Law-Abiding Causal Decision Theory*

Timothy L Williamson & Alexander Sandgren

January 5, 2021

Abstract

In this paper we discuss how Causal Decision Theory should be modified to handle a class of problematic cases involving deterministic laws. Causal Decision Theory, as it stands, is problematically biased against your endorsing deterministic propositions (for example it tells you to deny Newtonian physics, regardless of how confident you are of its truth). Our response is that this is not a problem for Causal Decision Theory *per se*, but arises because of the standard method for assessing the truth of certain counterfactuals. The truth of deterministic laws is 'modally fragile' on the standard semantics for counterfactuals: if determinism is true and you were to do otherwise, the laws would be different. We provide two ways of avoiding this problem: 1) supplement the standard semantics for counterfactuals with impossible worlds, or 2) introduce rigid designators into the description of problematic decision situations. We argue that both of these approaches are well-motivated and can be readily incorporated into Lewisian Causal Decision Theory.

1 Introduction

Causal Decision Theory (CDT) has problems with the laws, deterministic laws of nature in particular. One such problem comes from Arif Ahmed [2013, 2014b], who argues that defenders of CDT cannot rationally endorse, or at least bet on, deterministic laws. But if you

*This is a preprint version of an article that has been accepted for publication in *The British Journal for the Philosophy of Science*, published by Oxford University Press.

are highly confident that some deterministic system of laws L holds, then you should, all else being equal, be permitted to endorse L. CDT therefore advises you to act against your credences, which is irrational. Or so the argument is supposed to go.[1]

We show that CDT can be adjusted to adequately handle cases involving deterministic laws. The problem with CDT as usually formulated is that it makes its recommendations based on how things would go were you to act in various ways, but such facts can be misleading in deterministic cases (for example, because doing otherwise might involve a law-violation). What we need then is to formulate CDT such that facts about the actual world (for example, whether some deterministic law L holds) appropriately constrain facts about what would happen were you to do otherwise (for example, what would happen were you to endorse L). Taking David Lewis' CDT as our backdrop, we provide two ways of achieving this.

Firstly (Section 5.1), we argue that CDT yields sensible verdicts when supplemented with an impossible worlds semantics for counterfactuals. By adding impossible worlds to our framework, we can consider worlds where the laws determine that you perform some act A but you do not do A. This allows us to treat the laws robustly: your doing otherwise would not make a difference to the laws. Given that robustness, CDT gives the right verdicts for the right reasons; indeed, the problem cases for CDT here are closely related to the cases the motivate the introduction of impossible worlds elsewhere. Secondly (Section 5.2), we demonstrate that a similar result can be achieved in the possible worlds framework by introducing rigid designators into the description of problem cases. This strategy might be appealing to those

[1]This is a distinct challenge from another in the neighbourhood: given the possibility of determinism, CDT might advise you to bet on the impossible (see Seidenfeld [1984], Ahmed [2013, 2014a], and Solomon [forthcoming]); for responses, see (Sobel [1988], Joyce [2016], Sandgren & Williamson [forthcoming], and Kment [unpublished]). Section 6 outlines the relationship between that problem and the present discussion. There are several challenges to CDT that we will not discuss here. In particular, we will not address the widely-discussed issue of decision instability (see, say, Briggs [2010] and Fusco [2017] for philosophers who take instability to be a challenge to standard formulations of CDT; see, say, Joyce [2012] and Williamson [forthcoming] for defenses of traditional CDT.)

who are skeptical of impossible worlds. It does, however, go wrong in a range of cases explicitly designed to block the introduction of rigid designators; Section 5.2.1 discusses how an act-state independence constraint avoids this worry.

In Section 6, we show that our proposal dovetails with a plausible solution to a distinct set of deterministic challenges to CDT raised by Ahmed [2014a], [2014b]. In Section 7, we sketch how our approach might be extended to non-counterfactual formulations of CDT.

The causalist can bet sensibly on the truth of laws. Doing so requires them to anchor the truth of certain counterfactuals to facts about the actual world. This can be achieved in a fairly conservative way, the result of which is that the causalist ends up on the right side of the laws.

2 Causation and Choice

Causal Decision Theorists are united by the conviction that rational agents act to causally promote the good. We take as our general framework Lewis' [1981] dependency hypothesis formulation of CDT.[2] Lewis builds causation into decision theory at the level of state-spaces: the states in a decision situation are causally independent of your choice and describe your influence over the things you care about (see Lewis 1981, p. 13). Lewis calls these causal state descriptions 'dependency hypotheses'. This framework is general in that it builds causation into decision theory without forcing us to adopt any particular analysis of causation. It will also play a crucial role in allowing us to distinguish between legitimate and illegitimate ways of spelling out certain decision problems.

Formally, a decision situation is a quintuple, $< \mathcal{A}, \mathcal{K}, O, Cr, u >$. Four of these elements are familiar from most decision theories: $\mathcal{A}$ is a set of acts that are available options; O is a set of outcomes over which you have preferences (think of these as the most fine-grained propositions that you care about); $Cr(\cdot)$ is a credence function representing your degrees of belief; and $u(\cdot)$ is a utility function representing your preferences. $\mathcal{K}$ is then the set of dependency hypotheses. In this setup, each dependency hypothesis $K \in \mathcal{K}$ is 'a maximally specific proposition about how the things [the agent] cares about do and do not depend causally on [their] present actions' (Lewis [1981], p. 11). We take $o_{A,K}$ to be the outcome of

[2]See, say, Thoma [2019] Section 3.3 for a general overview of the motivation for CDT.

doing A in K.[3] We then define an act's expected utility as:

$$U(A) = \sum_K Cr(K) \cdot u(o_{A,K})$$

CDT says that A is permissible if and only if $U(A)$ is maximal.

So what exactly is a dependency hypothesis? In particular, what does it mean to specify how outcomes do and do not causally depend on acts? In many cases, we can get by simply with an intuitive understanding of dependence: whether it rains or not is independent of my choice to take an umbrella, but my getting wet does depend on my choice. But for more exotic cases, we will need to adopt a more precise understanding of dependence.

One natural (and Lewisian) thought is that counterfactuals capture patterns of causal dependence and independence, so we should take dependency hypotheses as specifying the counterfactual consequence of each act. If we accept this thought, we can then think of each dependency hypothesis K as a conjunction of act-outcome counterfactuals, $A \mathrel{\square\!\!\rightarrow} o_{A,K}$. This, with some rearranging, allows us to calculate expected utility explicitly in counterfactual form:

$$U_{CF}(A) = \sum_K Cr(A \mathrel{\square\!\!\rightarrow} o_{A,K}) \cdot u(o_{A,K})$$

CDT is often simply presented in this form. But bear in mind that U_{CF} is an interpretation of U.[4]

3 Betting on the laws

Ahmed ([2013], [2014b]) presents the following case:

[3] In indeterministic cases, each dependency hypothesis might be compatible with multiple outcomes given each act. We will not deal with such cases here, so for simplicity we assume that each dependency hypotheses determines a unique outcome given each act.

[4] When we talk about 'causation' and 'causal dependence', we are talking about the sense of causation and dependence relevant relevant for decision-making. This is compatible with the thought that non-counterfactual analyses of causation might have applications in other contexts.

Betting on the Laws: You are a scientist about to publish a groundbreaking paper on whether the universe is governed by some system of deterministic laws, L. The paper will contain several strong theoretical arguments in favour of L and the results of numerous experiments confirming L. Your own credence in L is very high (only just short of 1). You just have to write the conclusion to the paper, in which you have two options: endorse L (call this O_1) or deny L (call this O_2). You care only about truth, such that the outcomes are:

	L	$\neg L$
O_1	Win	Lose
O_2	Lose	Win

Let $u(\text{Win}) = 1$, and $u(\text{Lose}) = 0$. Given this assumption, you clearly ought to endorse L. Indeed given the symmetry of the payoffs, you should endorse L whenever you are more confident in it than its negation.

The problem is that CDT seemingly permits you to deny L. Ahmed's argument ([2013], pp. 294-5) is essentially:

Premise 1: $(O_1 \mathrel{\square\!\!\rightarrow} \text{Win})$ entails $(O_2 \mathrel{\square\!\!\rightarrow} \text{Win})$.

Premise 2: If $(O_1 \mathrel{\square\!\!\rightarrow} \text{Win})$ entails $(O_2 \mathrel{\square\!\!\rightarrow} \text{Win})$, then $Cr(O_2 \mathrel{\square\!\!\rightarrow} \text{Win}) \geq Cr(O_1 \mathrel{\square\!\!\rightarrow} \text{Win})$.

Premise 3: If $Cr(O_2 \mathrel{\square\!\!\rightarrow} \text{Win}) \geq Cr(O_1 \mathrel{\square\!\!\rightarrow} \text{Win})$, then CDT permits O_2.

Conclusion: CDT permits O_2.

This argument seeks to show that you should be more confident in the claim 'I would win by denying L' than the claim 'I would win by endorsing L', and that CDT takes this as a reason to deny L.

Premise 3 follows by noting:

$$U_{CF}(O_1) = Cr(O_1 \mathrel{\square\!\!\rightarrow} \text{Win}) \cdot 1 + Cr(O_1 \mathrel{\square\!\!\rightarrow} \text{Lose}) \cdot 0$$

$$= Cr(O_1 \mathrel{\square\!\!\rightarrow} \text{Win})$$

And:

$$U_{CF}(O_2) = Cr(O_2 \mathrel{\square\!\!\rightarrow} \text{Win}) \cdot 1 + Cr(O_2 \mathrel{\square\!\!\rightarrow} \text{Lose}) \cdot 0$$

$$= Cr(O_2 \mathrel{\square\!\!\rightarrow} \text{Win})$$

Premise 2 follows straight from the probability calculus. **Premise 1** follows from some basic claims about the semantics of counterfactuals and the nature of determinism. Firstly, we characterise determinism in line with Ahmed ([2013], p. 290), following Lewis ([1979], p. 460): if two possible worlds are governed by the same deterministic laws, then if those worlds agree with respect to matters of particular fact at a time, they agree with respect to matters of particular fact at all times. On Lewis' [1979] semantics for counterfactuals, the following two claims are therefore true:

Claim 1: If you do A and L is true, then $\neg A \mathrel{\square\!\!\rightarrow} \neg L$ is true.

Claim 2: If you do not do A and $A \mathrel{\square\!\!\rightarrow} L$ is true, then $\neg L$ is true.

Both of these claims follow from the fact that the standard semantics evaluates counterfactuals by considering worlds that diverge from our own due to a small miracle. To see that Claim 1 holds, say that $A \wedge L$ is true. Since L is determinsitic, then L and facts about the past determine that you do A. The nearest $\neg A$-worlds are worlds that match the actual world until a small miracle just before A would have occurred. Since such worlds match the actual world at pre-miracle times, they cannot be L-worlds by our characterisation of determinism. So the nearest $\neg A$-worlds are $\neg L$-worlds. This gives $\neg A \mathrel{\square\!\!\rightarrow} \neg L$.

Claim 2 holds for similar reasons. If you do not do A, the nearest A-worlds match the actual world until a small miracle just before A. Since those A-worlds are L-worlds, they are worlds in which you are determined to do A. But those worlds match our world at pre-miracle times, so if L is true in the actual world, then you do A in the actual world. But you do not do A in the actual world, so the actual world is a $\neg L$-world.

Having established claims 1 and 2, we can easily see that $(O_1 \mathrel{\square\!\!\rightarrow} \text{Win})$ entails $(O_2 \mathrel{\square\!\!\rightarrow} \text{Win})$. There are two ways that $(O_1 \mathrel{\square\!\!\rightarrow} \text{Win})$ can be true:

$$O_1 \wedge (O_1 \mathrel{\square\!\!\rightarrow} \text{Win}) \tag{3.1}$$

$$\neg O_1 \wedge (O_1 \mathrel{\Box\!\!\rightarrow} \text{Win}) \tag{3.2}$$

Take (3.1) first. $O_1 \mathrel{\Box\!\!\rightarrow}$ Win requires the nearest O_1-worlds to be L-worlds, and since you actually do O_1, this requires the actual world to be an L-world. Therefore we have an instance of the antecedent in **Claim 1**, so can conclude that $O_2 \mathrel{\Box\!\!\rightarrow} \neg L$. This gives $O_2 \mathrel{\Box\!\!\rightarrow}$ Win.

Next, take (3.2). Here we have an instance of the antecedent of **Claim 2**. We can therefore conclude that the actual world is a $\neg L$ world. So $O_2 \mathrel{\Box\!\!\rightarrow} \neg L$ is trivially true, and hence $O_2 \mathrel{\Box\!\!\rightarrow}$ Win. So in both cases, (1) and (2), ($O_1 \mathrel{\Box\!\!\rightarrow}$ Win) entails ($O_2 \mathrel{\Box\!\!\rightarrow}$ Win), which completes the argument.

Informally, the fact that L is deterministic means that it is 'modally fragile'. The standard semantics for counterfactuals licenses the following kind of reasoning: 'say that L is true and I am determined to endorse L—then I would do no worse by denying L (because L would have to be false for me to deny it)! Or say that I am determined to deny L—then I would do no better by endorsing L (because L would have to be false for me to endorse it)!'. Standard formulations of CDT vindicate this reasoning and so give the wrong verdict in Betting on the Laws.

4 A Lewisian worry

To see how the causalist can handle this kind of case, begin by noting that we get the wrong verdict in Betting on the Laws based on a U_{CF} calculation. But, as already emphasised, this is a particular interpretation of Lewisian CDT. So we ought to ask whether we can reach the same verdict based on U. To do so requires us to specify the dependency hypotheses in Betting on the Laws.

What then are the dependency hypotheses in Betting on the Laws? Forget for a moment that you have ever thought carefully about the semantics for counterfactuals and the modal fragility of L. If you simply read the case, you would likely think that the obvious candidate for the partition is $\mathcal{K} = \{L, \neg L\}$ (call this the L-partition). Moreover, we followed Ahmed (2013, p. 291) in using the L-partition to specify the columns of our decision table. But note what happens if we calculate U in line with the L-partition:

$$U(O_1) = Cr(L) \cdot 1 + Cr(\neg L) \cdot 0$$

$$= Cr(L)$$

While:

$$U(O_2) = Cr(L) \cdot 0 + Cr(\neg L) \cdot 1$$

$$= Cr(\neg L)$$

This says that you should endorse L whenever you are more confident in L than its negation—the correct verdict! So we have a puzzle. We get two different verdicts when we use U and U_{CF} respectively, but these are supposed to be statements of the same theory.

The solution is that the U_{CF} calculation above implicitly involves rejecting the L-partition.[5] This raises some important questions: which is the correct partition to use, the L-partition or the partition implicit in the U_{CF} calculation? And is there a way of spelling out the contents of of the L-paritition such that its members qualify as dependency hypotheses? We answer that the L-partition is the correct partition and can be spelled out in a way amenable to the causalist.

5 Motivating the L-Partition

If the members of the L-partition are appropriate candidates for dependency hypotheses, then CDT gets things right. So, the question becomes whether L and $\neg L$ are appropriate candidates for dependency hypotheses. We think they are. First, note that everyone thinks that the laws are causally independent of your choice and outside of your influence. Moreover, it is the truth of L that determines the outcome of your choice. If L is true, then that means you win by endorsing L or lose by denying L. Similarly, if L is false, then that means you win by denying L or lose by endorsing L. These are the commonsense judgements that are underwriting our

[5]Instead, it assumes that the dependency hypotheses are $C_1 = \{O_1 \mathbin{\square\!\!\rightarrow} \text{Win}, O_2 \mathbin{\square\!\!\rightarrow} \text{Lose}\}$ and $C_2 = \{O_1 \mathbin{\square\!\!\rightarrow} \text{Lose}, O_2 \mathbin{\square\!\!\rightarrow} \text{Win}\}$, where these counterfactuals are interpreted in line with the standard Lewisian semantics.

intuitions in Betting on the Laws, and it is these judgements that our decision theory should respect. The truth of L tells you how the things you care about do and do not depend on your decision in an intuitive and plausible way, meaning that the members of the L-partition are good candidates for dependency hypotheses.

Sections 5.1 and 5.2 make the above thought precise. Section 5.1 introduces impossible worlds as an addition to the standard semantics for counterfactuals; this allows us to consider worlds that have the same laws as ours but in which you do other than you are determined to. Section 5.2 considers a plausible reading of Betting on the Laws involving rigid designators; when we interpret the case as a bet on whether the actual is an L-world, we guarantee that the truth of L is modally robust. Each of these strategies gives us the L-partition.

These strategies serve as proof of concept (in particular, we will not argue that they are the only ways of securing the L-partition). What matters is that the L-partition can be spelled out in a way amenable to the causalist; the key is to make the laws modally robust.[6]

5.1 Option 1: Impossible Worlds

CDT goes wrong because the standard semantics for counterfactuals makes L modally fragile: if L determines that you do A, then L is false at nearby possible worlds in which you do not do A. Now, you cannot causally influence the laws, but the standard semantics for counterfactuals tells us that your doing otherwise would involve the laws being different. This creates a tension: on the one hand, we are trying to analyse causal dependence in terms of counterfactuals; on the other hand, we are trying to block the inference from 'if I were to do otherwise, the laws would be different' to 'the laws depend on my acts'. The fact that Lewis analyses the latter in terms of the former might strike some as reason to reject counterfactual analyses of causation.

There is, however, another route. Your doing otherwise involves different laws at nearby

[6]Why introduce two solutions? Because, as we will see, they have different theoretical costs/benefits, so will appeal to different people. Moreover, they are not quite extensionally equivalent, though the differences do not seem to us to constitute a clear argument for one over the other.

possible worlds. But some are already convinced that we need impossible worlds to make sense of the relationship between action and the laws of nature. Nolan [2017] in particular argues that impossible worlds are needed to make sense of certain plausible judgements about counterfactuals. He argues:

> When considering a counterfactual situation, we initially assume the same fundamental principles of nature are at work [. . .] for most of these antecedents [i.e. antecedents of counterfactuals involving things going differently from the way they actually do], the relevantly similar worlds where they are true are not ones where the laws of nature differ from the actual laws. ([2017], p. 19)

Nolan is saying that while counterfactual scenarios involve things going differently, those differences do not generally go all the way to the laws of nature. My dropping a cup might make a difference to whether the cup breaks, but not to whether Newtonian physics is true. Of course, we could keep the laws the same if we were to tweak the distant past. But it also sounds wrong to say 'if I had dropped the cup, the initial stages of the big bang would have been different' (more on this in Section 5.1.1.). But then we have a puzzle: if the laws along with the past determine that I do not drop the cup, how could I drop the cup without different laws or a different past? This is Nolan's trilemma (see [2017], Section 3): tweak the past, tweak the laws, or give up on the possibility of doing otherwise.

Impossible worlds resolve the trilemma: we say that even if L along with the past determines A, there are L-worlds that match our own with respect to the past but in which A is false.[7] Instead of making the laws or past dependent on your choice, we enrich our conceptual toolbox with impossible worlds.

If we want claims like 'if I were to drop the cup, the past and laws of physics would be the same' to come out true, we have to say something about similarity. In particular, we will want it to be the case that the closest worlds in which you drop the cup are impossible worlds with

[7]Or, as a referee points out, this might not resolve the trilemma so much as sweeten the third horn: even if it is, in some sense, impossible for agents to do otherwise, impossible worlds still allow us to make sense of what would happen if they did otherwise.

the same laws and past as ours. Nolan crucially denies that the closest worlds are generally possible worlds by denying the following ([1997], p. 550):

> **Strangeness of Impossibility Condition:** Every possible world is closer to every other possible world than any impossible world is to any possible world.

But those who want to get the most out of impossible worlds had better deny this condition anyway (see Nolan [2017], p. 29). We can then retain the core of Lewis' similarity semantics for counterfactuals: $X \boxdot\!\!\rightarrow Y$ is true if and only if the closest X-worlds are all Y-worlds (where the closest X-worlds might be possible or impossible worlds). The crucial thing then is that we prioritise match with respect to the laws and past matters of fact, guaranteeing that your doing otherwise would not involve different laws. (We will not defend any precise similarity ranking here. Provided that it guarantees the modal robustness of the past and laws, the correct similarity ranking is independent of whether we include impossible worlds or not.) This leaves one further condition to spell out:

> **Localised Impossibility**: If the actual world is possible, then pro tanto, the more impossible facts or events (for example contradictions or law-violations) there are at a world w, the less close to the actual world w is. (What does 'pro tanto' mean here? At least that if the antecedent involves a law-violation at time t, then minimising impossible facts and events generally trumps securing match with respect to particular matters of fact from t onward.)

Nolan articulates a good reason to accept **Localised Impossibility**:

> The 'explosion' world—the impossible world where every proposition is true—is very dissimilar from our own [. . .] On the other hand, the world which is otherwise exactly like ours, except that Hobbes succeeded in his ambition of squaring the circle (but kept it a secret), is far less dissimilar. (Nolan [1997], p.544)

Localised Impossibility ensures that the explosion world is indeed more distant from our own since it has far more impossible facts and events (the impossibilities are extremely non-local). Hobbes' secretly squaring the circle, however, is comparatively close to ours.

Generally (see Nolan [1997], p.544)), we want counterfactuals with non-actual antecedents to have non-trivial truth values (that is, we do not want them all to come out as trivially true). And if simply adding extra contradictions to a world did not threaten to make it less similar to the actual world, then every such counterfactual would be true. For any counterfactual and any world world where the antecedent is true and the consequent is false, we could simply add that the consequent is also true without making that world less close to the actual world, thus making the whole counterfactual true.[8]

We think that the closest worlds generally have the same laws and past as ours. To ensure this, we might need to consider impossible situations. But apart from an initial 'divergence impossibility', which guarantees that the laws and past are unchanged, we want those worlds look as much like possible worlds as possible.

We are now in a position to see how this impossible worlds framework solves Betting on the Laws. In the impossible worlds framework, $L \wedge O_1$ no longer entails $O_2 \boxrightarrow \text{Win}$. If L is true, then the nearest worlds where you endorse L are always L-worlds, hence worlds in which you win; similarly, if L is false, then the nearest worlds where you deny L are similarly worlds in which you win. It follows that:

$$(O_1 \boxrightarrow \text{Win}) \iff L$$

[8]This all has implications for decision theory. Say that you will actually endorse L in Betting on the Laws, then consider the closest worlds in which you do not endorse L. But for **Localised Impossibility**, those worlds might be worlds in which you endorse L and do not endorse L (such worlds match the actual world with respect to an important fact: you endorse L). If the closest not-endorse-worlds are also endorse-worlds, then CDT is going to make odd recommendations; after all, in L-worlds where you both endorse L and do endorse L, you still win by endorsing L! So, you would do no worse by denying L. **Localised Impossibility** blocks this kind of absurd reasoning. Worlds in which you both endorse and do not endorse have extra impossibilities; so, despite increasing match with respect to matters of particular fact, they are more distant to the actual world than those in which you merely do not endorse (and hence, do not win). Thanks to an anonymous referee for raising this point.

$$(O_2 \boxrightarrow \text{Win}) \iff \neg L$$

So we get the L-partition: $L = \{O_1 \boxrightarrow \text{Win} \wedge O_2 \boxrightarrow \text{Lose}\}$, $\neg L = \{O_1 \boxrightarrow \text{Lose} \wedge O_2 \boxrightarrow \text{Win}\}$, where these counterfactuals are evaluated in line with the impossible worlds semantics for counterfactuals.[9,10]

Note the striking similarity between Nolan's motivations for introducing impossible worlds

[9]Ahmed argues any theory deserving to be called causal must endorse:

> **Weak Causal Dominance:** For options A and B, suppose that any A-world W_A is at least as good (for you) as any B-world W_B that matches W_A over all matters of particular fact that are causally independent of your choice between them. Then it is rational for you to realize A when B is the only alternative. ([2013], p. 301)

He then argues that any theory endorsing **Weak Causal Dominance** must (irrationally) endorse O_2. The impossible worlds proposal avoids this challenge. In Betting on the Laws, the antecedent condition of **Weak Causal Dominance** does not hold, so the principle imposes no restrictions in that case. Weak Causal Dominance says that *if* any O_2-world is better than every O_1-world that matches over causally independent matters of particular fact, then you ought to take O_2. Once impossible worlds are in the picture, not every O_2-world is at least as good as the O_1-worlds matching over causally independent matters of particular fact (there are O_2-worlds where you lose with the same past and laws as O_1 worlds where you win).

[10]In a blog post from 2014 'Decision-making under determinism', Wolfgang Schwarz suggests a response to Betting on the Laws that has much in common with our proposal. We are sympathetic to many of Schwarz's suggestions. Like our proposal, Schwarz's makes use of 'incomplete or inconsistent' outcomes. At the end of his post, however, Schwarz is somewhat uncomfortable with (or tentative about) appealing to such outcomes:

> Another aspect of these proposals that bothers me is that they make use of outcomes that are either incomplete or inconsistent.... But how do we assign values to contradictory outcomes? Or, how do we calculate the value of option A under condition K if the counterfactual consequences are not closed under conjunction? (Schwarz [2014])

and the commonsense intuition in Betting on the Laws. Nolan is concerned with the strangeness of judgements like 'had I done differently, the laws would be different'—the modal fragility of the laws at the linguistic level. Betting on the Laws is problematic because of precisely this fragility. We might even say that Betting on the Laws is a practical analogue to the linguistic problems that Lewisian semantics generates for counterfactuals. Just as it is foolish to assert that the laws would have been different had you acted differently, it is foolish to act differently so that the laws would have been different. Adding impossible worlds saves us from both kinds of folly.[11]

5.1.1 Why go Impossible?

Ahmed's case is only a challenge to CDT when combined with a semantics for counterfactuals on which the laws are fragile. We have argued that Nolan's impossible worlds semantics meets that challenge. But you might wonder whether there are other semantics for counterfactuals that avoid the fragility of the laws.[12] Indeed, Goodman [2015] and Dorr [2016] defend views on which 'had I done differently, the laws would be the same' comes out as true. Moreover, their account avoids introducing impossible worlds since they allow the past to vary at worlds in which you do otherwise. Since such semantics could underwrite the

Indeed on the account we have proposed, counterfactual consequences are not closed under conjunction, and impossible worlds get assigned utilities based on which outcomes hold at those worlds. Partly for the reasons we discuss above, we also agree with Schwarz that standard Savage-style decision theories do not help here. One can understand our proposal as a way of putting some flesh on the bones of Schwarz's proposal and answering some of the crucial questions with which he finishes his post. Thanks to an anonymous referee for drawing this post to our attention.

[11]We cannot here weigh the balance between impossible worlds' virtues and vices. Interested readers can consult Nolan [1997] for background. Bernstein [2016] discusses another application of impossible worlds to causal counterfactuals, those involving omissions. We want to make sure 'providing the foundation for a reasonable causal decision theory' is duly added to the tally of virtues.

[12]Thanks to an anonymous referee for pushing us to consider this point.

L-partition, would it not be simpler to adopt one of these possible worlds account (see Dorr pp. 274-6 for a discussion of Betting on the Laws)?

We might reply by suggesting, as Nolan does ([2019], Section 3.2), that it is odd to say that 'had I dropped the cup, distant past stages of the universe would be different', just as it is odd to say that 'had I dropped the cup, Newtonian physics would be false'.

Dorr responds (p. 252) to this kind of worry by noting that the past differences involved are small and involve highly specific facts (for example, the precise arrangement of atoms). And it is not absurd to say that *those* facts would have been different. People spend their lives trying to understand the laws, and we hope that they are able to succeed; by contrast, people tend not to spend their lives trying to understand the exact structure of the universe at past times. So, at the linguistic level, Dorr argues that the robustness of the past is less important than the robustness of the laws.

We might reply that even if the robustness of the laws is *more* important than the robustness of the past, the robustness of the past is also important. True, we tend not to be as concerned with the precise state of the universe at a time, but it is still a cost to say that 'had I dropped the cup, the initial state of the universe would have been different'. All else being equal, we should prefer an account that secures the robustness of the past *and* laws to one that secures only the robustness of the laws.

There is, however, a deeper problem with the Goodman-Dorr account: you are still forced to bet against your credences. Consider the following:

> Betting on History: You are about to publish a groundbreaking paper on whether the universe was ever in a particular state (call the proposition that the universe was at some point in that state '*H*'). The paper contains several strong arguments in favour of *H* and the results of numerous experiments confirming *H*. Your own credence in *H* is very high (only just short of 1). You just have to write the conclusion to the paper, in which you have two options: *endorse H* or *deny H*. You are also convinced that determinism is true.

Betting on History is structurally similar to Betting on the Laws. You are extremely confident in *H*, and your credences should guide your betting behaviour. The rational thing

therefore is to endorse H.

The Goodman-Dorr approach, however, permits you to deny H. Just as on Lewis' semantics you would not do worse by denying L, by allowing the counterfactual past to change you would not do any worse by denying H.[13] Note that this is about betting behaviour. Even if Dorr is correct that it is harmless to say, 'were I to do otherwise, H would be false', we can surely agree that it is irrational to bet against H while being overwhelmingly confident in H. If we attach monetary prizes to outcomes, then it is cold comfort to lose money while reflecting on the fact that your theory of counterfactuals captures ordinary linguistic judgements! This highlights the importance of considering not just our ordinary judgements about which counterfactuals are true, but the role that counterfactuals play in practical reasoning when deciding on the correct semantics. No decision theory should tell you to bet against your credences, and so even if the different-past view is plausible at the linguistic level, it cannot play the decision-theoretic role that an adequate semantics should play.

Might someone respond that Betting on History is less worrying because it is more far-fetched than Betting on the Laws? We think not. CDT still advises you to do something irrational in Betting on History, and the correct decision theory never recommends the irrational. Moreover, most causalists are already motivated by the desire to have a decision theory that can handle far-fetched cases (Newcomb's problem is itself quite far-fetched). Since there is nothing incoherent about Betting on History, it falls within the purview of decision theory and we ought to be able to say something sensible about it. The impossible worlds proposal tells you to bet on H, the correct verdict, while Goodman-Dorr style accounts do not.

5.2 Option 2: Laws Actually

Impossible worlds allow us to anchor the outcome of your choice to facts about the actual world. Here we present another way of achieving this within the standard possible worlds semantics.

What we care about in Betting on the Laws is whether L is true, and by that we plausibly mean that we care about whether the *actual* world is an L-world. Note that this reading of the

[13] The proof essentially follows that in Section 3, with H playing the role of L.

case makes reference to a particular world, the actual world. When scientists debate the laws, we should interpret them as debating not the laws in nearby worlds where people act differently, but the laws in *this* world. We ought to interpret Betting on the Laws to reflect this concern for the way things actually are.

To make this precise, take L to be the set of worlds at which L is true, but take $L_@$ to be the proposition that the actual world, rigidly designated, is an L-world. Initially we interpreted Betting on the Laws as a decision about whether or not to endorse L, but perhaps we should take it to be about $L_@$.[14]

Ahmed's argument against CDT fails on the $L_@$ reading. To see this, say that the actual world is an L-world. Then were you to act otherwise, L would be false. *But* the designated actual world is still an L-world. So, even within the possible worlds framework, we can get it to come out as true that 'if I acted differently, L would still be true', provided that we interpret that reference to the laws as a reference to the actual laws. It then follows that you would win by endorsing $L_@$ just in case $L_@$ is true (your endorsing $L_@$ would not change the truth of $L_@$). So we again get the biconditionals required for the L-partition:

$$(O_1 \boxdot\!\!\rightarrow \text{Win}) \iff L_@$$

$$(O_2 \boxdot\!\!\rightarrow \text{Win}) \iff \neg L_@$$

$\mathcal{K} = \{L_@, \neg L_@\}$ is therefore an appropriate set of dependency hypotheses. So CDT

[14]This requries that we can assign non-trivial probabilities to propositions like $L_@$. This might be problematic on some possible worlds accounts of probability theory, which assume that a proposition gets assigned probability 1 if it is true at every world. If $L_@$ refers the truth of L at the actual world, then you might think that $L_@$ is either necessarily true or necessary false, so must be assigned probability 1 or 0. If so, then that is a problem for the possible worlds account of probability theory: it cannot accommodate the necessary *a posteriori*. We assumed no particular account of credences in our framework, so we take it as a (reasonable) precondition for employing the rigid designator strategy that your theory can assign non-trivial probabilities to $L_@$.

correctly recommends endorsing whichever laws you are more confident (actually) hold.

How natural is the rigidified reading of the case? Two things can be said in its favour. Firstly, it strikes us as plausible that what we care about in our ordinary bettings and endorsings is how things actually are. At the very least, when we are betting on the truth of L, we take ourselves to be betting on the truth of something modally robust (that is, something whose counterfactual truth tracks how things actually are). $L_@$ guarantees this robustness. By contrast, the simple L reading involves endorsing something like 'the world I would end up in were I to act differently is an L-world', and it is unclear that we really care about propositions like that.[15] Secondly, distinguishing between rigidified and non-rigidified readings of cases

[15]We often talk about the actual world via an indexical; in particular, we take it for granted that we can talk about @ without being able to describe @ in great detail (for example , without knowing what the laws are at @). When asked which world we inhabit, we often just say '*this* one'. This means that though $L_@$ is modally robust, expressing $L_@$ (and being able to rationally bet on $L_@$) is modally fragile, in the sense that if someone tried to pick out @ from some other world with the use of an indexical, they would fail. (Agents at world W_1 saying 'L holds at the actual world' are expressing L_{W1} rather than $L_@$). This results in a kind of strangeness (thanks to a referee for raising this point and pushing us to consider it): say that L is true and you win by endorsing $L_@$, then your counterpart in nearby $\neg L$-worlds who denies L *loses* by denying $L_@$, even though the world they call 'the actual world' is a $\neg L$ world. We agree that this is an odd situation for your counterpart ('I said that the world is not actually L, and it's not! So why am I wrong?'). But we do not think that this strangeness undermines the general strategy of rigidification; it merely highlights the strangeness of thinking about indexical expressions that *you* express from somebody else's perspective. What matters for our purposes is that A) you inhabit @ and so can express $L_@$ via an indexical, and B) having expressed $L_@$, interpreting Betting on the Laws as a bet on $L_@$ yields the right verdicts. Notwithstanding the strange situation your counterparts find themselves in, it is incontestable that whatever you are betting on in Betting on the Laws is modally robust. The fact that $L_@$ is modally robust but L is not therefore counts as good reason to think that $L_@$ is the right kind of thing to be betting on. Of course, if you maintain that the strangeness described above is

rarely makes a difference. The things we typically bet on (whether it rains or not, who the next prime minister will be) would not change regardless of how you bet. More generally, dependency hypotheses specify what the outcome of your acts would be, so they are typically causally upstream of your choice and would be unaffected by your doing otherwise. So, for most decision situations, it makes no difference whether we adopt the proposition K or its rigidified counterpart $K_@$, assuming, of course, that the bet is being made from the actual world (@). Only when betting on odd things like laws does the distinction become relevant. Given then that our ordinary interests underdetermine the correct reading of the case, we should not hold too dogmatically to one reading. Since introducing rigid designators gets things right, we have good reason to introduce them.[16]

5.2.1 Objection: You Can't Stipulate a Reading

You might here be tempted to object: sure, CDT does fine on the rigidified reading of the case, but what about the non-rigidifed reading?[17] Can we not stipulate that the bet concerns L, not $L_@$? Call the case with this stipulation Betting on Laws*, in which case is Betting on Laws* not a counterexample to CDT?

In response, the causalist should insist on a plausible act-state independence requirement, which Betting on Laws* violates. So the possible worlds response is a divide and conquer one: get the right verdict in Betting on the Laws and deny that Betting on Laws* is a genuine decision on the grounds that it violates act-state independence.

Act-state independence is a familiar condition from Savage-style decision theories. The basic idea is that states tell us about aspects of the world that are unaffected by your choice, while acts are the part of the decision situation under your control. Since the world provides the states and you provide the acts, which state holds must be independent of your act. In the

overly puzzling, or you are more suspicious than we are about the kind of work indexicals can do, then that may push you towards taking the impossible worlds route.

[16]Note that Weak Causal Dominance ceases to apply on this strategy for the same reasons it did not apply on the impossible worlds strategy. There are worlds where L is false but you still win your bet on $L_@$.

[17]Thanks to Boris Kment for pushing us to respond to this point.

context of counterfactual CDT, we analyse independence in counterfactual terms as follows (following Gibbard [1986], discussed in Joyce [1999], Section 7.1):

> **Act-State Independence:** For any dependency hypothesis K: K if and only if $A \boxrightarrow K$ for all acts A.[18]

This says that if some dependency hypothesis holds, then it would hold regardless of how you act. In a slogan: you can't make a difference to how you can make a difference.[19]

Now, most formulations of CDT do not insist on Act-State Independence, so it is an extension to CDT as usually understood. But it is a well-motivated extension. One reason that it is rarely insisted on is that it is natural to think that dependency hypotheses satisfy act-state independence *automatically* (as does Lewis [1981], p. 13). Which act-outcome counterfactuals are true will typically be settled by facts apart from your choice (the outcome of each act is typically determined by factors outside your control). So, even though typical Lewisian CDT does not endorse act-state independence, we would prefer to think of it as *not needing* to endorse it under ordinary circumstances. But the fact that we can get by without act-state independence in ordinary cases does not settle whether we should endorse it in exotic cases. And if, as we will argue, act-state independence solves cases like Betting on Laws*, then we have good reason to accept it.

We now show that Betting on Laws* does indeed violate act-state independence. The candidate dependency hypotheses, which we will call the C-partition, are:

$$C_1 = \{O_1 \boxrightarrow \text{Win}, O_2 \boxrightarrow \text{Win}\}$$

$$C_2 = \{O_1 \boxrightarrow \text{Lose}, O_2 \boxrightarrow \text{Win}\}$$

$$C_3 = \{O_1 \boxrightarrow \text{Win}, O_2 \boxrightarrow \text{Lose}\}$$

[18]Might one claim that the causalist should take act-state independence to be about causal rather than counterfactual independence? That seems unmotivated at this point in the dialectic; having accepted that the relevant act-outcome dependence is counterfactual dependence, we should characterise act-state dependence in the same way.

[19]This, of course, leaves open statistical correlations between acts and outcomes.

$$C_4 = \{O_1 \boxrightarrow \text{Lose}, O_2 \boxrightarrow \text{Lose}\}$$

To see that this violates act-state independence, it suffices to show that there are situations in which some C_i holds but there is some O_j such that the corresponding $O_j \boxrightarrow C_i$ fails. Take the situation in which $C_1 \wedge L$ is true. This means that L determines that you take O_1, so $L \wedge O_1$ holds and we can derive $O_2 \boxrightarrow \neg L$ from **Claim 1**.[20] Given $O_2 \boxrightarrow \neg L$, consider those nearest $O_2 \wedge \neg L$ worlds. The counterfactual ($O_1 \boxrightarrow$ Win) need not be true at those worlds (they are $\neg L$-worlds, and given $\neg L$, it is false that you win your bet if you were to do endorse L).[21] So: if ($O_1 \boxrightarrow$ Win) is true in the actual world, that counterfactual dependence can cease to hold in nearby O_2-worlds. Hence:

$$\neg(C_1 \supset (O_2 \boxrightarrow C_1))$$

So the C-partition violates act-state independence, hence is not an appropriate set of states, provided we insist on act-state independence. We can therefore deny that Betting on Laws* is a coherent decision.[22]

[20]Recall: If you do A and L is true, then $\neg A \boxrightarrow \neg L$ is true.

[21]For example, O_1 might be compatible with the laws at some of those $\neg L$-worlds; or, a violation of the laws at those worlds need not result in L being true at all of the subsequent closest worlds.

[22]Maybe Ahmed can insist that the dependency hypotheses themselves should be rigidly designated, for example:

$$(C_1)_@ = \{(O_1 \boxrightarrow \text{Win})_@, (O_2 \boxrightarrow \text{Win})_@\}$$

$$(C_2)_@ = \{(O_1 \boxrightarrow \text{Lose})_@, (O_2 \boxrightarrow \text{Win})_@\}$$

$$(C_3)_@ = \{(O_1 \boxrightarrow \text{Win})_@, (O_2 \boxrightarrow \text{Lose})_@\}$$

$$(C_4)_@ = \{(O_1 \boxrightarrow \text{Lose})_@, (O_2 \boxrightarrow \text{Lose})_@\}$$

This partition does not violate act-state independence and gives the wrong verdict for the causalist. But this move again seems unnatural - why insist that L cannot be interpreted rigidly

What should the causalist say about Betting on Laws* then? Given that the C-partition fails, the causalist can either utilise impossible worlds, which recovers the L-partition, or stick to the possible worlds framework and deny that Betting on Laws* is a genuine decision. If the latter, then CDT offers no advice in Betting on Laws*.

Is this silence itself a problem? After all, if the correct verdict is to endorse L, then giving no advice still falls short of giving you the *right* advice.

We think that defenders of possible worlds formulations of CDT can reasonably deny that the correct advice in Betting on Laws* is to endorse L. This means providing an error theory for the intuition that you should endorse L: it might seem like you ought to endorse L, but that seeming is accounted for by the fact that you ought to endorse $L_@$ (and we can ordinarily equivocate between the two). But when you think carefully about what it means to bet on L, you are really betting on whether L is true *at the actual world*. As stated previously, a bet on L non-rigidly designated is something like a bet on 'the world I would end up in were I to act differently would be an L-world'. We have no strong intuition or argument that that is the kind of thing rational agents should be able to bet on in the first place. It therefore seems like no great cost to deny that Betting on Laws* is a coherent decision situation, which means that CDT provides the right advice in all coherent decision situations. (Note that this error theory does not deny that scientists can rationally bet on or endorse deterministic laws; it just means that we should interpret them as betting on the truth of the actual laws.)

This discussion highlights a difference between the two approaches we have suggested in response to Betting on the Laws. The impossible worlds framework did not distinguish between a bet on L and a bet on $L_@$ for the simple reason that prioritising match in laws means

---

but that the dependency hypotheses must be? Moreover, 'dependency hypothesis' is a term of art so the causalist can stipulate what they mean by it. You could interpret dependency hypotheses as lists of counterfactuals $A \boxrightarrow O$ or lists of counterfactuals $(A \boxrightarrow O)_@$. Betting on Laws* shows that we should not use the second reading, so we simply insist that dependency hypotheses are lists of non-rigidly designated counterfactuals. Again, these two understandings of dependency hypotheses typicaly coincide, so this stipulation is no great departure from standard CDT.

the truth of $L_@$ guarantees that nearby worlds are L-worlds. So the impossible worlds strategy can treat Betting on the Laws and Betting on Laws* uniformly as genuine decisions. We find both strategies plausible, so we will not opt for one over the other. But you can decide for yourself whether you think that one strategy has a greater balance of virtues over vices.

6 States that Inform and Determine

CDT can say sensible things about cases like Betting on the Laws. Earlier, we mentioned that CDT faces another important challenge: sometimes CDT advises you to bet on the impossible. More precisely, when dependency hypotheses carry information about what you are determined to do, some versions of CDT force you to consider what would happen if you were to act in a way that you know you are determined not to. Ahmed ([2014a], [2014b]) has recently shown that this leads to absurd verdicts. Here is an example (originally from Ahmed [2014b], adapted from Sandgren & Williamson [forthcoming], p. 6):

> Betting on the Past: You are choosing between two bets. A_1 pays out \$10 if P and costs \$1 if $\neg P$. A_2 pays out \$2 if P and costs \$10 if $\neg P$. P is the proposition that the actual universe at some past time was in state H and the laws are L; you know that $H \wedge L$ determines that you take A_2 and $\neg(H \wedge L)$ determines that you take A_1.

	P	$\neg P$
A_1	10	1
A_2	2	−10

Now, it seems clear here that you should *not* bet on P.[23]

One response you might have is that this case violates act-state independence and so does not constitute a genuine decision. If so, then CDT does not make the wrong recommendation.[24]

But the responses to Betting on the Laws developed here allow us to formulate Betting on the Past as a coherent decision, either by rigidifying P to $P_@$ or adopting a semantics for

[23] See the arguments in (Ahmed [2014a], [2014b], Solomon [forthcoming], and Sandgren & Williamson [forthcoming]).

[24] For a similar response see (Joyce [2016]).

counterfactuals on which the truth of P is robust.[25] And if CDT provides a verdict in Betting on the Past, then it provides the wrong verdict. (A_1 dominates A_2, and CDT never recommends a dominated act.) So, by providing the tools to deal with Betting on the Laws, you might think that we have taken the causalist out of the frying pan and into the fire.

Indeed, we can add some details to Betting on the Laws to cause similar problems.[26] Say that we stipulate:

> You face Betting on the Laws, except that you have now consulted your physicist friend. They tell you that they have made a remarkable discovery: the truth of L determines that you will not bet on L. What should you now do?

Adding this stipulation changes nothing with respect to your credences in L or your utility function. So, CDT will recommend the same thing in this version of the case as in the original: endorse L. But that is clearly absurd.

Things get worse on the rigid designator account.[27] Consider the following:

> Betting on Betting: Let $B_@$ be the proposition that you actually take the bet that I am offering you. This bet costs \$10 if $B_@$ is true but pays out \$100 if $B_@$ is false. If you turn down the bet, you break even. What should you do?

	$B_@$	$\neg B_@$
Bet	-10	100
Don't Bet	0	0

Again, CDT seems to go badly wrong here. If you think that you are unlikely to bet (perhaps you are not the kind of person who gambles much), then standard formulations of CDT recommend betting. But that is clearly absurd! No rational agent could bet against their own betting.

We agree that each of these cases is a counterexample to standard formulations of CDT. Essentially, once we can represent impossible outcomes in a decision table, standard

[25] This is hardly surprising since Ahmed himself originally formulated this case using rigid designators (see Ahmed [2014a], pp. 669-671).

[26] Thanks to a referee for pushing us to consider this case.

[27] Again, thanks to a referee for this case.

formulations of CDT take those impossible outcomes too seriously and so cannot distinguish between, say, the original Betting on the Laws and the variant in which you know what the laws determine. This brings us back to our initial dilemma: some want to reject CDT (for example, Ahmed [2013]), while others think that deliberation requires us to reject the possibility of determinism, at least while making decisions (for example, Solomon [forthcoming]).

Again, however, we want to remain causalists and accept the possibility of determinism. To that end, we sketch a recent proposal we have developed elsewhere (Sandgren & Williamson [forthcoming]) and show that our response to Betting on the Laws is compatible with the view proposed in that paper. In that paper we introduce a fairly conservative modification to CDT, 'Selective Causal Decision Theory' (SDT), which agrees with CDT in the majority of cases while departing from it in cases where states determine and inform. We can combine our approach here with SDT to get the correct verdict in both Betting on the Laws and Betting on the Past.

The heart of SDT is the observation that not every true counterfactual is relevant for practical deliberation (even if it is represented in the intuitive tabular representation of the situation). For example, in Betting on the Past, 'if P is true and I were to take A_1, I would do better' is true, but this fact is *irrelevant*. Why? Because if P is true, then your doing A_1 would violate the laws, and 'a counterfactual can be true but irrelevant when the only way of making its consequent true given its antecedent involves a violation of the actual laws' (p. 4). SDT disregards irrelevant counterfactuals and so gives the correct verdict in cases like Betting on the Past. Formally (p. 5), for act A, denote the set of law-violating outcomes D_A (an outcome $o_{A,K}$ is law-violating if you know that $A \wedge K$ violates the actual laws). The relevant measure of A's choiceworthiness is then its renormalised causal expected utility, denoted:

$$U_R(A) = \sum_K C(K|\neg D_A) \cdot u(A \wedge K)$$

This allows us to distinguish between the original Betting on the Laws, in which you ought to bet on L, and the variant Betting on the Laws in which you know that L determines you will not bet on L. In the original Betting on the Laws, L is deterministic but uninformative;

from your deliberative stance, both L and $\neg L$ are compatible with your taking either option. So, it makes sense to ask both, 'given L, what would the world be like if I were to endorse L?', and 'given L, what would the world be like if I were to deny L?'.[28] But things change when you know that L determines that you do not endorse L. In that case, the question 'given L, what would the world be like if I were to bet on L?' is no longer relevant. In particular, the outcome in which you win by endorsing L is law-violating, so SDT calculates:

$$U_R(\text{Endorse L}) = Cr(L|\neg L) \cdot 1 + Cr(\neg L) \cdot 0 = 0$$

$$U_R(\text{Deny L}) = Cr(L) \cdot 0 + Cr(\neg L) \cdot 1 = Cr(\neg L)$$

SDT therefore recommends denying L in the in which you know that L determines that you do not endorse L, which is correct.

We can similarly deal with Betting on Betting. Since $B_@$ determines that you do not actually bet, not betting when $B_@$ is true involves doing something you are certain you are determined not to, so you ought to disregard the outcome Bet $\wedge \neg B_@$. This means that $U_R(\text{Bet}) < 0$, meaning that SDT never recommends betting.

The point of this section has been to show that we can handle two distinct challenges for CDT together. Indeed, there are really two distinct questions here: 1) How do we ensure that the laws (and whatever else is outside your influence) are modally robust?, and 2) How do we ensure that reasoning counterfactually does not lead you to irrationally bet on what you are certain is impossible? It is important not to conflate these questions. Betting on the Laws raises the first question, and we have provided an answer to it in this paper. The second question is raised by Betting on the Past, and we address it with SDT. These moves are compatible. CDT and SDT both presuppose some account of dependency hypotheses, but neither need be wedded to the standard Lewisian method. One attractive package of views then is SDT combined with an impossible worlds counterfactual theory of dependency

[28] Since no outcomes are law-violating in the original Betting on the Laws, SDT coincides with CDT, which we have argued means endorsing whichever system of laws your are most confident in.

hypotheses, or the rigidified interpretation of cases like Betting on the Laws.

7 Non-counterfactual Causal Dependence

We have taken the contents of dependency hypotheses to be act-outcome counterfactuals. But there are plenty of other approaches to CDT arising from different ways of analysing dependence. We end by noting that the broad strategies adopted here are unlikely to be restricted to the counterfactual causalist. While we will not spell out the details of different proposals, what matters is that anyone who can motivate the L-partition has a solution to Betting on the Laws. In the broadest terms, dependency hypotheses should reflect the causal propensities of your acts (for example Skyrms [1982], pp. 696-697). And since everybody agrees that the laws are causally independent of your choices, dependency hypotheses should capture claims like 'if L is true, then betting on L results in me winning my bet' and 'if L is false, then betting against L results in me losing my bet'. To deny those claims would be to put L under your causal influence, which is precisely what the causalist should avoid.

In the broadest terms, the dependency hypotheses in Betting on the Laws should be:

$$L = \{\text{Win if } O_1, \text{Lose if } O_2\}$$

$$\neg L = \{\text{Lose if } O_1, \text{Win if } O_2\}$$

The goal for defenders of different accounts of causal dependence is to spell out the details of their theory such that the L-partition holds.

Skyrms, for example, adopts an account on which dependency hypotheses specify the chance of each outcome conditional on your act. Letting $Ch_K(X|Y)$ denote the chance of X conditional on Y given dependency hypothesis K, it seems reasonable to take conditional chance functions to be Popper functions (to be defined on chance 0 events) and then introduce the constraint $Ch_L(\text{Win}|O_1) = Ch_{\neg L}(\text{Win}|O_2) = 1$. This could most obviously be achieved by using the rigidification strategy, though there are other routes. Or we could follow those in the causal modelling literature who take causal dependence as primitive. Instead of trying to say precisely what it takes for outcomes to depend on acts, we start with a commonsense

understanding of when one event causes another and analyse more complex causal relationships in terms of more simple ones.[29] As we construct our causal models then, we take the fact that L is causally unaffected by your choice as a constraint, so that on our model O_1 (or O_2) results in Win (or Lose) just in case L, which is causally independent of your choice, is true (false).

Clearly much more needs to be said to precisify any of the above suggestions. We simply want to emphasise that those who prefer other frameworks are free to consider whether they can capture the L-partition. If they can, that is good news: we have a general solution to make CDT law-abiding. If not, that is also good news: we have a strong argument that our particular framework is the right one.

8 Conclusion

CDT can give the right advice, even when you are betting on the truth of deterministic propositions like L. We have argued that this can be achieved by supplementing the standard similarity semantics for counterfactuals with impossible worlds or by interpreting cases with rigid designators. These two approaches are not quite equivalent, but they share a common motivation and will agree in a wide range of cases. While determinism does not undermine CDT, it does highlight some interesting choice points in our formulation of CDT and some important issues surrounding the relationship between choice and determinism.

Funding

This research was supported by a Vetenskapsrådet Research Project Grant (2019-02786); Timothy Williamson was supported an Australian Government Research Training Program Scholarship Stipend.

[29] See, for example, Stern [2017].

Acknowledgements

Thanks to Christopher Bottomley, Alan Hájek, Christian Löw, Daniel Nolan, Wlodek Rabinowicz, Wolfgang Schwarz, Toby Solomon, Katie Steele, Jeremy Strasser, James Willoughby, Caroline Touborg, and audiences at the Umeå University's higher seminar in Philosophy.

Timothy Luke Williamson

Australian National University

School of Philosophy

Office 6.64, RSSS Building,

New Acton, Australia, 2601

timothy.williamson@anu.edu.au

Alexander Sandgren

Umeå Universtiy

The Department of Historical, Philosophical and Religious Studies

Umeå, Sweden, 901 87

alexander.sandgren@umu.se

References

[1] Ahmed, A. [2013]: 'Causal Decision Theory: A Counterexample', *Philosophical Review*, **122**, pp. 289-306.

[2] Ahmed, A. [2014a]: 'Causal Decision Theory and the Fixity of the Past', *The British Journal for the Philosophy of Science*, **65**, pp. 665-85.

[3] Ahmed, A. [2014b]: *Evidence, Decision and Causality*, Cambridge: Cambridge University Press.

[4] Bernstein, B. [2016]: 'Omission impossible', *Philosophical Studies*, **173**, pp. 2575-89.

[5] Briggs, R. [2010]: 'Decision-Theoretic Paradoxes as Voting Paradoxes', *Philosophical Review*, **119**, pp. 1-30.

[6] Dorr, C. [2016]: 'Against Counterfactual Miracles', *Philosophical Review*, **125**, pp. 241-86.

[7] Fusco, M. [2017]: 'An Inconvenient Proof: The Gibbard-Harper Collapse Lemma for Causal Decision Theory', *Proceedings of the 21st Amsterdam Colloquium*.

[8] Gibbard, A. [1986]: 'Characterisation of Decision Matrices that Yield Instrumental Expected Utility', in L. Daboni, A. Montesano, and M. Lines *(eds)*, *Recent Developments in the Foundations of Utility and Risk Theory* , Boston: Dordrecht Reidel.

[9] Goodman, J. [2015]: 'Knowledge, Counterfactuals, and Determinism', *Philosophical Studies*, **172**, pp. 2275-8.

[10] Joyce, J. [1999]: *The Foundations of Causal Decision Theory*, Cambridge: Cambridge University Press.

[11] Joyce, J. [2012]: 'Regret and Instability in Causal Decision Theory', *Synthese*, **187**, pp. 123-45.

[12] Joyce, J. [2016]: 'Arif Ahmed: Evidence, Decision and Causality', *Journal of Philosophy*, **113**, pp. 224-32.

[13] Kment, B. [unpublished]: 'Decision, Causality, and Predetermination'.

[14] Lewis, D. [1981]: *Counterfactuals*, Blackwell.

[15] Lewis, D. [1979]: 'Counterfactual Dependence and Time's Arrow', *Nos*, **13**, pp. 445-76.

[16] Lewis, D. [1981]: 'Causal Decision Theory', *Australasian Journal of Philosophy*, **59**, pp. 5-30.

[17] Nolan, D. [1997]: 'Impossible Worlds: A Modest Approach', *Notre Dame Journal of Formal Logic*, **38**, pp. 535-72.

[18] Nolan, D. [2017]: 'Causal Counterfactuals and Impossible Worlds', in C. Hitchcock and H. Murphy (*eds*), *Making a Difference*, Oxford: Oxford University Press.

[19] Sandgren, A. and Williamson, T. [forthcoming]: 'Determinism, Counterfactuals, and Decision', *Australasian Journal of Philosophy*, pp. 1-17.

[20] Schwarz, W. [2014]: 'Decision-making Under Determinism', available at <https://www.umsu.de/wo/2014/603>.

[21] Seidenfeld, T. [1984]: 'Comments on Causal Decision Theory', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, **2**, pp. 201-12.

[22] Skyrms, B. [1982]: 'Causal Decision Theory', *The Journal of Philosophy*, **79**, pp. 695-711.

[23] Sobel, J.H. [1988]: 'Infallible Predictors', *Philosophical Review*, **97**, pp. 3-24.

[24] Solomon, T.C.P. [forthcoming]: 'Causal Decision Theory's Predetermination Problem', *Synthese*, pp. 1-32.

[25] Stalnaker, R. [1968]: 'A Theory of Conditionals', in N. Rescher (*ed.*), *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, Oxford: Blackwell.

[26] Stern, R. [2017]: 'Interventionist Decision Theory', *Synthese*, **194**, pp. 4133-53.

[27] Thoma, J. [2019]: 'Decision Theory', in *The Open Handbook of Formal Epistemology*, The PhilPapers Foundation, available at <https://jonathanweisberg.org/pdf/open-handbook-of-formal-epistemology.pdf>.

[28] Williamson, T.L. [Forthcoming]: 'Causal Decision Theory is Safe From Psychopaths', *Erkenntnis*, pp. 1–21