$ is consistent with $\neg P$, $\Diamond(A \mathrel{\Box\!\!\rightarrow} B)$ is consistent with $\neg(A \mathrel{\Box\!\!\rightarrow} B)$, which by CEM is equivalent to $A \mathrel{\Box\!\!\rightarrow} \neg B$.

[37] Stalnaker (1981a, 99) says that the most important datum concerns conjunctions of a *might*-counterfactual

A full defense of CEM would go beyond the scope of this paper (see Hawthorne 2005, Williams 2010, Moss 2013, and see Hájek 2020 for a recent rebuttal), but I find it plausible enough for me to favor the original, simple version of counterfactual decision theory.

7 Conclusion

Counterfactual decision theory was developed prior to causal decision theory, with both intended largely to justify two-boxing in Newcomb's Problem. But counterfactual decision theory has typically been considered a specific formulation of causal decision theory, rather than as a separate and distinctive decision theory. This is a mistake. Indeed, there is a real irony here. The relationship between counterfactual decision theory and causal decision theory is rather like the relationship between causal decision theory and evidential decision theory. Correlation and causation often go together. But not always. And when they come apart, we face a choice as to which corresponding decision theory (if either) is correct. Similarly, causation and counterfactuals often go together. But not always. And when they come apart, as in cases of overdetermination, constitution, and determinism, it is counterfactual decision theory which gives the right verdicts, and for the right reasons.[38]

with the negation of the corresponding *would*-counterfactual, for instance, 'If I had flipped the coin, it might have landed heads, but it's not the case that if I had flipped it, it would have landed heads.' Such conjunctions are infelicitous. Stalnaker's theory accounts for this infelicity, since when the 'might' is read epistemically, they are Moore-paradoxical, with the form 'It might be that P, but it is not the case that P.' By contrast, on Lewis' account such conjunctions are consistent and also not Moore-paradoxical. Williams (2010) gives further evidence for CEM involving counterfactuals under extensional quantifiers.

[38]For helpful discussion, I would like to thank Andrew Bacon, Zach Barnett, Helen Beebee, Kevin Dorst, Alan Hájek, Caspar Hare, Daniel Muñoz, Philip Pettit, Daniel Stoljar, and Timothy Luke Williamson, as well as audiences at the Australian National University, the University of Leeds, the London School of Economics, Cambridge University, Oxford University, the University of St. Andrews, the University of Edinburgh, the Dianoia Institute of Philosophy at the Australian Catholic University, and the University of Sydney.

References

Ahmed, Arif. 2013. 'Causal Decision Theory: A Counterexample.' *Philosophical Review* 122(2): 289-306.

Ahmed, Arif. 2014a. *Evidence, Decision and Causality*. Cambridge: Cambridge University Press.

Ahmed, Arif. 2014b. 'Causal Decision Theory and the Fixity of the Past.' *British Journal for the Philosophy of Science* 65(4): 665-85.

Bacon, Andrew. forthcoming. 'Actual Value in Decision Theory.' *Analysis*.

Beebee, Helen. 2000. 'The Non-Governing Conception of Laws of Nature.' *Philosophy and Phenomenological Research* 61(3): 571-94.

Beebee, Helen. 2003. 'Local Miracle Compatibilism.' *Noûs* 37(2): 258-77.

Beebee, Helen. 2004. 'Causing and Nothingness.' In Collins, Hall, and Paul (eds), *Causation and Counterfactuals*, Cambridge, MA: MIT Press, 291-308.

Beebee, Helen and Mele, Alfred. 2002. 'Humean Compatibilism.' *Mind* 111(442): 201-24.

Bennett, Jonathan. 1984. 'Counterfactuals and Temporal Direction.' *Philosophical Review* 93(1): 57-91.

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. New York: Oxford University Press.

DeRose, Keith. 2010. 'The Conditionals of Deliberation.' *Mind* 119(473): 1-42.

Dorr, Cian. 2016. 'Against Counterfactual Miracles.' *Philosophical Review* 125(2): 241-86.

Eells, Ellery. 1981. 'Causality, Utility, and Decision.' *Synthese* 48(2): 295-329.

Eells, Ellery. 1984. 'Causal Decision Theory.' *Proceedings of the Biennial Meeting of the Philosophy of Science Association, vol 2*: 177-200.

Egan, Andy. 2007. 'Some Counterexamples to Causal Decision Theory.' *Philosophical Review* 116(1): 93-114.

Gallow, J. Dmitri. 2020. 'The Causal Decision Theorist's Guide to Managing the News.' *Journal of Philosophy* 117(3): 117-49.

Gallow, J. Dmitri. forthcoming. 'Causal Counterfactuals without Miracles or Backtracking.' *Philosophy and Phenomenological Research*.

Gibbard, Allan and Harper, William. 1981. 'Counterfactuals and Two Kinds of Expected Utility.' In Harper, Stalnaker, and Pearce (eds), *Ifs: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht: D. Reidel, 153-90.

Goodman, Jeremy. 2015. 'Knowledge, Counterfactuals, and Determinism.' *Philosophical Studies* 172(9): 2275-8.

Greco, Daniel. 2017. 'Cognitive Mobile Homes.' *Mind* 126(501): 93-121.

Hájek, Alan. 2003. 'What Conditional Probability Could Not Be.' *Synthese* 137(3): 273-323.

Hájek, Alan. 2020. 'Contra Counterfactism.' *Synthese* 199(1-2): 181-210.

Hájek, Alan. ms. *Most Counterfactuals are False*. Australian National University.

Hall, Ned. 2004. 'Two Concepts of Causation.' In Collins, Hall, and Paul (eds), *Causation and Counterfactuals*, Cambridge, MA: MIT Press, 225-76.

Harper, William and Skyrms, Brian. 1988. 'Introduction.' In Harper and Skyrms (eds), *Causation in Decision, Belief Change, and Statistics, II*. New York: Kluwer, ix-xix.

Hawthorne, John. 2005. 'Chance and Counterfactuals.' *Philosophy and Phenomenological Research* 70(2): 396-405.

Hedden, Brian. 2012. 'Options and the Subjective *Ought*.' *Philosophical Studies* 158(2): 343-60.

Hitchcock, Christopher. 2013. 'What is the 'Cause' in Causal Decision Theory?' *Erkenntnis* 78(1): 129-46.

Hitchcock, Christopher. 2016. 'Conditioning, Intervening, and Decision.' *Synthese* 193: 1157-76.

Horgan, Terence. 1981. 'Counterfactuals and Newcomb's Problem.' *Journal of Philosophy* 78(6): 331-56.

Hurka, Thomas. 1993. *Perfectionism*. New York: Oxford University Press.

Hurka, Thomas. 2014. *British Ethical Theorists from Sidgwick to Ewing*. New York: Oxford

University Press.

Jeffrey, Richard. 1983.*The Logic of Decision*, 2nd Ed. Chicago: University of Chicago Press.

Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Joyce, James. 2016. 'Review of Arif Ahmed, Evidence, Decision and Causality.' *Journal of Philosophy* 113(4): 224-32.

Khoo, Justin. 2017. 'Backtracking Counterfactuals Revisited.' *Mind* 126(503): 841-910.

Kratzer, Angelika. 2012. *Modals and Conditionals*. New York: Oxford University Press.

Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell Press.

Lewis, David. 1974. 'Radical Interpretation.' *Synthese* 23: 331-44.

Lewis, David. 1981. 'Why Ain'cha Rich?' *Noûs* 15(3): 377-80.

Lewis, David. 1986a. 'Counterfactual Dependence and Time's Arrow.' In *Philosophical Papers: Volume II*, New York: Oxford University Press, 32-66.

Lewis, David. 1986b. 'Probabilities of Conditionals and Conditional Probabilities.' In *Philosophical Papers: Volume II*, New York: Oxford University Press, 133-156.

Lewis 1986c. 'Causation.' In *Philosophical Papers: Volume II*, New York: Oxford University Press, 159-213.

Lewis 1986d. 'Events.' In *Philosophical Papers: Volume II*, New York: Oxford University Press, 241-69.

Lewis, David. 1986e. 'Are We Free to Break the Laws?' In *Philosophical Papers: Volume II*, New York: Oxford University Press, 291-298.

Lewis, David. 1986f. 'Causal Decision Theory.' In *Philosophical Papers: Volume II*, New York: Oxford University Press, 305-339.

Lewis, David. 2004. 'Void and Object.' In Collins, Hall, and Paul (eds), *Causation and Counterfactuals*, Cambridge, MA: MIT Press, 277-90.

Loewer, Barry. 2007. 'Counterfactuals and the Second Law.' In Price and Corry (eds), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford

University Press, 293-326.

Maslen, Cei. 2004. 'Causes, Contrasts, and the Nontransitivity of Causation.' In Collins, Hall, and Paul (eds), *Causation and Counterfactuals*, Cambridge, MA: MIT Press, 341-58.

Moss, Sarah. 2013. 'Subjunctive Credences and Semantic Humility.' *Philosophy and Phenomenological Research* 87(2): 251-78.

Nolan, Daniel. 1997. 'Impossible Worlds: A Modest Approach.' *Notre Dame Journal of Formal Logic* 38)4): 535-72.

Nozick, Robert. 1969. 'Newcomb's Problem and Two Principles of Choice.' In Rescher (ed), *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel, 114-46.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Pettit, Philip. 2018. 'Three Mistakes about Doing Good (and Bad).' *Journal of Applied Philosophy* 35(1): 1-25.

Sandgren, Alexander and Williamson, Timothy Luke. 2021. 'Determinism, Counterfactuals, and Decision.' *Australasian Journal of Philosophy* 99(2): 286-302.

Schoenfield, Miriam. 2014. 'Permission to Believe: Why Permissivism is True and What it Tells Us About Irrelevant Influences on Belief.' *Noûs* 48(2): 193-218.

Skyrms, Brian. 1980. *Causal Necessity*. New Haven: Yale University Press.

Skyrms, Brian. 1982. 'Causal Decision Theory.' *Journal of Philosophy* 79(1): 695-711.

Sobel, Jordan Howard. 1994. *Taking Chances: Essays on Rational Choice*. Cambridge: Cambridge University Press.

Solomon, Toby. 2019. 'Causal Decision Theory's Predetermination Problem.' *Synthese* 2019, 1-32.

Stalnaker, Robert. 1968. 'A Theory of Conditionals.' In Rescher (ed), *Studies in Logical Theory* (*American Philosophical Quarterly Monographs 2*, 98-112.

Stalnaker, Robert. 1981a. 'A Defense of Conditional Excluded Middle.' In Harper, Stalnaker, and Pearce (eds), *Ifs: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht: D. Reidel, 87-104.

Stalnaker, Robert. 1981b. 'Letter to David Lewis.' In Harper, Stalnaker, and Pearce (eds), *Ifs: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht: D. Reidel, 151-2.

Williams, J. Robert. 2010. 'Defending Conditional Excluded Middle.' *Noûs* 44(4): 650-68.

Williamson, Timothy Luke, and Sandgren, Alexander. Forthcoming. 'Law-Abiding Causal Decision Theory.' *British Journal for the Philosophy of Science*.