Noname manuscript No.
(will be inserted by the editor)

Causal Decision Theory's Predetermination Problem*

Toby C.P. Solomon

Received: date / Accepted: date

Abstract It has often been noted that there is some tension between engaging in decision-making and believing that one's choices might be predetermined. The possibility that our choices are predetermined forces us to consider, in our decisions, act-state pairs which are inconsistent, and hence to which we cannot assign sensible utilities. But the reasoning which justifies two-boxing in Newcomb's problem also justifies associating a non-zero causal probability with these inconsistent act-state pairs. Put together these undefined utilities and non-zero probabilities entail that expected utilities are undefined whenever it is a possibility that our choices are predetermined. There are three ways to solve the problem, but all of them suffer serious costs: always assume that, contrary to our evidence, the outcome of our present decision-making is not predetermined; give up the reasoning that justifies unconditional two-boxing in Newcomb's problem; or allow epistemically impossible outcomes to contribute to expected utility, leading to the wrong results in a series of cases introduced by Arif Ahmed (2014a,b). However they choose to respond, causal decision theorists cannot remain silent: the intuitive tension between decision-making and the possibility of predetermination can be made precise, and resolving it will require giving up something. Causal decision theorists have a predetermination problem.

Keywords Causal Decision Theory · Decision Theory · Determinism · Predetermination · Free Will

* This is a post-peer-review, pre-copyedit version of an article published in *Synthese*. The final authenticated version is available online at: http://dx.doi.org/10.1007/s11229-019-02425-0 Please cite the published version when possible.

This research was supported by an Australian Government Research Training Scheme Scholarship.

Toby C.P. Solomon
Australian National University
E-mail: toby.solomon@anu.edu.au

1 Introduction

When making a decision, it is always possible that the outcome of your decision-making is predetermined—that the past and the laws of nature already determine what the outcome of your decision-making will be. This possibility makes trouble for causal decision theory (CDT). Roughly, the problem goes as follows.

If predetermination is possible then we must, at least sometimes, consider *inconsistent* act-state pairs in our decision-making—specifically the pairs consisting of an act and each of the states in which we are predetermined *not* to perform that act. But, because these act-state pairs are inconsistent, we cannot assign them any sensible utility—there is no fact of the matter about how good it will be if a contradiction is true. Call the conjunction of these two claims *Undefined Utilities*.

However, the proposition that you are predetermined not to perform an act is a proposition about the past and the laws of nature, both of which are—according to the reasoning that justifies two-boxing in Newcomb's problem—not causally influenced by your choice. CDT tells us that, when calculating expected utilities, the probability of any state which is not causally influenced by our choice should be equal to the unconditional probability of that state obtaining.[1] As such, in CDT's expected utility calculations, each of our inconsistent act-state pairs—performing an act together with each of the states where we are predetermined not to perform it—should be assigned a probability equal to the unconditional probability of the respective state obtaining. So long as we don't rule out predetermination in general, these unconditional probabilities should not be zero. Call the claim that these inconsistent act-state pairs have non-zero probability in CDT's expected utility calculations *Non-Zero Probabilities*.

Together these claims—Undefined Utilities and Non-Zero Probabilities—entail that CDT's expected utility calculations will involve summing over the product of an undefined term and a non-zero term. But such products and sums are undefined. As long as predetermination is a possibility, then, causal decision theory gives us no advice. This is the *Predetermination Problem*.[2]

This paper is an exploration of this problem and the possible solutions to it, with two related aims: First, to show that we can make precise the commonly voiced suspicion that there is some tension between rational decision-making and believing that one's choices are predetermined—it is not a phantom tension.[3] In particular, I will

[1] While Evidential Decision Theory tells us that what matters is the evidence that our acts give us about which state obtains, CDT tells us that what matters is what our acts will causally promote. An act can be evidence for a state which it does not causally promote. For example, our choice is evidence about, but does not causally influence, the prediction in Newcomb's problem. EDT tells us that in such situations we must use the conditional probability of the state obtaining if the act is performed. CDT, on the other hand, tells us that in such situations we should use the unconditional probability of that state obtaining—or, equivalently, that we should use causal probabilities defined so that they coincide with unconditional probabilities when there is no causal influence from our act to the state—since our act cannot make the state any more or less likely in the relevant causal sense.

[2] In fact this may only be *a* predetermination problem—I do not wish to rule out the possibility that predetermination might cause other problems for decision-making, or CDT specifically (perhaps to do with ought-implies-can principles or similar)—but it is *the* predetermination problem this paper is about.

[3] For some discussions of this tension see, for example, (Nozick, 1969, 141), or (Fernandes, 2016) and references therein. There is also a debate on the relationship between belief in determinism and decision-making (usually called deliberation) in the free will literature, see (McKenna and Pereboom, 2016, §12.3)

show that the Predetermination Problem applies even when we define CDT in a way that remains neutral on as many controversial issues as possible. Doing so shows that the tension is a fundamental one for CDT.

The second aim is to show that all of the solutions to this problem have costs; none of them will be acceptable to *all* causal decision theorists, even if each is acceptable to *some* causal decision theorists. Of course, different theorist's intuitions will favor different solutions. I take it, however, that which intuitions we should save in the end (and which we can explain away), and hence which solution is best, is as yet unsettled.[4]

We will proceed as follows: in §2 I introduce a version of CDT which only places minimal constraints on its key objects—acts, states, causal probabilities, and utilities. Non-Zero Probabilities and Undefined Utilities are precisified and defended in §2.1 and §2.2 respectively; thus setting up the problem. I then turn to the three possible solutions to the problem: First, in §3, the suggestion that we should, in making our decisions, assume that our choices are not predetermined—denying that Non-Zero Probabilities holds. I argue that, if so, we must either give up the idea that our degrees of belief should track our evidence or the standard picture of the relationship between practical and theoretical reasoning. I call this the *Freedom For Current Decision* (FFCD) solution. Second, in §4, I discuss (re)defining causal probabilities so that the inconsistent act-state pairs are associated with zero probability, while maintaining that predetermination is possible—again denying that Non-Zero Probabilities holds. I argue that doing so forces us to give up the standard reasoning that justifies two-boxing, no matter how high your credence that your choice is predetermined, in Newcomb's problem. I call this the *Predetermination Possible, Inconsistency Impossible* (PPII) solution. Third, in §5, I discuss assigning utilities to the the inconsistent act-state pairs—primarily by appeal to subjunctive conditionals—and thus denying that Undefined Utilities holds. This solution clashes with the idea that impossible outcomes cannot be relevant to practical decision-making, giving the wrong results in cases of a kind recently introduced by Arif Ahmed (2014a,b). I call this the *Subjunctive Utilities* (SubU) solution.

Causal decision theorists are thus faced with a trilemma: they must choose one of these options, but none of them are entirely attractive. In a slogan: causal decision theorists have a predetermination problem.

---

for a short introduction to this debate and further references. Finally, there is the debate about whether it is possible to have credences about one's own acts while engaged in decision-making. The negative answer to this question seems to be at least partly motivated by the worry that believing that one's choices are predetermined might conflict with decision-making. See (Hájek, 2016) for a critical introduction to that debate. Note that most of these discussions are explicitly phrased in terms of global *determinism*, rather than *predetermination* of particular choices. This is, however, a mistake. It is no help to my decision-making that there might be indeterministic events at other times and places in the universe; the relevant question is always whether this particular choice is predetermined.

[4] I do not intend the Predetermination Problem, or the trilemma it gives rise to, as an objection to CDT, but as a challenge to be solved by further argument and refinement of our intuitions. Those already skeptical of CDT might, perhaps, turn it into an objection to CDT. But, even if not, I hope that they can learn something from understanding the Predetermination Problem and the challenge it raises for CDT.

2 CDT and The Predetermination Problem

To make the Predetermination Problem precise, and to understand the arguments for Non-Zero Probabilities and Undefined Utilities, I first need to introduce a definition of CDT. This is complicated by the fact that CDT is not a single theory, but rather a family of theories which agree on the central claim that you should take both boxes in Newcomb's problem, no matter how high your credence that the prediction is correct (I shall take this qualification as read going forward). There are many further issues on which the members of this family disagree. Here I will introduce a schematic version of CDT that allows us to understand the Predetermination Problem while remaining neutral on as many issues as possible. Doing so will allow us to see that the problem applies to any formulation of CDT which meets a very minimal set of conditions, and, in particular, to all available versions of CDT. This is not to say that none of these formulations have a solution to the problem—indeed solutions to the Predetermination Problem are easy to find—but that all such formulations must solve the problem by (implicit) appeal to one of the solutions discussed below, and must, therefore, face the costs that apply to that solution.[5]

Unfortunately there is one issue on which it is impossible to remain neutral: there is no formulation, however schematic, of CDT that can accommodate both *partition-invariant* and *partition-dependent* versions. A decision theory is partition-invariant iff it gives the same result for a particular decision problem no matter how we divide up the possible ways the world might be into a set of exclusive and exhaustive states (a partition). Partition-dependent formulations allow expected utility to depend on our choice of partition but pick out one (kind of) partition as the privileged one that we must use. The schematic version of CDT I will use is a partition-invariant formulation. Partition invariance is a valuable property if we can have it. Nonetheless, partition-dependent formulations of CDT are influential and I will discuss in §6 why they do not escape the Predetermination Problem. I leave this to the end of this paper for the sake of convenience; it is easiest to understand how the problem applies to such formulations after we already understand how it applies to partition-invariant formulations.

I will here define CDT as the theory that one may rationally perform an *act*, A, if and only if it maximizes the quantity $EU(A)$ when compared to all other available acts in a decision problem.[6] For any partition $\mathbb{S}$ of the possible *states*, $EU(A)$ is defined as:

$$EU(A) = \sum_{S_i \in \mathbb{S}} P(S_i \| A) U(A \wedge S_i)$$

[5] For example, the most obvious understanding of Lewis (1981) is committed to using subjunctives to define utilities as in the SubU solution; interventionist decision theories like Pearl (2000) are plausibly understood as being committed to the FFCD solution; as, more explicitly, is Joyce (2016, 226); and at least two authors, Cantwell (2010, 2013) and Edgington (2011), have suggested versions of CDT which endorse the PPII solution.

[6] A decision problem is sometimes just taken to be a set of acts and states. I prefer to think of a decision problem as a concrete situation facing a decision maker, and the corresponding set of acts and states as (one among) the correct description(s) of that situation.

What makes this formula distinctive of *causal* decision theory—rather than expected utility maximizing theories generally—is that we understand the probabilities $P(S_i||A)$ causally. Exactly what *causal probabilities* amount to, apart from being a kind of credence, is a matter of some controversy, and not something I can settle here. Very roughly, causal probabilities are just the credences which encode an agent's beliefs about the causal structure of the world. I can, however, give some *constraints* on any account of causal probabilities that will allow us to understand the Predetermination Problem. Before doing so it is important to define the key terms *act* and *state*.

The acts that decision makers are concerned with are all the physical movements and mental operations which they could initiate by deciding to do them, at the time of decision-making—or, in a formal model, the propositions which express the performance of these things. There are, however, two importantly different ways of understanding acts more formally in the literature, and I should be clear about which one I am using: First, following Savage (1972), one might understand acts as functions from states to *outcomes*, where outcomes are the direct objects of utility evaluations. Second, following Jeffrey (1983) and Joyce (1999), one might understand acts as simply those propositions which the decision maker can make true *at will*. The discussion which follows assumes the second picture of acts. However, everything which is said here can be translated into a Savage style understanding, which does not help to solve the problem.

In the present context we need to be careful about how we understand the claim that a proposition can be made true at will: in particular we must allow that an agent can engage in decision-making about acts which they are not *certain* they can perform at will. In other words, we must allow acts to be possible relative to some states *but not others* in a decision problem. We should understand an act as a proposition that the decision maker believes they *may* be able to make true at will. If, instead, we required an agent be certain that they can perform an act (as, for example, Sobel (1994, 101) suggests) then it would be impossible to consider the possibility of predetermination in decision-making; considering that possibility *entails* considering the possibility that only one act is compatible with how the actual world is. Requiring certainty about which acts can be performed would, therefore, amount to ruling out by definitional fiat the possibility of taking account of predetermination. But this will not do—even if it turns out that we must assume our choices are not predetermined in order to engage in rational decision-making, as the FFCD solution asserts, the argument for this must be given on substantial grounds, not by fiat definition of what counts as an act.[7][8]

---

[7] This might be different if the argument concerned an intuitive and pre-theoretic notion of 'act'. But in the present context we are clearly using 'act' as a theoretical term, and we should not try to settle debates by defining our technical terms to make our preferred position trivially or analytically true.

[8] Hedden (2012), amongst others, suggests that we can solve a variety of problems to do with the epistemic availability of options by taking the options in decision making to be the decisions themselves, instead of the actions which issue from them. I am sympathetic to this position in so far as it solves a number of problems involved with trying and failing to perform an act—since one generally cannot try and fail to make a decision—and have no objection to its being substituted here. However, it will not help to solve the Predetermination Problem. The Predetermination Problem arises because the world might be such that we are predetermined to choose or perform some option; making the options in decision making the decisions themselves will not help, because we might be predetermined to make a particular decision.

Moving on to states: decision makers are concerned with the relevant ways the world might be. We will understand states as propositions which express that one among some set of these ways the world might be obtains. These propositions can be more or less fine-grained. The most fine-grained states pick out fully specific ways the world might be or, in other words, individual possible worlds. Less fine-grained states specify some, but not all, of the facts about how the world might be, and pick out sets of possible worlds. The relevant sense of 'might' is *possible according to the decision maker's epistemic state*.

The states in any particular decision problem must form a partition: they must be exhaustive—so that every epistemically possible world is contained in at least one state—and exclusive—so that no epistemically possible world is contained in more than one state. Since we are assuming partition invariance, for now, we will allow any such partition of the possible worlds. When we discuss partition-dependent formulations in §6 we will introduce restrictions on which particular propositions can define a state, but will not alter our basic understanding of what states are. I take it that this way of understanding states is shared by all the major formulations of CDT.

For finite partitions of the ways the world might be into states we can identify the relevant sense of 'might' with the decision maker assigning non-zero credence to a state. There is, however, a difficulty with doing this for infinite partitions: in some infinite partitions each individual state is a set of measure zero and must, therefore, have zero probability. How to characterize the relevant sense of epistemic possibility when sets of measure zero are involved is a difficult problem I cannot address here, but it will not make any difference for our purposes. Nothing in the Predetermination Problem or its solutions depends on infinities in any way—we could explain the problem, its solutions, and their costs with a finite space of possible worlds.

To understand the Predetermination Problem two things remain to be explained: First, the constraints on causal probabilities required to ensure two-boxing in Newcomb's problem and why these constraints commit us to Non-Zero Probabilities. Second, why the most obvious account of utilities commits us to Undefined Utilities.

2.1 Non-Zero Probabilities

To begin understanding Non-Zero Probabilities it is helpful to remember the reason that CDT was introduced in the first place—namely, to get the right results in Newcomb's problem, which goes like this[9]:

> **Newcomb's Problem:** You are on a game show and there are two boxes in front of you. One of them is transparent and contains $1,000. The other is

Hedden (2012, 354) suggests that if an agent is predetermined to make a particular decision then subjective oughts of the kind that CDT is interested in do not apply to them. He does not suggest what an agent should do when they are unsure if this is their situation; this is just what is at stake in the Predetermination Problem. We can, charitably, assume that he is suggesting something like the FFCD solution. But that is an additional commitment over and above the claim that the options in decision making are decisions themselves' a solution to the Predetermination Problem is still needed on such an account.

[9] See (Nozick, 1969) for the canonical statement of Newcomb's problem. My version differs only in inconsequential details.

> opaque, but you are told that it contains either nothing or $1,000,000. You must choose whether to take just the opaque box (one-boxing), or the opaque and transparent boxes (two-boxing). The opaque box contains $1,000,000 if and only if the show's producers have predicted that you will take only the opaque box. The prediction was made 10 minutes ago during the ad break, and the money has already been placed in the box or returned to the bank. You believe that the producers of the show are extremely likely to be right. They take detailed questionnaires before the show, scan your brain, run psychological tests, and observe your behaviour during the show to determine whether you will take one or both boxes—and they have never gotten it wrong before.

Should you take one or two boxes? The causal decision theorist tells you to take both, and they reason like this:

> The producers have already made their prediction and either put the money in the box or returned it to the bank—nothing you do now can change the past or make money magically appear in the box (violating the laws of nature). Now, if there is already $1,000,000 in the opaque box then taking both boxes will get you $1,000 more than taking just one box. And if there is nothing in the opaque box then taking both boxes will again get you $1,000 more than taking just one box. So, either way, you should take both boxes, because doing so is guaranteed to get you an extra $1,000 and does not affect whether you will get $1,000,000 as well—even though you think that taking both boxes is good *evidence* that you will walk away with only $1,000.

This reasoning is, I will assume, sound—you should take both boxes.[10]

Now, we can extract from this argument three conditions that any account of causal probability must satisfy, if it is to justify the claim that we should take both boxes in Newcomb's problem. These conditions are not sufficient to define causal probabilities—much more must be said to do that—but they are *necessary* conditions for any account of causal probabilities that is suitable for CDT. Unfortunately, as we shall see, any account of causal probabilities that satisfies these three conditions will also commit us to Non-Zero Probabilities. The three conditions are:

1. $P(S_i||A) = P(S_i)$ whenever S_i is *not causally influenced by* A. That is, the causal probability of S_i if A is performed, when S_i is not causally influenced by A, is equal to the unconditional credence of S_i—or, in other words, $P(S_i||A)$ tracks causal influence. (This entails the next condition.)
2. $P(S_i||A)$ is not always equal to $P(S_i|A)$—that is, the causal probability of S_i obtaining if I do A is not just the conditional (evidential) probability of S_i given A.
3. Neither the past nor the laws of nature are causally influenced by our choices.[11]

[10] Of course, opponents of CDT will not agree, but they can still gain something from the rest of this paper: understanding the Predetermination Problem can help us understand CDT even if we disagree with it by helping us to understand costs the CDT must take on which do not arise directly from Newcomb's problem. I will also assume throughout that we can use monetary values as a directly proportional substitute for utilities. This is obviously unrealistic, but it will not harm the case to be made below.

[11] It would be more standard to use the phrase "is causally independent of" here instead of "are not causally influenced by". However, in the present context it is very important to bear in mind the asym-

Put together 1 and 3 entail that the causal probability that the past and the laws of nature will be some particular way if we perform an act ϕ is just the unconditional probability of the past and the laws of nature being that way. That is, for all ϕ and all propositions P&L about the past and/or the laws of nature, $P(\text{P\&L}\|\phi) = P(\text{P\&L})$.

Since the prediction in Newcomb's problem is just a fact about the past this entails that the causal probability of the prediction being either way is independent of our choice and, hence, that the dominance reasoning which justifies two-boxing is valid. These three conditions validate the reasoning in Newcomb's problem. Causal decision theorists cannot give them up without giving up their *raison d'etre*: the argument for two-boxing in Newcomb's problem.

To understand what goes wrong with CDT when predetermination comes into the picture we will continue with our story about Newcomb's problem.

> You are deliberating about whether to take one or two boxes and the following thought comes to mind while: What if my decision, not just whether the money is in the box, is already determined? Perhaps the conditions of the universe 10 minutes ago, during the ad break, together with the laws of nature, entail that I will take only one box. Or perhaps they entail that I will take both boxes. What should I do? I can't change what the conditions of the universe 10 minutes ago were, nor the laws of nature, nor what they together entail—not any more than I can change the fact that the prediction has been made and that this, together with the laws of nature, either entails that there is $1,000,000 in the opaque box or entails that there is not.

This reasoning certainly looks cogent: in Newcomb's problem causal decision theorists rely on the fact that the past and the laws of nature are not under our (present) control to argue that whether or not there is money in the opaque box is not under our (present) control. But both the facts, if there are any, which determine which decision I will make, and the facts that determine what the prediction was, are just facts about the past and the laws of nature—indeed, they might be *the very same facts*. From the perspective of CDT all facts about the past and the laws of nature should be on a par. Hence, if the prediction in Newcomb's problem is not causally influenced by our acts then neither is what, if anything, we are predetermined to choose. To say that what, if anything, we are predetermined to choose depends on our choice requires denying one of the three conditions on causal probabilities above, and if we do that we will have to give up the argument for two-boxing that motivates endorsing CDT instead of EDT in the first place.

This establishes the first of our two key claims: *Non-Zero Probabilities*. The very same reasoning that justifies two-boxing in Newcomb's problem ensures that the proposition that the past and the laws of nature entail I will not perform some act ϕ must not be causally influenced by my choice over whether to perform ϕ. Together with the two other conditions on causal probabilities this entails that the credence I should use in my decision-making for the act-state pair consisting of an act and a

metrical nature of the relationship we are interested in: it is possible for which state obtains to causally influence one's choice while at the same time one's choice has no causal influence on which state obtains. The phrase "causally independent" tends to suggest a symmetry we should avoid suggesting here. I thank an anonymous reviewer for pointing out this possible confusion.

state in which I am predetermined not to perform that act is just my unconditional credence that I am predetermined not to perform that act. And, so long as we take predetermination to be generally possible, this credence will not be zero.[12] Hence, if predetermination is a possibility then there are inconsistent act-state pairs in our decision problems—specifically the act-state pairs consisting of an act and the states in which we are predetermined not to perform it—which have non-zero causal probability.

A brief note: we should be careful about the nature of causal probabilities here. The probability of an act occurring and a state obtaining must be zero when that act and state are inconsistent—-the axioms of probability require that every contradiction has zero probability. However unlike (evidential) conditional probabilities, causal probabilities are not necessarily defined in terms of the probability of any conjunction. Causal probabilities are a theoretical construction designed to do special work in causal decision theory. We are, therefore, free to define them in a way that allows the causal probability of some inconsistent act-state pairs to be non-zero, so long as we meet the three conditions above which get us the right results in Newcomb's problem. Below I will, for ease of expression, refer to the 'causal probability *of* an act-state pair', but the reader should bear in mind that this is not the probability of the conjunction of that act and that state.

Now that we see that any account of causal probabilities which justifies two-boxing in Newcomb's problem also commits us to Non-Zero Probabilities, we will keep going with our little story: As a good causal decision theorist you know you need to construct a partition of the epistemically possible worlds. Since you are worrying about predetermination you want to take that into account, so you use the following six states:[13]

S_1: I am now determined to one-box, and the prediction is one-boxing.
S_2: I am now determined to one-box, and the prediction is two-boxing.
S_3: I am now determined to two-box, and the prediction is one-boxing.
S_4: I am now determined to two-box, and the prediction is two-boxing.
S_5: My choice is not yet determined, and the prediction is one-boxing.
S_6: My choice is not yet determined, and the prediction is two-boxing.

Since each of these states can be specified completely by facts about the past and the laws of nature, you know that none are causally influenced by your choice. Each of these states is just the conjunction of a prediction, which is not causally influenced by your choice—if the argument for two-boxing in Newcomb's problem is right—and a fact about what, if anything, you are predetermined to choose—which, as I have just argued, is also not causally influenced by your choice. The conjunction of any two things which are not causally influenced by your choice cannot be causally influenced by it. You know, then, that you should use your unconditional credence of

[12] Specific agents in specific situations might be sure their choice is not predetermined. But CDT is supposed to apply to all rational agents. So unless it is *irrational* to believe your choice might be predetermined the Predetermination Problem will arise.

[13] It is sometimes suggested that something is wrong with this partition of states. However, since, for the moment, we are assuming partition invariance it is necessarily the case that if CDT gives an undefined expected utility with this partition it will do so with any other.

each state, $P(S_1)$, $P(S_2)$, etc., as the causal probabilities in this decision problem. To complete our understanding of the Predetermination Problem I now turn to examining the utilities involved.

2.2 Undefined Utilities

Let's keep going with the story. At this point you need to find the utilities for your act-state pairs, so you think to yourself:

> If I one-box and the prediction is one-boxing I will get \$1,000,000 (whether or not that choice was predetermined). If I one-box and the prediction is two-boxing I will get \$0 (whether or not that choice was predetermined). If I two-box and the prediction is one-boxing I will get \$1,001,000 (whether or not that choice was predetermined). And If I two-box and the prediction is two-boxing I will get \$1,000 (whether or not that choice was predetermined). So, I can fill in some of the utilities in the decision matrix, as in Table 1.

	S_1	S_2	S_3	S_4	S_5	S_6
One-Box	\$1,000,000	\$0			\$1,000,000	\$0
Two-Box			\$1,001,000	\$1,000	\$1,001,000	\$1,000

Table 1 Decision Matrix for Newcomb's problem.

So far, so easy. All this is justified by the intuitive thought that the utility of an act-state pair is just the utility of their conjunction—to work out how good it would be if I did A in state S_i I just need to work out how good it would be if I did A and state S_i obtained. Most versions of CDT appeal to Jeffrey's (1983) account of desirability as their account of utilities. I have no objections to this way of construing utilities; it will justify the reasoning so far. Now you are, however, faced with a problem:

> What about the blank squares? What will happen if I am already determined to take two boxes, but I take only one box? And what if the past and the laws of nature entail that I will take one box, but I take both boxes? Hmm.

Your puzzlement is well founded. You need to know what the outcome of two-boxing will be if *true facts entail* that you will one-box (and vice versa). Anything that is entailed by true facts must be true. So this is equivalent to knowing what will happen if you both do and do not take two boxes. However, there is no fact of the matter about this, and, hence, no sensible utility we can associate with the inconsistent act-state pairs of an act and the state where you are predetermined not to perform that act. On the one hand we might think that there is no fact of the matter about what will happen if a contradiction is true, because no contradiction could in fact be true. Or, if we are more classically minded, we might think that *everything* will be true if a contradiction is true. Either way there is no sensible way to say how good it will be if a contradiction actually is true.[14] Undefined Utilities is justified by the fact that it makes no sense to ask how good it will be if a contradiction is true

[14] If you are *paraconsistently* minded you might object at this point that we can make sense of the truth of contradictions in the actual world. Unfortunately I do not have the space here to discuss using a

Now, unfortunately, there is no way for us to satisfactorily complete our story and find out which act you, as a causal decision theorist, should choose. We cannot fill in the blank squares in the decision matrix in Table 1. But the corresponding act-state pairs have non-zero causal probabilities. So your expected utility for one boxing will be (and *mutatis mutandis* for two boxing):

$$\begin{aligned} EU(\text{One-Box}) &= P(S_1) \times \$1,000,000 + P(S_2) \times \$0 + P(S_3) \times \text{undefined} \\ &\quad + P(S_4) \times \text{undefined} + P(S_5) \times \$1,000,000 + P(S_6) \times \$0 \\ &= \text{undefined} \end{aligned}$$

The suspicion that there is some tension between rational decision-making and believing that your choices might be predetermined is borne out: Any account of causal probabilities which justifies two-boxing in Newcomb's problem commits us to Non-Zero Probabilities. And the most obvious understanding of the utility of an act-state pair commits us to Undefined Utilities. Put together these two claims entail that whenever predetermination is a possibility there are act-state pairs with non-zero probability but undefined utility in our expected utility calculations. But the product of an undefined term and a non-zero term is undefined, and the sum of any number of defined terms with an undefined term is undefined. Hence, CDT falls silent when we consider the possibility that the outcome of our current choice is predetermined. This is causal decision theory's Predetermination Problem.

Of course, there would be no problem if the probability associated with the inconsistent act-state pairs were zero. For the purposes of expected utility calculations, we must assume that an undefined term multiplied by zero is equal to zero (as (Lewis, 1981, 14) notes). Hence, if Non-Zero Probabilities were false and the inconsistent act-state pairs all had zero probability the undefined utilities would be cancelled out. We need this assumption to deal with a variety of cases involving inconsistent act-state pairs that do not raise the same issues as the Predetermination Problem. Most notably cases in which an act will *cause* the world not to be in some state, so that the act and state are inconsistent but their causal probability is zero. We can justify the assumed rule by appeal to supervaluation: if an undefined term is multiplied by zero then we will get the same result no matter what value we substitute for the undefined term. We can now also see why evidential decision theory does not face the Predetermination Problem: the evidential probability that I am predetermined to do A, given that I do not do A, is zero, and hence all of the problematic act-state pairs have zero probability and do not contribute to the expected utility calculations.

The Predetermination Problem arises from the conjunction of Undefined Utilities and Non-Zero Probabilities.[15] If we do not want to give up CDT we must deny one of

paraconsistent logic to solve the Predetermination Problem. I take it, however, that this possibility will be attractive to very few—giving up classical logic (or even just the law of non-contradiction) would be a very high price to pay to solve the Predetermination Problem.

15 You might be suspicious that the Predetermination Problem is really about simply considering what we will do in the future; that it would arise with any state which *entails* that you will perform some particular act, not just those in which you are *predetermined* to perform some particular act. For example, in Newcomb's problem, you might think that the state "I will two-box and the prediction is two-boxing" would be a problem—because it is inconsistent with one-boxing—even though it has nothing specifically to do with predetermination. However, when the fact that I will take two boxes is a fact about the future

these claims. There are two ways to deny Non-Zero Probabilities: We could require that every rational decision maker assign zero credence to their present decision being predetermined. Or, we could give up either the idea that causal probabilities should track only causal influence, or that the past and the laws of nature cannot be causally influenced by our choices and suggest a new way of defining causal probabilities so that the inconsistent act-state pairs always have zero probability. To deny Undefined Utilities we must give some account of how a contradictory act-state pair can be assigned a sensible utility; the obvious way to do so is to appeal to subjunctive conditionals to find an answer to the question "supposing I am predetermined not to do A, what would happen if I were to do A?" and use this to define utilities. Below I will consider all three possibilities and argue that they all face substantial costs—setting up a trilemma for causal decision theorists.[16]

3 The Freedom For Current Decision Solution

The first way to solve the Predetermination Problem is to insist that, in order to rationally deliberate, decision makers must assume that predetermination is impossible. Indeed, there is a long tradition of such thinking, often traced back to Kant's claim that we must *act under the idea of freedom* (Nelkin, 2011, 117). Specifically, we can solve the Predetermination Problem by requiring that every rational decision maker give zero credence to the outcome of their *current* decision being predetermined. That is, all rational decision makers have $P(S_i) = 0$ whenever the past and laws of nature of S_i entail that they will do ϕ or entail that they will not do ϕ, for any option ϕ under consideration in their current decision.[17] (That my earlier decisions might have been, or my later decisions will be, determined ahead of time is no challenge for thinking about what I ought to do now.) The inconsistent act-state pairs would then all have zero causal probability, solving the Predetermination Problem. This is the Freedom For Current Decision (FFCD) solution.

The first problem for the FFCD solution is that it seems to endorse ignoring possibilities relevant to our decisions. The FFCD solution excludes all act-state pairs involving predetermination—because it entirely excludes a state whenever it entails that *any* action will or will not be performed—even though only some such act-state

not entailed by the past and the laws of nature, it is sensible to say that if I take only one box it will *cause* me not to be in a world where I take both boxes; the causal probability of taking two boxes if I take one box is zero when my taking two boxes is not entailed by anything, including the past and the laws of nature, outside my causal control. Only when the fact that I will two-box is entailed by the past and the laws of nature, or something else outside my causal control, is the causal probability that I will two-box if I one-box non-zero. The Predetermination Problem is specifically to do with predetermination, not the mere existence of facts about what we will do.

[16] All three solutions can be understood in either a revisionary mood, suggesting a change to CDT that will solve the problem, or in a descriptive mood, as pointing to something that CDT is already, more or less implicitly, committed to; I will remain neutral on this distinction here—though the reader should bear in mind that this might make a difference to the plausibility of the solutions.

[17] We could, equivalently, exclude such S_i from the probability space over which causal probabilities are defined. This might be preferable on the grounds that it allows us to distinguish between propositions with zero probability because they pick out sets of measure zero and propositions which the decision maker takes to be impossible.

pairs cause a problem. This seems to involve ignoring relevant information. Supporters of the FFCD solution owe us an account of why it is that ignoring this information won't lead us into trouble, or why the trouble it leads us into is less trouble than we would get into with either of the other solutions.

The second problem for the FFCD solution is that a rational agent can, and likely must, have a non-zero credence that the outcome of their current decision is predetermined. This possibility is compatible with any set of evidence available to agents like us.[18] Indeed it is compatible with our actual evidence, which justifies a non-zero credence in deterministic theories like Bohmian mechanics. We cannot, therefore, rule out this possibility without ignoring our evidence.

Perhaps one might dispute this and claim that our evidence in fact supports a zero credence in our choices being predetermined. Taken as a claim about our actual empirical evidence this claim is extremely implausible. But one might (continuing the Kantian mood) argue that we have (synthetic) a priori reason to believe our choices are not predetermined and therefore to reject as mistaken or misleading any evidence to the contrary.[19] Not being a Kant scholar the details of such an account are somewhat opaque to me. However, few causal decision theorists are likely to endorse such a claim—it would require denying that facts about the past and the laws of nature are epistemically contingent, a claim that is central to a naturalistic understanding of the world.

How then can we claim that predetermination is not a possibility in decision-making? There are two basic strategies: First, we could deny that credences—qua both degrees of belief *and* decision-making probabilities—must always be proportioned to our evidence. Second, we could deny that each agent has a single credence function that plays both the degree of belief and decision-making probability role—we could become credal pluralists. I will examine these options in turn.

Rational degrees of belief must, intuitively, have something to do with how much our evidence supports various propositions. At the very least it seems that rational degrees of belief should be proportioned to our evidence to some degree, even if our evidence is not all that determines them: when we have evidence in favor of a proposition our degree of belief in it should not be zero. Without such a minimal responsiveness to evidence it is hard to see how credences could count as degrees of belief. Denying this entails that an agent could truthfully, and without any incoherence or irrationality, assert that "the evidence is that it is possible that my choices are predetermined, but I do not believe it, in fact I believe it is impossible that my choices are predetermined". (The analogy with Moore's paradox is intentional.) The cost of endorsing this kind of statement is to make it unclear what degrees of belief really are.

Turning now to our second possibility: we might, instead, endorse credal pluralism and distinguish between credences *qua* degrees of belief and credences *qua* probabilities for decision-making. It will then be far more plausible to suggest that for the purposes of CDT we must assume that predetermination is impossible, because doing so will be compatible with having any degree of belief that our choices

[18] Perhaps God can have evidence that is incompatible with predetermination, but no mere mortal could.

[19] I thank an anonymous reviewer for pointing out this possibility.

are predetermined. Our degrees of belief can remain responsive to our evidence, while the probabilities we use in decision-making sometimes depart from it on pragmatic grounds. The primary challenge facing this suggestion is to provide a detailed account of the two kinds of credence and their interaction if they are not always equal. This is especially difficult because separating degrees of belief and the probabilities used in decision-making forces us to give up on many elements of the standard (Bayesian) picture of the relationship between theoretical and practical reasoning. The success of this picture is therefore an objection to making such a distinction.

For example, separating these credences undermines the standard picture of the representation theorems which are often taken to be central to decision theory. While representation theorems may still be sound, they will not be able to tell us (directly) about an agent's epistemic state—they will only be able to tell us about the probabilities that the agent uses in their decision-making. We might use these probabilities to infer things about an agent's epistemic state, but we will no longer have a tight link between an agent's preferences (or choices) and their degrees of belief. This will be particularly problematic if our representation theorems must rely on external constraints on the probabilities involved derived from epistemic considerations, as, for example, Joyce (1999) does when he requires that an agent have both a preference ranking and a relative probability ranking of all propositions in order to secure the result that credences must be probabilistic. In such cases it is plausible that constraints on preferences will transfer to decision-making probabilities, while epistemic constraints will transfer to degrees of belief—but then it may not be possible to show that decision-making probabilities need be probabilities at all. Representation theorems will still be useful in defining and understanding an agent's decision-making probabilities. But any further link to their degrees of belief cannot be taken for granted and will require justification *external* to the representation theorem. Of course such justification may well be possible, but I take it that the departure from the standard picture on which representation theorems directly inform us about degrees of belief is a cost. Similar considerations will complicate arguments which move from practical constraints on decision-making credences—avoid dutch books for example—to constraints on degrees of belief.

Note that whether we choose to divorce degrees of belief from evidence or endorse credal pluralism, we are saying much more than that it is acceptable to treat as false some proposition which our evidence suggests is very unlikely for pragmatic reasons. One might be sympathetic to the idea that, at least for non-ideal agents like us, it is acceptable to treat as false in our decision-making some claims with a low enough probability, because doing so will reduce cognitive burden (and thus the probability of mistakes) without introducing a high risk of going wrong. For example, it is plausibly rational to assume, when planning a holiday, that all the world's cats will not suddenly develop super intelligence and rise up to enslave humanity—the probability of such an occurrence is extremely low and entertaining it is very likely a waste of time for a cognitively limited agent like me. Indeed it might even be reasonable to believe that this will not occur in epistemic contexts where pragmatics are less of a concern. But neither suggestion here is like this. Both require assuming our choice is not predetermined, even if our evidence justifies a credence close to certainty that it is—both require thinking it can be rational assigning zero credence to what we take to be *far and away the most likely scenario*. And this only makes worse the worry that ruling out predetermination involves ignoring relevant information.

The FFCD solution violates the intuition that our decision-making should take into account all relevant possibilities. An explanation of why this intuition is mistaken or why predetermination is not relevant to our decision making is needed. The FFCD solution also forces us to give up either a plausible understanding of the relationship between credences and evidence, or to endorse credal pluralism and give up the standard picture of the relationship between theoretical and practical rationality. This is a high price to pay to solve the Predetermination Problem.

4 The Predetermination Possible, Inconsistency Impossible Solution

The past and the laws of nature are not causally influenced by anything I can do now. Yet surely the causal probability that I am predetermined not to perform some act if I do perform that act is zero. There is a tension between these two claims: causal probabilities were introduced to track causal influence (so as to achieve the correct results in Newcomb's problem), but both claims cannot be true if causal probabilities track causal influence. The next solution to the Predetermination Problem sides with the second claim and defines causal probabilities so that the probability of being predetermined not to perform an act, on the assumption that act is performed, is zero—undermining Non-Zero Probabilities and solving the Predetermination Problem. Call this solution the Predetermination Possible, Inconsistency Impossible (PPII) solution.

The heart of the PPII solution is captured by replacing either the first or third condition on causal probabilities from §2.1 with the following two conditions (call these the PPII conditions). The Inconsistency Impossible condition solves the Predetermination Problem, while the Predetermination Possible condition ensures that the PPII solution is distinct from the FFCD solution.

Predetermination Possible Condition: It is *not always* the case that $P(S_i||A) = 0$ when S_i entails A—that is, the causal probability of being predetermined to do A, if you do A, is, at least sometimes, non-zero.

Inconsistency Impossible Condition: $P(S_i||A) = 0$ when S_i entails $\neg A$—that is, the causal probability of being predetermined not to do A, if you do A, is zero. Meeting this condition solves the Predetermination Problem.

These two conditions leave a lot of room for how we might define causal probabilities, and hence there are many variations of the PPII solution. One relevant question is whether we give up the first or third condition in §2.1, and, hence, whether the PPII solution is suggesting that causal probabilities track more than just causal influence, or that our acts can causally influence the past and/or the laws of nature. But, as we shall see, all versions of the PPII solution face the same problem: giving up the reasoning that justifies two-boxing, no matter how high your credence that your choice is predetermined, in Newcomb's problem. Call this *unconditional two-boxing*.

To show this I will show that any definition of causal probabilities that meets the PPII conditions and the remaining conditions from §2.1 will sometimes endorse one-boxing in a kind of case I call *Bonus Money Newcomb Problems* (BMNPs). It might seem needlessly complicated to introduce these new BMNP cases instead of directly arguing that the PPII solution must give up unconditional two-boxing in Newcomb's problem. The reason for doing so is that there are ways of defining causal probabilities as the PPII solution suggests which will ensure that it endorses unconditional two-boxing in Newcomb's problem.[20] However, there is no way of defining causal probabilities to satisfy the PPII solution that will endorse unconditional two-boxing in every BMNP. And the reasoning that justifies unconditional two-boxing in Newcomb's problem also justifies unconditional two-boxing in BMNPs—any way of distinguishing the *arguments* that justify unconditional two-boxing in the two cases is ad hoc. And it is the argument, not merely the result, that matters to decision theory—decision theories should produce their results from coherent reasoning patterns, not by mere accident. Hence, giving up the argument in favor of unconditional two-boxing in BMNPs is as bad as giving it up in Newcomb's problem. Now, consider the following case:

> **$1,000,000 Bonus Newcomb Problem:** You are on a game show and there are two boxes in front of you. One of them is transparent and contains $1,000. The other is opaque, but you are told that it contains either nothing, $1,000,000, or $1,000,000 plus a bonus $1,000,000. You must choose whether to take just the opaque box, or the opaque and transparent boxes. The opaque box contains $1,000,000 if the show's producers have predicted that you will take only the opaque box (regardless of whether your choice is predetermined). The prediction was made 10 minutes ago during the ad break and the money has already been placed in the box or returned to the bank. You believe that the producers of the show are extremely likely to be right. They take detailed questionnaires before the show, scan your brain, run psychological tests, and observe your behavior during the show to determine whether you will take one or both boxes—and they have never gotten it wrong before. The opaque box *also* contains the bonus $1,000,000 if and only if a separate team of physicists have established that the past and the laws of nature already determine that you will choose to take only the opaque box.

The only difference between this problem and Newcomb's problem the possibility of the bonus. Now, should you take one or two boxes? Causal decision theorists must *prima facie* endorse taking both boxes. Nothing you can do now can change whether or not the bonus money is in the box, no more than it can change the prediction and, hence, whether the first $1,000,000 is in the opaque box. How much money is in the opaque box is not causally influenced by anything you can do now.

[20] For example, we might suggest that causal probabilities be defined so that the following equalities hold in Newcomb's problem, using the states from Table 1: $P(S_1||\text{One-Box}) = P(S_1) + 0.5 \times P(S_3)$, $P(S_2||\text{One-Box}) = P(S_2) + 0.5 \times P(S_4)$, $P(S_3||\text{One-Box}) = 0$, $P(S_4||\text{One-Box}) = 0$, $P(S_5||\text{One-Box}) = P(S_5)+0.5\times P(S_3)$, and $P(S_6||\text{One-Box}) = P(S_6)+0.5\times P(S_4)$. That is, we start with the assumption that the states are not causally influenced by our choice—so that the causal probabilities are just the unconditional probabilities of the states. Then we set the causal probabilities of the inconsistent act-state pairs to zero. Finally we ensure normalization by splitting the unconditional probability of the inconsistent act-state pairs equally among the consistent act-state pairs which agree about the prediction (S_1, S_3, and S_5 all agree that the prediction is one-boxing; S_2, S_4, and S_6, that it is two-boxing). And *mutatis mutandis* for two-boxing. Using the causal probabilities so defined will always lead to two-boxing in Newcomb's problem, no matter how high your credence that your choice is predetermined.

As such, you can ignore the possibility of the bonus—nothing you can do now will improve your chances of getting it, so it should make no difference to your choice. It is as though someone has first given you a lottery ticket that pays out $1,000,000 if you win, and then you are faced with Newcomb's problem, and the conditions under which you win the lottery—which have already been settled—just accidentally happen to be connected to causal precursors of your choice. The fact that you are holding the lottery ticket, which has either already won or lost, should not influence the decision you now face. Of course, if you get the bonus, i.e. win this lottery, that will be a good thing. But similarly it will be a good thing, in Newcomb's problem, if it turns out that the prediction is one-boxing. Causal decision theorists insist, however, that this should not influence your decision-making because you cannot (now) change whether or not the prediction is one-boxing. If we ignore the possibility of the bonus being in the opaque box, then the $1,000,000 Bonus Newcomb Problem is no different to Newcomb's problem. Hence, you should take both boxes.

Causal decision theorists cannot maintain, in any non-ad hoc way, that you should sometimes one-box in the $1,000,000 Bonus Newcomb Problem but should unconditionally two-box in Newcomb's problem. The existence of the bonus does not change the situation in any way relevant to the argument in Newcomb's problem. Endorsing one-boxing in the $1,000,000 Bonus Newcomb Problem will undermine the argument which causal decision theorists use to motivate unconditional two-boxing in Newcomb's problem. Further, the amount of the bonus involved in the $1,000,000 Bonus Newcomb Problem is arbitrary: nothing in the above argument depends on the actual dollar value. Hence, if causal decision theorists endorse unconditional two-boxing in the $1,000,000 Bonus Newcomb Problem they should do so no matter what the value of the bonus is—even if it is zero. If CDT endorses one-boxing in *any* BMNP it will have undermined the reasoning that leads us to unconditional two-boxing in Newcomb's Problem.

The proof that there is always a BMNP for which any account of causal probabilities meeting the PPII conditions endorses one-boxing is quite simple. First, define the states S_1 through S_6 as for Newcomb's problem above. Then the decision matrix for an arbitrary BMNP, with bonus $\$a$ is as in Table 2.

	S_1	S_2	S_3	S_4	S_5	S_6
One-Box	$1,000,000 + $a	$a			$1,000,000	$0
Two-Box			$1,001,000	$1,000	$1,001,000	$1,000

Table 2 Decision Matrix for an arbitrary Bonus Money Newcomb Problem.

To meet the Inconsistency Impossible Condition—and avoid the Predetermination Problem—we must have $P(S_3||\text{One-Box}) = 0$, $P(S_4||\text{One-Box}) = 0$, $P(S_1||\text{Two-Box}) = 0$, and $P(S_2||\text{Two-Box}) = 0$. Substituting these into the expected utility formula we find:

$$\begin{aligned} EU(\text{One-Box}) = {} & P(S_1 \| \text{One-Box}) \times (\$1,000,000 + \$a) + P(S_2 \| \text{One-Box}) \times \$a \\ & + P(S_5 \| \text{One-Box}) \times \$1,000,000 \\ EU(\text{Two-Box}) = {} & (P(S_3 \| \text{Two-Box}) + P(S_5 \| \text{Two-Box})) \times \$1,001,000 \\ & + (P(S_4 \| \text{Two-Box}) + P(S_6 \| \text{Two-Box})) \times \$1,000 \end{aligned}$$

We also know that $EU(\text{Two-Box}) < \$1,001,000$, because $EU(\text{Two-Box})$ is maximized when $P(S_3 \| \text{Two-Box}) + P(S_5 \| \text{Two-Box})$ is maximized, and, since our credences must be normalized, $P(S_3 \| \text{Two-Box}) + P(S_5 \| \text{Two-Box}) \leq 1$. Together with the utility values this entails $EU(\text{Two-Box}) \leq \$1,001,000$. We can now, by simple rearrangement, find a formula for a that guarantees $EU(\text{One-Box}) > EU(\text{Two-Box})$:

$$\begin{aligned} & EU(\text{One-Box}) > EU(\text{Two-Box}) \\ & P(S_1 \| \text{One-Box}) \times (\$1,000,000 + \$a) + P(S_2 \| \text{One-Box}) \times \$a \\ & \quad + P(S_5 \| \text{One-Box}) \times (\$1,000,000) > EU(\text{Two-Box}) \\ & (P(S_1 \| \text{One-Box}) + P(S_2 \| \text{One-Box})) \times \$a + (P(S_1 \| \text{One-Box}) \\ & \quad + P(S_5 \| \text{One-Box})) \times (\$1,000,000) > EU(\text{Two-Box}) \end{aligned}$$

$$\begin{aligned} & (P(S_1 \| \text{One-Box}) + P(S_2 \| \text{One-Box})) \times \$a > EU(\text{Two-Box}) - (P(S_1 \| \text{One-Box}) \\ & \quad + P(S_5 \| \text{One-Box})) \times (\$1,000,000) \\ & \$a > \frac{EU(\text{Two-Box}) - (P(S_1 \| \text{One-Box}) + P(S_5 \| \text{One-Box})) \times (\$1,000,000)}{P(S_1 \| \text{One-Box}) + P(S_2 \| \text{One-Box})} \end{aligned}$$

When this final inequality is satisfied CDT will endorse one-boxing. But, for any definition of causal probability which satisfies the PPII conditions, there is *always* a value of a which satisfies this inequality. This is guaranteed by the fact that $EU(\text{Two-Box}) \leq \$1,001,000$, and that the probabilities must be between zero and one. This entails that the numerator of the right hand side is a real number less than $\$1,001,000$. We know, therefore, that the above inequality is always satisfied when the following, simpler, one is:

$$\$a > \frac{\$1,001,000}{P(S_1 \| \text{One-Box}) + P(S_2 \| \text{One-Box})}$$

And the Predetermination Possible condition guarantees that there will be some agents for which $P(S_1 \| \text{One-Box}) + P(S_2 \| \text{One-Box})$ is strictly greater than zero—so there is always some agent for which the right hand side of this inequality is defined and finite. We will, then, always be able to find a value for a, given a value for $P(S_1 \| \text{One-Box}) + P(S_2 \| \text{One-Box})$, which guarantees $EU(\text{One-Box}) > EU(\text{Two-Box})$.[21] The PPII solution entails that an agent with this credence function should one box in a BMNP

[21] This *might* fail if the probabilities $P(S_1 \| \text{One-Box})$ and $P(S_2 \| \text{One-Box})$ decrease with increasing *utility* of the associated act-state pairs for every rational agent. This is not a requirement that causal decision theorists should rush to place on causal probabilities. First, it would require that *all* rational agents consider increasing utility always less likely, which is an implausible rational constraint (it must be all rational agents since causal decision theorists are committed to all rational agents two-boxing, not just those with particular beliefs about the relationship between utilities and probabilities). Second, allowing such depen-

where the bonus is equal to the relevant value of a.[22] Even if a plausible definition of causal probabilities can be found which succeeds in guaranteeing unconditional two-boxing when $a = 0$ (Newcomb's problem), there will still be values of the bonus for which CDT will endorse one-boxing. And since there is no non-ad hoc distinction between the arguments in favor of unconditional two-boxing in BMNPs and in Newcomb's problem, all versions of the PPII solution undermine the reasons which justify unconditional two-boxing in Newcomb's problem.

Of course, proponents of the PPII solution might be willing to pay this cost—solving the Predetermination Problem is going to require giving up some plausible claims no matter what we do, and giving up unconditional two-boxing might turn out to be the most palatable option.

However, to make it plausible that we should give up the intuitions in favor of unconditional two-boxing, supporters of the PPII solution owe us two things: First, an account of a non-ad hoc distinction between the conditions that predetermine our actions and those that determine what the prediction is which can justify the claim that the former but not the later have causal probabilities that depend on our choices. Both are, *prima facie*, simply facts about the past and the laws of nature; they might even be the very same facts. Since the PPII solution is committed to allowing the causal probability of the former to depend on our choice—as the BMNPs show—some such distinction must be drawn if the PPII solution is not to collapse into EDT. Second, in tension with the first requirement, supporters of the PPII solution need to provide an argument that justifies two-boxing in Newcomb's problem when you are certain (enough) that your choice is *not* predetermined—otherwise we will have no reason to give up EDT in favor of CDT. And this argument cannot just be the standard argument from §2.1, since that argument justifies *unconditional* two-boxing.[23]

We cannot accept the PPII solution without accepting one-boxing in some BMNPs, and, thereby, giving up the argument for unconditional two-boxing in Newcomb's problem. To accept the PPII solution is to accept that the past and the laws of nature can be causally influenced by our choices or that more than causal influence is

dence between these probabilities and utilities causes all sorts of problems elsewhere, particularly for the Principal Principle and normalization of credences, because it allows the probability of a state to depend on something—its utility—which is not a feature of that state (the utility a state has is a relationship between that state and features of the agent in the actual world).

[22] Note that while simplifying the numerator of the inequality above to its maximum value $1,001,000 is useful in understanding what is going on, the inequality which must be satisfied for one-boxing to be endorsed is the more complex one. The relevant value of a in the more complex inequality can be equal to, or less than, zero—that is, there are many definitions of causal probability meeting the PPII solution conditions that will directly endorse one-boxing in Newcomb's problem when you are sure enough that your choice is predetermined.

[23] Unless, perhaps, supporters of the PPII solution can explain why, despite appearances, our intuitions in the standard Newcomb's problem are based on an assumption that our choice is not predetermined. Note that supporters of the FFCD solution will also claim that our intuitions in Newcomb's problem are based on the *assumption* that our choice is not predetermined. But since they take this assumption to be required for rational decision-making it seems less difficult for them to explain why our intuitions are based on it than for supporters of the PPII solution who argue that we can reason perfectly well while considering the possibility of predetermination, and that when we do so we will sometimes come to endorse one-boxing. Supporters of the FFCD solution also maintain that one should two-box no matter how high one's degree of belief that one's choices are predetermined—it is merely that no matter what this degree of belief one's decision-making probability should be zero in one's choice being predetermined.

relevant to decision-making. This is a high price to pay to solve the Predetermination Problem.

5 The Subjunctive Utilities Solution

Above I assumed that we cannot assign sensible utilities to contradictory act-state pairs—that there is no sensible answer to the question "supposing I am predetermined not to do something, what will happen if I do it". This assumption was justified by the claim that the utility of an act-state pair is just the utility of the conjunction of that act and that state. However, we can think about utilities in a different way—namely, by introducing the concept of *outcomes* as the direct object of utility assessments, and introducing a mapping from act-state pairs to outcomes. Outcomes are most naturally thought of as the propositions which express the consequences of performing an act when the world is in a particular state. On this picture the Predetermination Problem occurs when we assume, naturally, that the mapping from act-state pairs to outcomes is such that *both the act and the state must be entailed by the outcome*.[24] Or, in other words, that this mapping respects the answers to the indicative question "supposing S_i, what *will* happen if I *do* A?" for each act state pair. There is no answer to this question when the act and state in question are inconsistent, and so no outcome we can associate with the act-state pair, justifying Undefined Utilities.

But it is not necessarily the case that utilities must be understood in this indicative way. We can instead think of outcomes as answers to the *subjunctive* question "supposing S_i, what *would* happen if I were to do A".[25] There may be answers to such a question even when S_i and A are inconsistent. Our last solution to the Predetermination Problem appeals to this distinction between indicative and subjunctive ways of thinking about outcomes and suggests that the appropriate mapping from act-state pairs to outcomes tracks the subjunctive understanding. This can ensure that even inconsistent act-state pairs are associated with consistent outcomes, to which sensible utilities can be assigned. This solves the Predetermination Problem by undermining Undefined Utilities, without requiring any changes to the account of causal probabilities from §2.1. Call this the Subjunctive Utilities (SubU) solution.

There are, again, many variations on this solution; each proposing a different way of mapping inconsistent act-state pairs to consistent outcomes. The most obvious will involve direct appeal to subjunctive conditionals, but this need not be the case. What matters is that they assign consistent outcomes to the inconsistent act-state pairs in our decisions—which entails giving up an indicative understanding of what the outcomes of an act-state pair are. They will all (with one exception) face the same cost: they force our decision-making to take into account epistemically impossible outcomes, and this violates the intuition that decision-making is about working out what is the

[24] It is natural to think of outcomes as giant conjunctions, in which case this will be the case when the act and state are each conjuncts.

[25] I am here using the terms 'subjunctive' and 'indicative' rather loosely. All I take the former to imply is that the relevant answers to this question need not hold fixed what the decision maker takes to be the facts. In particular, it may be that what would happen if I did A on the supposition that S_i requires a violation of that very supposition. Whereas answers to indicative questions must hold any supposition fixed.

best thing to do in the actual world. Before showing this I will briefly discuss the exceptional case.

The simplest version of the SubU solution is the version which entails a uniform utility value for all inconsistent act-state pairs. The obvious choice is zero utility, since one might think that an inconsistent outcome is equivalent to nothing at all happening, and presumably nothing at all happening is neither bad nor good. In any case, since utilities only make a difference to CDT's recommendation up to linear transformation we can transform any uniform value into zero. However, assigning a uniform value to inconsistent act-state pairs in this way forces us to give up two-boxing, no matter how certain we are that the prediction is correct, in Newcomb's problem—as the PPII solution did. To confirm this the reader can easily calculate the expected value of one and two-boxing in Newcomb's problem, using the states from Table 1, assuming that the inconsistent act-state pairs have zero utility, and that $P(S_1) > P(S_3)$.[26] Whatever motivational or intuitive differences there might be between this solution and the PPII solution both must give up unconditional two-boxing, and we have already seen that that is a high cost to pay to solve the Predetermination Problem. As such I will from here on assume that the SubU solution assigns non-uniform utility values to inconsistent act-state pairs.

Here I will focus on one particular suggestion for doing so by appeal to subjunctive conditionals. But the reader should bear in mind that any way of assigning non-uniform utilities to inconsistent act-state pairs faces the same problem because it will allow for construction of counterexamples with the same structure as the one to be introduced below. The problem that faces all such suggestions is that they weigh epistemically impossible outcomes in favor of, or against, our choices and this is incompatible with the purpose of decision-making—to help us make the best decision for how the actual world is.

The most obvious suggestion for a way to use the subjunctive question "supposing S_i, what would happen if I were to do A" to undermine Undefined Utilities is to simply take the conjunction of all true answers to this question as the outcome of the act-state pair S_i and A. Formally, we can understand the idea in two steps. First, instead of defining utilities directly for act-state pairs as $U(A \wedge S_i)$, we instead use $U(\mathbb{O}[A, S_i])$ where $\mathbb{O}[A, S_i]$ is the proposition which expresses all the consequences of performing A in S_i. The Predetermination Problem arises because it is natural to assume that the consequences of performing an act A in S_i must include the facts that A is performed and that S_i obtains (as an indicative understanding of the relevant question would imply). If so, when A and S_i are inconsistent—as in the Predetermination Problem—their outcome will be contradictory and cannot be assigned any sensible utility. But thinking in subjunctive terms we can instead define:

$$\mathbb{O}[A, S_i] = \mathbb{C}_{A \wedge S_i} = \bigwedge c \text{ for all } c \text{ such that } A \mathrel{\Box\!\!\rightarrow} c \text{ is true at all worlds in } S_i$$

That is to say, the outcome of an act-state pair is just the conjunction of all the consequents of the subjunctive conditionals '$A \mathrel{\Box\!\!\rightarrow} c$' which are true at all worlds in

[26] Indeed, this suggestion is mathematically equivalent to a version of the PPII solution. Namely, the one where we assign the inconsistent act-sate pairs zero causal probability and then simply normalize the other causal probabilities in proportion to the unconditional probabilities of the states involved.

S_i. Supplied with an appropriate account of the subjunctive conditional '$\Box\!\!\rightarrow$'—for example, the popular account in terms of similarity relations in Lewis (1973)—there will be a consistent set of c for every act-state pair in the Predetermination Problem. We can then simply find the utility of $\bigwedge c$ and Undefined Utilities will be false.

The problem with the SubU solution is that it asks us to take into account outcomes which we are certain will not occur. When I am deciding whether to do something—when I am making a practical decision—I am concerned to find out which option is best given how the actual world is, not if the world were some way it is not. Of course I do not have direct access to the facts about how the world actually is, so I must use my best guess about how the world is (and this is where the need for expected utility theory comes in to the picture). That is, I must consider the various ways the actual world might be that I have not ruled out, and what the consequences of my actions would be for each of those ways. But my best guess about how the world is can rule out possibilities, and in so far as it does so, those possibilities are irrelevant. The SubU solution denies this because it accepts that consequences can be relevant, even though my best guess about how the world is entails they will not occur.

For example, consider the choice of whether to drive or ride a bicycle to work. It is epistemically impossible that riding a bicycle to work will cause the laws of nature to be broken such that I can travel faster than the speed of light and therefore reach work earlier than I left. While the outcome of riding my bike if this did happen would be very nice—I could sleep in very late and still make it to work on time—this is irrelevant to my decision. But the SubU solution will sometimes tell us to take such (im)possibilities into account and must give up the very intuitive claim that only epistemically possible outcomes are relevant to practical decision-making.

These consequences are nicely drawn out by a series of cases recently introduced by Arif Ahmed (2014a,b). Ahmed takes his examples to be counterexamples to CDT. However, his analysis is (more or less implicitly) based on assuming the SubU solution to the Predetermination Problem. His examples, and the very plausible intuitions they are based on, are therefore counterexamples to the SubU solution. Here I will examine a modified version of one of Ahmed's examples, show that using the SubU solution gives the result that Ahmed suggests, and argue that this is the wrong result. Our example goes as follows (adapted from (Ahmed, 2014a, 666)):

> **Ahmed's Bet:** In my pocket, Billy says, I have a slip on which is written a proposition, P. You must choose between two bets. Bet 1 is a bet which pays \$10 if P is true and costs \$1 if P is false. Bet 2 is a bet which pays \$1 if P is true and costs \$10 if P is false. Before you choose whether to take bet 1 or bet 2, I should tell you what P is. It is the proposition that the state of the world yesterday was such as to determine that you now take bet 2.

If we understand outcomes indicatively Billy's description of the two bets here is inconsistent: bet 1 cannot pay out \$10 if P is true, because P being true entails that I will not take bet 1. But if we understand outcomes subjunctively it seems natural to accept Billy's description. Only a minor miracle—violation of the laws of nature—is required to make it that case that P is true and yet I take bet 1. And it is very plausible that the closest worlds to P worlds in which I take bet 1 are worlds in which bet 1 still pays out $10 when P is true—there is no obvious reason that the outcome of bet 1 would need to vary to allow me to take it when P is true.

It is possible to construct versions of the SubU solution which avoid the problematic result in this particular case by giving different utilities to the one's suggested by Billy's description. However, it is also possible to construct a parallel case which will be a counterexample to any such version of the SubU solution. All we need to do is specify a proposition P' which entails that we will take one or the other bet, and which gives the right utility values when we ask the question "supposing P', what would happen if I took bet 1'. Substituting this proposition for P will a counterexample with the same structure as Ahmed's Bet. It is hard to see how a non-ad hoc way of assigning outcomes to act-state pairs could avoid there being such propositions. I will, therefore, assume that we can rely on Ahmed's suggested utility values. Hence, we will have the decision matrix in Table 3.

	I am now determined to take bet 2 (and, therefore, P is true).	I am now determined to take bet 1 (and, therefore, P is false).	The outcome of my decision is not yet determined (and, again, P is false).
Take bet 1	$10	-$1	-$1
Take bet 2	$1	-$10	-$10

Table 3 Decision matrix for Ahmed's Bet.

Now, our choice has no causal influence over whether P is true; nor does it causally influence whether, if P is false, we are predetermined to take bet 1 or not yet determined either way. The truth of P, and what, if anything, we are predetermined to choose, are entirely settled by facts about the past and the laws of nature—over which we have no causal control. According to the SubU solution we can, therefore, rely on dominance reasoning. And it is obvious that bet 1 dominates bet 2—bet 1 is better no matter how the world turns out to be. Hence, the SubU solution endorses taking bet 1, *regardless of what our credences are*.

But, as Ahmed argues, this is intuitively the wrong result[27]: if we have a high enough credence that our choice is predetermined, then we should take bet 2. We are certain that if our choice is predetermined then taking bet 2 will get us $1—because P is just the proposition that we are predetermined to take bet 2, so if our choice is predetermined and we take bet 2 then P must be true. Similarly we are certain that if our choice is predetermined then taking bet 1 will lose us $1. Hence, on the assumption that our choice is predetermined it is better to take bet 2. Of course, on the assumption that our choice is *not* yet determined it is better to take bet 1. But whether or not our choice is predetermined is not causally influenced by it. So, if we are sufficiently sure that our choice is predetermined (specifically with these utilities, if our credence that our choice is predetermined either way is greater than $\frac{9}{11}$) we should take bet 2.

[27] I am somewhat modifying Ahmed's actual argument here. In particular I am including the possibility that our choice is not yet determined, which Ahmed ignores. The modification does no harm and makes the argument more general.

There are two worries that one might raise at this point: First, one might worry, as Joyce (2016) does, that Ahmed's reasoning must be mistaken, because it, falsely, assumes that we can make sense of decision-making under the assumption that our choice is predetermined. Second, one might be suspicious that a parallel argument could be constructed for one-boxing in Newcomb's problem. If so, causal decision theorists might have good reason to reject Ahmed's intuitions. I will consider these two worries in turn.[28]

First, Joyce argues that we can understand Ahmed's Bet as a case in which the agent is unsure which of two choices they face: with probability ϵ, they face a choice between the two bets on the assumption that their choice is *not predetermined*, and, with probability $1-\epsilon$, they face a 'choice' between the two bets on the assumption that their choice *is* predetermined. The problem, he contends, (and the reason for the scare quotes) is that we cannot see the latter choice as a real practical decision:

> An agent who deliberates about a decision which is framed so that each state *entails* a single act (and outcome), is engaging in an epistemic exercise, not an agential one (Joyce, 2016, 226).

This is because to see ourselves as faced with a real practical decision we must be able to compare how good both options (or their outcomes) are *for each way the world might be*. Otherwise we are comparing apples to oranges; comparing how good one act will be on one assumption against how good a different act will be on a different assumption. But, if predetermination is in the picture, we cannot compare each option for each way the world might be, because some of the relevant act-state pairs are impossible. Joyce then argues that the agent should therefore make her choice as though ϵ was equal to 1—that is, on the assumption that her choice is not predetermined. But to do this is just to accept the FFCD solution. If Joyce is right here he has shown that Ahmed's reasoning goes astray, and so cannot be an objection to the SubU solution, but only by showing that the FFCD solution must be accepted instead.

Moving to our second worry, the reasoning Ahmed relies on sounds very similar to the following reasoning in Newcomb's Problem: on the assumption that the predictor is infallible (i.e. incapable of being wrong), you are certain that if you one-box you will receive \$1,000,000, and that if you two-box you will receive only \$1,000. Hence you should one-box if you are sure enough that the predictor is infallible. (Note that the predictor being infallible is not the same as your being certain that the predictor is correct, see Sobel (1994, Ch. 5) for an illuminating discussion of the difference)

If this argument and Ahmed's argument really are parallel then it appears that causal decision theorists have a reason to reject Ahmed's reasoning: it forces us to give up unconditional two-boxing. I will assume here that this reasoning really is parallel, but it is worth pointing out that the problem with the SubU solution is fundamentally about considering (epistemically) impossible outcomes in our decision-making, and it may be possible to find cases where this gives the wrong result which do not share the structure of Ahmed's Bet or allow for a parallel argument to be constructed in Newcomb's problem.

Now, there are two ways that the predictor might be infallible. The first way is that there is some deterministic connection from our choice to the prediction (if the

[28] I thank an anonymous reviewer for this journal for encouraging me to deal with both of these worries.

prediction is in the past then this will have to be a kind of backwards causation). But that case is not of interest to us here, since everyone (whether they support CDT or EDT) agrees that we should at least sometimes one-box if our choice will causally influence the prediction. The other way is that the prediction is deterministically connected to (possibly by being identical too) the deterministic causes of our choice. If so, our choice must be predetermined. It is not possible for the prediction to be infallible, not causally influenced by our choice, and for our choice not to be predetermined. Hence, the relevance cases of infallibility all involve predetermination, and supporters of the FFCD and PPII solutions will not agree that the alleged parallel is a problem.

If the above arguments really are parallel, this shows that causal decision theorists have reason to reject Ahmed's intuitions on the assumption that causal decision theorists should endorse two-boxing in the face of an infallible predictor. But supporters of the PPII solution will reject this assumption; they will argue that we should sometimes one-box against infallible predictors. It is important to note that they do so because the PPII solution entails that we should sometimes one-box if we are sure enough that our choice is predetermined—which infallibility entails—and not because you are certain that the prediction is correct.

There are two intuitions here that are in conflict: on the one hand the intuition that we should two-box against infallible predictors, and on the other the intuition that only epistemically possible outcomes can be relevant to decision-making. Which intuition we should save is not obvious: It is a cost to give up unconditional two-boxing, since our original intuitions in Newcomb's problem seemed to support that. But it is also a cost to give up the intuition, which underlies Ahmed's judgments, that only epistemically possible outcomes can be relevant to decision-making.

Furthermore, even if causal decision theorists have good reason to reject one-boxing against infallible predictors, and hence to reject Ahmed's reasoning above and the PPII solution, this is not good reason to accept the SubU solution over the FFCD solution—the FFCD solution rejects the SubU solution's judgment in Ahmed style cases without rejecting two-boxing against infallible predictors.

To see that the FFCD solution and the SubU solution conflict, despite agreeing that we should unconditionally two-box, take the following modified version of Ahmed's Bet: the case is exactly as Ahmed's bet, except that if your choice is not predetermined then the payoffs are -\$10 if you take bet 1 and -\$1 if you take bet 2 (i.e. they are swapped), as shown in Table 4.

	I am now determined to take bet 2 (and, therefore, P is true).	I am now determined to take bet 1 (and, therefore, P is false).	The outcome of my decision is not yet determined (and, again, P is false).
Take bet 1	\$10	-\$1	-\$10
Take bet 2	\$1	-\$10	-\$1

Table 4 Decision matrix for modified Ahmed's Bet according to the SubU solution.

The SubU solution will endorse taking bet 1 in this case if your credence is high enough that your choice is predetermined (that one of the first two states obtains).

But this again looks like the wrong result. Bet 1 only looks better than bet 2 when we take into account the impossible outcome of getting \$10 by taking bet 1 when we are predetermined to take bet 2. Without this (im)possibility bet 1 is sure to lose us money. And it is sure to lose us more than bet 2—because it would only lose us less than bet 2 if it were, contrary to the facts, possible for us to lose \$10 by taking bet 2 when we were predetermined to take bet 1. Bet 2 is the better option in this case.

The FFCD solution agrees because it only takes the outcomes when our choice is not predetermined into account (i.e. it only assigns the right-most column positive probability). And supporters of the FFCD solution can appeal to something like Joyce's reasoning above to argue that the SubU solution gets things wrong *without* agreeing that you should one-box when faced with an infallible predictor. Hence, even if the alleged parallel between Ahmed's reason and one-boxing against infallible predictors holds *and* we have good reason (contra the PPII solution) to reject one-boxing against infallible predictors, this is not enough to show that the SubU solution is superior to the FFCD solution.

We have seen that using the SubU solution to undermine Undefined Utilities leads us into conflict with very plausible intuitions in the kind of cases introduced by Ahmed. These intuitions are justified by the thought that decision-making should only ever take into account outcomes which are epistemically possible, because practical decision-making is about making the right decision for how the world is, not how it would be if things were different. Endorsing the SubU solution forces us to give up this very intuitive link between decision-making and the actual world. This is a high price to pay to solve the Predetermination Problem.

6 Partition Dependence

There are formulations of CDT which do not appeal to causal probabilities, instead appealing to a privileged partition of states which allows us to capture the argument for two-boxing in Newcomb's problem with only unconditional evidential probabilities, most notably Lewis's (1981) influential dependency hypothesis formulation. You might wonder whether using such a *privileged-partition* formulation of CDT can escape the Predetermination Problem. If this were the case it would be a partial vindication of CDT, but still an important result in showing that partition-invariant formulations of CDT face problems that privileged-partition formulations do not—especially because partition invariance has several advantages that suggest we should want, if possible, to maintain it (see Joyce (1999, §5.5)). However, as it turns out, employing a privileged partition to avoid causal probabilities will not help. In this section I will examine Lewis's (1981) formulation of CDT and show that it does not avoid the problem, and that the same solutions are applicable to it. This demonstrates that partition invariance is not the root of the problem, and arguments similar to the one below will be applicable to any partition dependent formulation of CDT.

Lewis's decision theory requires that our states be *dependency hypotheses*. Dependency hypotheses are "maximally specific proposition[s] about how the things [the decision maker] cares about do and do not depend causally on [their] present actions" (Lewis, 1981, 11). The idea is that these dependency hypotheses will encode all the causal information we need, and then by calculating expected utility with the *unconditional* probability of these dependency hypotheses we can correctly distinguish real causal influence—which matters to decision-making—from merely evidential connections—which do not according to CDT. Lewis offers the following formula for expected utility (notation adjusted):

$$EU(A) = \sum_{K_i \in \mathbb{K}} P(K_i) \sum_{S_j \in \mathbb{S}} P(S_j | K_i \wedge A) U(S_j \wedge K_i \wedge A)$$

Here $\mathbb{K}$ is the partition of dependency hypotheses and $\mathbb{S}$ is any partition of states. $\mathbb{S}$ is required to deal with cases where the dependency hypotheses—which fix everything that is causally influenced by our acts—do not fix *everything* about how good the world will be. For example, in most realistic situations there will be many chancy events that influence how good the future is that are not causally related to our choices, using $\mathbb{S}$ allows us to account for this. In idealized cases, such as BMNPs, where no chancy processes are involved we can use $\mathbb{S} = \mathbb{K}$. We will then find, since $P(K_j | K_i \wedge A) = 0$ if $j \neq i$ and $P(K_j | K_i \wedge A) = 1$ if $j = i$, that:

$$EU(A) = \sum_{K_i \in \mathbb{K}} P(K_i) U(K_i \wedge A)$$

Now, to see how the Predetermination Problem arises for this formulation of CDT we will use again the \$1,000,000 Bonus Newcomb Problem from §4. To assess which action we should perform in this problem, using Lewis's formulation of CDT, we first need to know what the dependency hypotheses are. To work this out we can appeal to the partition of states we used in §4. Dependency hypotheses are complete specifications of how things we care about depend on our acts; one way to discover what these are is to take a partition of states—which necessarily includes every way the world might be—and work out for each state in the partition what will happen if we perform each act. This will give us a complete list of the outcomes our acts might have. We can then simply remove any duplicates from the list to find a partition of dependency hypotheses. (Of course this procedure is only easy to implement in cases where it is easy to work out the consequences of performing each act in every state, which will often not be the case.) Recall that our states are:

S_1: I am now determined to one-box, and the prediction is one-boxing.
S_2: I am now determined to one-box, and the prediction is two-boxing.
S_3: I am now determined to two-box, and the prediction is one-boxing.
S_4: I am now determined to two-box, and the prediction is two-boxing.
S_5: My choice is not yet determined, and the prediction is one-boxing.
S_6: My choice is not yet determined, and the prediction is two-boxing.

The question is, for each of these possibilities, what will the consequences of each choice be. For S_5 and S_6 this is easy to work out—since the consequences of our acts in these states only depend on what the prediction is—we will get two dependency hypotheses (numbered by the corresponding states that generates them):

K_5: If I one-box I will receive \$1,000,000 and if I two-box I will receive \$1,001,000.
K_6: If I one-box I will receive \$0 and if I two-box I will receive \$1,000.

What about the remaining states, where our choice is predetermined one way or the other? This is where we begin to run in to trouble. We need to know "what will happen if I one-box" (or, "what will one-boxing cause") and "what will happen if I two-box" for each of the remaining states. But if we understand these questions indicatively they have no answers. Consequently there is no maximally specific proposition about how the things the decision maker cares about are causally influenced by their present actions; if there were such a proposition it would have to tell us what will happen if I one-box in a world where I am predetermined to two-box, but that world is one where I do not one-box and so there is not fact of the matter about this. But without such facts we cannot find a complete set of dependency hypotheses and Lewis's formulation gives us no advice. This is the Predetermination Problem for Lewis's formulation of CDT.

The first solution that Lewis might avail himself of is the FFCD solution: require that all rational decision makers assume, in every decision, that their present choice is not predetermined, and hence that they assign zero credence to S_1–S_4. If so, we do not need any answer to these questions and K_5 and K_6 will be a complete set of dependency hypotheses (giving us the right result, by CDT's lights, that we should two-box). The costs of assuming that predetermination is impossible are no different for a partition-dependent decision theory than they are for a partition-invariant decision theory. Hence, doing so will incur all the costs of the FFCD solution discussed in §3.

However, we have another option (which Lewis himself would surely have endorsed): we can understand the dependency hypotheses *subjunctively*. Instead of asking "what *will* happen if I one-box and I am predetermined to two-box" in the indicative mood, we can ask "what *would* happen if I were to one-box and I were predetermined to two-box" in the subjunctive mood. Depending on which subjunctive conditionals we take to be relevant to answering this question the resulting version of CDT will be equivalent to the SubU or PPII solution.

In order to get CDT's result that we should two-box in Newcomb's problem the dependency hypotheses generated by S_3 and S_4 should be no different to the two already listed: Our being predetermined to two-box won't make any difference to what happens if we two-box. And one-boxing when we are predetermined to two-box cannot change what the prediction is without appeal to backtracking subjunctives conditionals—that is, subjunctive conditionals which appeal to worlds where the past is different. But allowing backtracking subjunctives into our dependency hypotheses undermines unconditional two-boxing. Instead our credence that our choice is predetermined will matter. For example, we might appeal to backtracking subjunctive conditionals to claim that if I one-box in S_4 then the prediction would have been one-boxing and so I would get \$1,000,000. We would then need a new dependency hypothesis:

K_4: If I one-box I will receive \$1,000,000 and if I two-box I will receive \$1,000.

The problem is that if we allow this sort of backtracking subjunctive in the \$1,000,000 Bonus Newcomb Problem there is no principled reason not to allow it in the case of the \$0 Bonus Newcomb Problem—that is, Newcomb's problem. But if this dependency hypothesis is allowed in Newcomb's problem then one-boxing will turn out to be the right option if the probability of this dependency hypothesis is high enough. This is analogous to using the PPII solution in a partition-invariant formulation of CDT. Any account of subjunctives that will guarantee unconditional two-boxing in Newcomb's problem will rule out backtracking subjunctive conditionals. And if the dependency hypotheses cannot appeal to backtracking subjunctive conditionals then K_5 and K_6 already account for S_3 and S_4 in BMNPs, and for all the states in Newcomb's problem. To avoid giving up unconditional two-boxing, and endorsing an analogue of the PPII solution, we will have to allow miracles—violations of the laws of nature—in assessing the relevant subjunctive conditionals, rather than differences in the past.[29]

K_5 and K_6 account for four of our states. But what about S_1 and S_2? We are considering the \$1,000,000 Bonus Newcomb Problem, and part of the specification of that case is that there will be an extra \$1,000,000 in the opaque box *if and only if* we are predetermined to one-box. If we do not wish to endorse the PPII solution we must allow miracles here. Presumably then, taking both boxes when we are predetermined to take one box will involve a minor miracle that leaves the bonus money in the box, and, hence, will get us \$2,001,000, if the prediction is one-boxing, and \$1,001,000 if the prediction is two-boxing (because the bonus will be in the box, but not the money tied to the prediction). Then we have our full partition of dependency hypotheses:

K_1: If I one-box I will receive \$2,000,000 and if I two-box I will receive \$2,001,000.
K_2: If I one-box I will receive \$1,000,000 and if I two-box I will receive \$1,001,000.
K_5: If I one-box I will receive \$1,000,000 and if I two-box I will receive \$1,001,000.
K_6: If I one-box I will receive \$0 and if I two-box I will receive \$1,000.

Now, we know that the causal probabilities of these dependency hypotheses are just their unconditional probabilities, since the whole point of using dependency hypotheses is that they are guaranteed not to be causally influenced by our choice. This will guarantee that we should two-box in BMNPs—two-boxing obviously dominates one-boxing given the above dependency hypotheses.

However, this is just the result that the SubU solution would have given us in a partition-invariant formulation of CDT. And now that we have allowed miracles into the picture, it is easy to see that we will get the wrong results in Ahmed's Bet. The first two states in Table 3 will serve as dependency hypotheses (taking bet 1 when P is true involving a minor miracle):

K_1: If I take bet 1 I will receive \$10 and if I take bet 2 I will receive \$1.
K_2: If I take bet 1 I will lose \$1 and if I take bet 2 I will lose \$10.

And given these dependency hypotheses we should take bet 1—conflicting again with Ahmed's very plausible intuitions and the idea that only epistemically possible outcomes can be relevant to decision-making.

We see then that Lewis's dependency hypothesis formulation of CDT faces the Predetermination Problem. We can solve it for Lewis by requiring predetermination

[29] Note that these are miracles relative to S_i, not relative to the worlds at which the subjunctive conditionals are assessed. That is to say, the account of subjunctives will have to allow that the worlds at which we assess the consequences of an act contain violations of the laws of nature of the worlds in the state we are considering, but not violations of their own laws—which are logically impossible.

to always receive zero credence—endorsing the FFCD solution. Or by using backtracking subjunctives in dependency hypotheses—endorsing the PPII solution. Or by allowing miracles in our dependency hypotheses—endorsing the SubU solution. The costs of endorsing these solutions are no different than they were for a partition-invariant formulation of CDT. Of course, this is not a refutation of Lewis's formulation of CDT—the Predetermination Problem must be solved somehow if we don't want to give up CDT—but appealing to partition-dependence has not made the Predetermination Problem any easier to solve or the costs any easier to bear.

7 Conclusion

The possibility that our choices are predetermined is a problem for causal decision theory. It gives rise to inconsistent act-state pairs in our decision problems. Intuitively there is no fact of the matter about what the outcomes of these inconsistent act-state pairs are, and hence no sensible utility can be assigned to them. But the causal probability of these act-state pairs is—according to the standard reasoning that justifies two-boxing in Newcomb's problem—non-zero. Hence, causal decision theory's expected utility calculations will include undefined terms multiplied by non-zero terms—leaving the entire sum undefined. We see then that the intuitive tension between believing that one's choices are predetermined and engaging in decision-making is not an illusion—it can be made precise.

The dialectic now looks like this: causal decision theorists want to maintain that you should two-box in Newcomb's problem. This rules out evidential decision theory because it endorses one-boxing even when your choice is not predetermined (whereas all three solutions here can agree that if your choice is not predetermined then you should two-box). However, predetermination raises another problem for CDT, and we must choose between one of the three ways to solve the problem:

The FFCD Solution: Require that the credences which all rational decision maker's use in their decisions are such that $P(S_i) = 0$ whenever the past and laws of nature of S_i entail that they will do ϕ or entail that they will not do ϕ, for any option ϕ under consideration in their current decision—that is, all rational decision makers assume, for the purposes of decision-making, that their present choice is not predetermined.

The PPII Solution: (Re)Define causal probabilities such that $P(S_i||A) = 0$ whenever S_i entails that we will not perform A—that is, accept that facts about the past and the laws of nature which determine what we will do are causally influenced by our choices or that causal probabilities must track more than causal influence.

The SubU Solution: Use subjunctive conditionals, or some other systematic method, to assign non-uniform utilities to inconsistent outcomes—accepting that epistemically impossible outcomes can be relevant to our decision-making.

And we have seen that none of these solutions will be acceptable to *all* causal decision theorists, even if every solution is acceptable to some.[30] Every solution must give up

[30] Does one of these solutions deserve the name *causal* decision theory better than others? I am not one to put much stock in names. But, more importantly, I think this question, and the implication that

a very plausible claim and take on costs associated with denying it—causal decision theorists face a trilemma (summarized in Table 5).

	Can we take into account the possibility of predetermination in making our decisions?	Should we two-box no matter how likely our evidence says predetermination is?	Should decision-making only take into account outcomes that might occur in the actual world?
FFCD	X	✓	✓
PPII	✓	X	✓
SubU	✓	✓	X

Table 5 The Trilemma Arising from the Predetermination Problem.

Giving up any one of these claims is a cost, so we must settle which is most important by further argument and refinement of our intuitions. When doing so we should bear in mind that insisting that we maintain any one of these three claims will not fully settle the issue: We must decide which two are most important to us—or, equivalently, which is the least important—to settle which horn causal decision theorists should grasp. I have not attempted to settle this question here. However we settle it, we must step on some toes and reject some intuitions—causal decision theorists have a predetermination problem.

Acknowledgements Thanks to Al Hájek, Tim L. Williamson, Jeremy Strasser, Wolfgang Schwarz, Alex Sandgren, the ANU decision theory reading group, and two anonymous reviewers for this journal for useful comments and discussion.

References

Ahmed A (2014a) Causal decision theory and the fixity of the past. *The British Journal for the Philosophy of Science* 65(4):665–685

Ahmed A (2014b) *Evidence, Decision and Causality*. Cambridge University Press, Cambridge

Cantwell J (2010) On an alleged counter-example to causal decision theory. *Synthese* 173(2):127–152

Cantwell J (2013) Conditionals in Causal Decision Theory. *Synthese* 190:661–679

Edgington D (2011) Conditionals, causation, and decision. *Analytic Philosophy* 52(2):75–87

Fernandes A (2016) Varieties of epistemic freedom. *Australasian Journal of Philosophy* 94(4):736–751

whichever does deserve the name therefore has a dialectic advantage, gets things the wrong way around. *Causal* decision theory was introduced to systematize the intuitions behind two-boxing in Newcomb's problem—we don't (or at least shouldn't) endorse two-boxing in Newcomb's problem because we have an antecedent commitment to the relevance of causal influence in decision-making. And we certainly shouldn't endorse two-boxing, or any particular solution to the Predetermination Problem, because of an antecedent commitment to some particular analysis of causation, say in terms of subjunctive conditionals. If it turns out that we have to give up the (exclusive) relevance of causation to best systematize all the relevant intuition then so be it.

Hájek A (2016) Deliberation welcomes prediction. *Episteme* 13(04):507–528

Hedden B (2012) Options and the subjective ought. *Philosophical Studies* 158(2):343–360

Jeffrey RC (1983) *The Logic of Decision*, 2nd edn. University of Chicago Press, Chicago

Joyce J (1999) *The Foundations of Causal Decision Theory*. Cambridge studies in probability, induction, and decision theory, Cambridge University Press, Cambridge ; New York

Joyce J (2016) Review of Arif Ahmed, Evidence, Decision and Causality. *The Journal Of Philosophy* 113:224–232

Lewis D (1973) *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts

Lewis D (1981) Causal decision theory. *Australasian Journal of Philosophy* 59(1):5–30

McKenna M, Pereboom D (2016) *Free Will: A Contemporary Introduction*. Routledge contemporary introductions to philosophy, Routledge, New York, NY

Nelkin D (2011) *Making Sense of Freedom and Responsibility*. Oxford University Press, Oxford

Nozick R (1969) Newcomb's problem and two principles of choice. In: Rescher N (ed) *Essays in Honor of Carl G. Hempel*, Springer Netherlands, Dordrecht, pp 114–146

Pearl J (2000) *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, U.K. ; New York

Savage LJ (1972) *The Foundations of Statistics*, 2nd edn. Dover Publications, New York

Sobel J (1994) *Taking Chances: Essays on Rational Choice*. Cambridge studies in probability, induction, and decision theory, Cambridge University Press, Cambridge ; New York