

# SOMETIMES IT IS BETTER TO DO NOTHING: A NEW ARGUMENT FOR CAUSAL DECISION THEORY

OLAV BENJAMIN VASSEND

*Inland Norway University of Applied Sciences*

It is often thought that the main significant difference between evidential decision theory and causal decision theory is that they recommend different acts in Newcomb-style examples (broadly construed) where acts and states are correlated in peculiar ways. However, this paper presents a class of non-Newcombian examples that evidential decision theory cannot adequately model whereas causal decision theory can. Briefly, the examples involve situations where it is clearly best to perform an act that will not influence the desired outcome. On evidential decision theory—but not causal decision theory—this situation turns out to be impossible: acts that an agent does not think influence the desired outcome are never optimal. Typically, sophisticated versions of evidential decision theory emulate causal decision theoretic reasoning by (implicitly) conditioning on causal confounders, but in the kind of example considered here, this trick does not work. The upshot is that there is more to causal reasoning than has so far been appreciated.

## 1. Introduction

The debate between Evidential Decision Theory (EDT) and Causal Decision Theory (CDT) concerns the extent to which causal information and presuppositions are necessary components of rational decision making.<sup>1</sup> In general, proponents of EDT hold that an act is rational for an agent if and only if it maximizes expected utility, where expected utility is typically calculated relative to a conditional probability distribution that represents the agent's degrees of belief. To proponents of EDT, causal relations have no special status in decision making. Indeed, historically, proponents of EDT have tended to find causality metaphysically suspect

---

1. At any rate, this is how I construe the debate. But the debate is notoriously slippery, because both theories come in multiple varieties, and there are no clearly agreed upon rules for what it would take for either position to come out on top.

**Contact:** Olav Benjamin Vassend <olav.vassend@inn.no>

and have consequently often wished to reduce or even eliminate its role in theorizing. Proponents of CDT, on the other hand, maintain that rational decision making needs to be explicitly based on causal information or assumptions and that an agent's (evidential) conditional probability distribution may fail to reflect accurately the causal information in a decision problem.<sup>2</sup> The main goal of this paper is to show that there are common scenarios in which causal information plays an indispensable role in rational decision making in a way that cannot be adequately modeled even in sophisticated versions of EDT.

Most of the debate between EDT and CDT has focused on whether each theory can get the "right verdict" in various scenarios, but the main point of this article is instead that there are certain natural *inferences* that cannot be modeled using the standard EDT framework. Already at this point, some readers may be tempted to object that EDT should not be in the business of trying to capture inferences anyway, since EDT is simply the view that rational decision making should be based on conditional probabilities, without being committed to any stance regarding how such conditional probabilities may be justifiably inferred. For reasons I explain in greater detail in Section 6, I do not think this position is tenable. In brief: (1) if there are no restrictions on how EDT theorists and CDT theorists are allowed to justify the probability distributions that they employ in expected utility calculations, then the difference between EDT and CDT vanishes. (2) In many scenarios, it is unrealistic to expect agents to have precise probabilities for various events, and in at least some such cases, the way agents make decisions will arguably be via the kind of inferences that will be discussed in this paper. (3) In any case, a complete theory of rational action should arguably also be able to capture the *reasons* why some acts are more rational than others. The focus should not merely be on getting the right verdicts, but getting those verdicts in the right way.

Section 2 gives an introduction to the versions of EDT and CDT that will be assumed in the paper. Section 3 lays out a new kind of example that I maintain EDT cannot adequately model. Section 4 shows that even sophisticated versions of EDT arguably cannot model the example correctly. Section 5 shows that CDT can. Section 6 responds to several objections. Finally, Section 7 concludes the paper with thoughts on why the type of example discussed in this paper has not been noticed earlier.

## 2. Background on Evidential and Causal Decision Theory

Although there are different formalizations of CDT, the version that I will assume in this paper uses the causal modeling framework due to Spirtes, Glymour, and

---

2. As I explain in greater detail later in the paper, this gloss on the distinction between EDT and CDT is not completely accurate, but it suffices as a rough picture.

Scheines (2000) and Pearl (2009) and is sometimes called “interventionist decision theory.” This version of CDT traces back to Meek and Glymour (1994) and is developed further in Pearl (2009), Hitchcock (2016), Stern (2017), Stern (2019), and Stern (2021).<sup>3</sup> In this framework, the “causal probability” of an effect  $E_j$  given an act  $A_i$  is sometimes denoted by using the notation  $P(E_j | do(A_i))$  (although we will see below that this notation is misleading).<sup>4</sup> Whereas the evidential probability  $P(E_j | A_i)$  represents the ordinary conditional probability that  $E_j$  will happen given that  $A_i$  happens,  $P(E_j | do(A_i))$  is supposed to be the probability that  $E_j$  happens on the condition that  $A_i$  is *made* to happen. Making  $A_i$  happen amounts to *intervening* on the variable  $A$ , thereby severing  $A$  from all its usual causes and forcing it to take on the particular value  $A_i$ .<sup>5</sup> Conceptually, conditioning on  $A_i$  and intervening on  $A_i$  seem to be quite distinct, but proponents of EDT tend to claim that in those cases in which CDT appears to get the right answer, everything captured with CDT’s causal probabilities can be captured with evidential probabilities as well (e.g., Eells 1982; Horwich 1987; and Jeffrey 1983).<sup>6</sup> At this point in the dialectic there is what may best be described as a stalemate between EDT and CDT, as there are well-known purported counterexamples to both views.<sup>7</sup> However, counterexamples to both CDT and EDT tend to be controversial and hard-to-interpret variants of the Newcomb problem where one’s actions and the desired effects are correlated due to a common cause structure. My main goal in this paper is to show that there is a fundamental difference between EDT and CDT that shows up in ordinary contexts that have nothing to do with Newcomb problems, and that CDT is the superior framework in those contexts.

Even though my argument does not depend on an analysis of Newcomb problems, I will start my discussion with a simple “Medical Newcomb” example, since a discussion of this example is a useful way of clarifying some of the major differences between the two frameworks.

Suppose an observational study with 1000 participants finds that the following inequality holds for the correlations between incidences of smoking and cancer:

---

3. See Joyce (1999) for a survey of other versions. Among philosophers, the consensus seems to be that the various versions of CDT are equivalent (this attitude is for instance expressed in Joyce 1999: 171 and Lewis 1981: 5). Pearl himself is, however, less sure of the equivalence of the different versions (see Pearl 2009: 240).

4. Here,  $E_j$  and  $A_i$  are specific values of the random variables  $A$  and  $E$ , each of which can take values in partitions  $\{A_1, A_2, \dots, A_n\}$  and  $\{E_1, E_2, \dots, E_m\}$ , respectively. More generally, the rest of this paper uses uppercase letters without subscripts to denote random variables and subscripted uppercase letters to denote specific values of those random variables.

5. I explain the *do*-operator in greater detail later in this section.

6. Price (2012) even claims that causal probability is just a kind of evidential probability.

7. See Egan (2007) for examples. Of course, Egan’s counterexamples have not gone unchallenged—see for instance Ahmed (2012).

$$P(\text{Cancer} | \text{Smoking}) > P(\text{Cancer} | \neg \text{Smoking}) \quad (1)$$

An agent who naively uses the above inequality to decide whether to smoke will conclude that it is better not to smoke, assuming the agent values avoiding cancer over the pleasure derived from smoking. Suppose, however, that genetic research suggests that, in fact, the following causal diagram is correct:

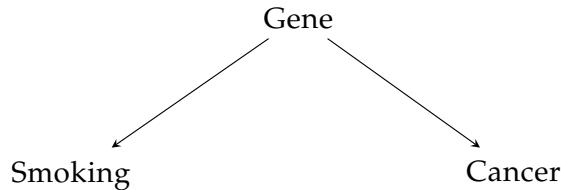


Figure 1: Counterexample to naive EDT.

The above diagram is an example of a Directed Acyclic Graph (DAG), a standard and convenient tool for representing causal relationships between variables.<sup>8</sup> According to the diagram, a gene is a (probabilistic) common cause of both cancer and smoking. More importantly, there is no arrow between Smoking and Cancer. The absence of such an arrow implies that Cancer and Smoking are causally independent even though they are probabilistically dependent. According to CDT, agents should be guided by causal dependencies, not probabilistic ones. Thus, CDT will say that the agent should smoke, just as long as the agent prefers smoking to not smoking. Naive EDT—that is, versions of EDT that use evidential probabilities uncritically—will neglect to take into account the agent’s causal knowledge and will issue the irrational recommendation that the agent abstain from smoking.

More sophisticated versions of EDT have the resources to handle the smoking gene example and other examples of the same kind, however. According to sophisticated EDT, one must (in effect) condition on variables that confound the correlation between acts and states of interest. In the smoking gene example, the Gene variable is a confounder since it completely mediates the correlation between Smoking and Cancer. When we condition on Gene ( $G$ ), the probabilistic dependence between Smoking ( $S$ ) and Cancer ( $C$ ) disappears—in other words,  $G$  screens off  $S$  from  $C$ :

$$P(\text{Cancer} | \text{Gene} \ \& \ \text{Smoking}) = P(\text{Cancer} | \text{Gene} \ \& \ \neg \text{Smoking}) \quad (2)$$

8. For thorough discussions of the causal interpretation of DAGs, see Pearl (2009) and Spirtes et al. (2000).

G is not the only potential variable we can condition on in order to remove the spurious correlation between S and C. For example, according to Eells’s “tickle defense” of EDT (Eells 1982), the variable that needs to be conditioned on in the Medical Newcomb problem might be a “tickle” for a cigarette that indicates to the agent whether he or she has the gene for smoking. In the above causal diagram, the “tickle” would be on the path between G and S; conditioning on the tickle variable would therefore have the same effect as conditioning on G itself.<sup>9</sup> In general, sophisticated EDT dictates that if one wants to know whether one should do  $A_i$  in order to bring about  $E_j$ , one cannot just look to  $P(E_j | A_i)$ . In order not to be fooled by spurious correlations between A and E, one must condition on some set of confounding variables between A and E. In other words, one must use  $P_C(E_j | A_i)$ , where the subscript C indicates that the relevant confounders are held fixed. Of course, defenders of EDT often do not put matters in terms of conditioning on confounding variables, but *in effect* that is precisely what they are doing from a causalist point of view. Eells’ tickle defense is one example of this strategy. Another example is Jeffrey’s ratifiability criterion (Jeffrey 1983: 16), according to which agents in the standard Newcomb problem need to condition on *having decided on an option*. In general, sophisticated EDT emulates CDT by finding an appropriate set of variables to condition on so as to get the same answer as CDT (in the cases where CDT intuitively gets the right answer—of course, in the original Newcomb case, many proponents of EDT hold that CDT does *not* get the right answer).

An interesting question, however, is whether a version of EDT that avails itself of conditioning on confounders is always able to imitate the verdicts of CDT (in cases where this is desirable). I will argue that the answer is “no”: there is more to CDT than just naive EDT + conditioning on confounders.

First, however, we must give a more careful statement of both EDT and CDT. Recall that standard decision theory assumes that the agent has a choice between a set of actions  $\{A_1, A_2, \dots, A_n\}$  that form a partition. Sophisticated EDT then says that the agent should choose the action  $A_i$  that maximizes so-called evidential expected utility (EEU-utility):

$$EEU(A_i) = \sum_j^m U(S_j, A_i) P_C(S_j | A_i) \tag{3}$$

Here,  $U(S_j, A_i)$  represents the utility of performing act  $A_i$  in state  $S_j$ , and  $P_C(S_j | A_i)$  is just a regular conditional probability.<sup>10</sup> The version of CDT propounded by Pearl (2009), on the other hand, recommends the action that has the highest CEU-utility, which is calculated as follows:

9. That is, conditioning on the tickle variable would also screen off S from C.

10. From now on, I will omit the subscript C.

$$CEU(A_i) = \sum_j^m U(S_j, A_i)P(S_j | do(A_i)) \tag{4}$$

The expression  $P(S_j | do(A_i))$  requires further explanation. Earlier I mentioned that  $P(S_j | do(A_i))$  is typically said to be the probability that  $S_j$  happens on the condition that  $A_i$  is made to happen. However, it is important to note that the expression  $do(A_i)$  does not refer to the proposition  $\langle A_i \text{ is made to happen} \rangle$ . Instead, the meaning of  $do(A_i)$  is best understood in the language of causal diagrams. As an illustration of what it means, suppose we have the following simple graph:



Suppose, moreover, that we have a joint probability distribution over  $A$ ,  $B$ , and  $C$ :  $P(A, B, C)$ . The graph then says that  $B$  depends causally on  $A$  (but not on  $C$ , except through  $A$ ), and that  $A$  causally depends on  $C$  (but not on  $B$ ). Given standard assumptions, the graph allows us to decompose  $P(A, B, C)$  as  $P(B | A) \times P(A | C) \times P(C)$ . Suppose we causally intervene on the variable  $A$  and set it to a particular value—thus we choose to  $do(A_i)$ . The causal graph is then transformed into the following one:



Figure 2: Example of a causal intervention.

The new joint probability distribution over  $A$ ,  $B$  and  $C$  is  $P(A, B, C | do(A)) = P(B | A) \times P(C)$ .<sup>11</sup>

As another illustration of what the  $do$ -operator is supposed to mean, consider again the DAG in Figure 1. The effect of  $do(\text{Smoking})$  on the DAG is to first change the DAG into the following one:

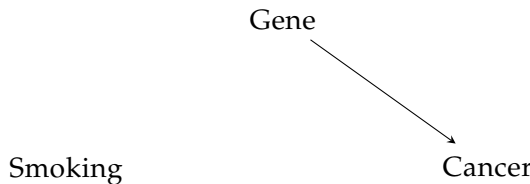


Figure 3: Counterexample to naive EDT.

In the *new* DAG represented in Figure 3, the effect of Smoking on Cancer is then found by setting the smoking variable to *Smoking*. The notation  $p(E_j | do(A_i))$  is misleading because it makes it seem as though we are conditioning on a

11. For a more extended discussion of how the  $do$ -operator differs from conditioning, see Pearl (2009: 73 and 242).



certain kind of proposition. However, as Hitchcock (2022) points out,  $do(A_i)$  is not an event in the original probability space at all and therefore cannot be conditioned on; he therefore suggests using the notation  $P_{do(A_i)}$  instead to denote the manipulated probability distribution that arises from performing the act  $do(A_i)$ . However, this notation is also somewhat infelicitous since it is already standard to use  $p_A$  to represent the probability distribution that arises by conditioning  $p$  on  $A$ . Zhang, Seidenfeld, and Liu (2019) instead use  $p^{do(A_i)}$ , which is the notation I will use in the remainder of the paper. Alternatively, I will write  $P^{A_i}$  where  $A_i = do(X_i)$  for some  $X_i$ , that is,  $A_i$  is an act that corresponds to a causal intervention (or lack thereof—see Section 3) in a DAG. Using the new notation, we get the following reformulation of causal expected utility (CEU-utility):

$$CEU(A_i) = \sum_j^m U(S_j, A_i) P^{A_i}(S_j) \tag{5}$$

We now have the necessary background in place to consider more carefully the differences between evidential and causal decision theory.

### 3. A Class of Non-Newcombian Examples

Suppose you have a computer that works fine most of the time, except that it sometimes freezes. You would like to fix your computer, but you have no particular technical expertise. What should you do? Obviously, the best thing would be to visit a repair shop. But suppose there is no repair shop nearby, so that the options are to try to repair the computer yourself or to leave it alone. It seems clear that it's best to leave the computer alone, because: (1) the computer already works fine most of the time, and leaving it alone is not going to change how it functions; (2) trying to repair the computer is likely to make it perform worse, given that you will not know what you are doing.

The above decision problem is an example of a broad class of problems for which the intuitively best option is not to interfere with the behavior of some system because interfering is likely to make its behavior worse. In many such problems, we do not have the ability to come up with justified probabilities for what is likely to happen given various acts we may choose. All we know—and all we need to know—is that the following two claims are true:

**Claim 1:** Not interfering with the system is not going to influence the probability that the system exhibits desirable behavior.

**Claim 2:** Interfering with the system in any way is going to decrease the probability that the system exhibits desirable behavior.

Claim 1 and Claim 2 jointly imply that not interfering with the system is the optimal act, so—despite not being able to come up with precise probabilities—we can often infer what we ought to do in this kind of decision problem. My claim is that an adequate formal decision theory needs to be able to capture this kind of inference. Somewhat surprisingly, Claim 1 and Claim 2 cannot both be true on EDT, so EDT cannot formally model this sort of inference. On the other hand, CDT is able to model the inference in a natural way.

It may be helpful to consider a concrete example. Consider the system represented in the following DAG (the simplest instance of the kind of example in which we are interested):

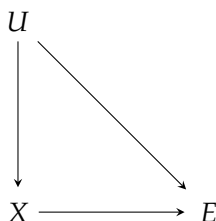


Figure 4: A causal system.

Here  $U$ ,  $X$ , and  $E$  are dichotomous variables, so that  $U$  can take the values  $U_1$  and  $U_2$ , and similarly for  $X$  and  $E$ . We may think of the DAG as representing part of some finely tuned mechanical or biological system. Note that the above graph implies that any joint distribution over  $X$ ,  $U$ , and  $E$  can be decomposed in the following way:  $P(X, U, E) = P(E | X, U) \times P(X | U) \times P(U)$ . Suppose—as is often realistic—that we do not have enough information to determine the precise probabilities of any of the events, but that we know (or assume) the following:

- (1)  $U_1$  and  $U_2$  have roughly the same (moderate) unconditional probability.
- (2)  $E_1$  if and only if  $U$  and  $X$  are in the same state, that is,
 
$$P(E_1) = P(X_2, U_2) + P(X_1, U_1)$$
- (3)  $P(X_2 | U_2)$  and  $P(X_1 | U_1)$  are both high.

$P$  may be an objective chance distribution, or it may represent our degrees of beliefs about the behavior of the system. For the sake of concreteness, I will assume in what follows that  $P$  represents our subjective degrees of belief. It follows from the preceding three facts that  $P(E_1)$  is high. In other words, our unconditional degree of belief is high that  $E_1$  will happen.

Suppose next that we have the option of interfering with the system and setting  $X$  to either  $X_1$  or  $X_2$  by force, or not interfering with the system, but instead doing some intuitively irrelevant act, such as going out to watch a movie at a theater. That is, the three possible acts are:



- $A_1$ : Intervene on the system and set  $X$  to  $X_1$
- $A_2$ : Intervene on the system and set  $X$  to  $X_2$
- $A_3$ : Do not interfere with the system/do some irrelevant act.

Suppose  $E_1$  has positive utility for us and that  $E_2$  has zero utility. More precisely, suppose that  $U(E_1, A_1) = U(E_1, A_2) = U(E_1, A_3) > 0$  and that  $U(E_2, A_1) = U(E_2, A_2) = U(E_2, A_3) = 0$ . Clearly, we prefer whatever action will make  $E_1$  as probable as possible. Given the above stipulations, I contend that both of the following claims are true:

**Claim 1:** Our degree of belief that  $E_1$  happens is unaffected by the supposition that  $A_3$  is chosen.

**Claim 2:** Our degree of belief that  $E_1$  happens is lowered by the supposition that either  $A_1$  or  $A_2$  is chosen.

These are of course more concrete versions of our earlier Claim 1 and Claim 2. As before, Claim 1 is plausible because not interfering with the system should not change its antecedent behavior, and hence our expectation about how the system will behave should not change on the supposition that we do nothing to the system. Claim 2 is plausible because our unconditional degree of belief that  $E_1$  will happen is already high, and given that we do not have much information about the system, it is unlikely that interfering with it will increase the probability that the desired outcome happens. Note that Claim 1 and Claim 2 jointly imply that  $A_3$  is the act that will maximize the probability of  $E_1$  and that  $A_3$  is therefore optimal.

It is important to note that there is not enough information to compute either the EDT-utility or CEU-utility of any of the acts. This is therefore plausibly a case where the *only* way we can arrive at the correct verdict about what to do is by justifying Claim 1 and Claim 2. The question, therefore, is whether Claim 1 and Claim 2 can be justified using the resources of either of the theories.

In the next section, we will see that Claim 1 and Claim 2 cannot both be true on EDT, and that EDT consequently is unable to adequately model this type of inference: on EDT, we must either hold that  $A_3$  influences the probability of  $E$ , or else we must accept that it cannot be the case that both  $A_1$  or  $A_2$  have a detrimental influence on the desired outcome. On the other hand, as is shown in Section 4, CDT vindicates both Claim 1 and Claim 2.

#### 4. EDT Is Inconsistent with the Conjunction of Claim 1 and Claim 2

On EDT, the object of interest will be some conditional distribution  $P(X, U, E | A)$ . Note that we do not want to assume that the various possible acts,  $A$ , have associated

probabilities, so we do not assume that we have a complete joint distribution over  $X, U, E$ , and  $A$ ; this assumption is not necessary in order to calculate the EEU-utility of various acts. However, we do need to assume that there is a conditional distribution  $P(X, U, E | A)$ , which is defined for each possible value of  $A$ .<sup>12</sup> The main goal in this section is to show that this conditional distribution—no matter how we decide to construct it—cannot satisfy both Claim 1 and Claim 2.

Now, on EDT, the claim that an act affects the probability of some outcome is naturally construed in terms of probabilistic dependence. Thus, to say that an act  $A_i$  affects the probability of an outcome  $O_j$  is to say that  $P(O_j | A_i)$  differs from  $P(O_j)$ , relative to some (or every) background context or confounder  $B$ . (Although note that in the example we are considering, there are no plausible confounders, so we can ignore this complication.) It follows, then, that the natural way of formalizing Claim 1 in the EDT framework is as follows:

**Claim 1 (EDT):** Our degree of belief that  $E_1$  happens is unaffected by the supposition that  $A_3$  is chosen:  $P(E_1 | A_3) = P(E)$ .

Note that the claim here is not that EDT is somehow committed to a reductive “probability raising” account of the causal influence that acts have on events. EDT is a theory of rational decision making and is in principle compatible with any metaphysical account of causal influence.<sup>13</sup> However, if the fact that  $A_3$  does not influence  $E_1$  is going to be evidentially relevant on EDT, then arguably the way this will enter into deliberation is via the above probabilistic equality.

Note also that the EDT version of Claim 1 is not saying that the *variable*  $A$  is probabilistically irrelevant to the variable  $E$ , since  $p(E | A_1)$  and  $p(E | A_2)$  are both plausibly different from  $p(E)$ . Instead, what the EDT version of Claim 1 is intended to capture is the idea that the *event*  $A_3$  will not affect the probability of the event  $E_1$  or  $E_2$ .<sup>14</sup>

The proper EDT precisification of Claim 2 is similarly straightforward:

**Claim 2 (EDT):** Our degree of belief that  $E_1$  happens is lowered by the supposition that either  $A_1$  or  $A_2$  is chosen:

- (i)  $P(E_1 | A_1) < P(E_1)$
- (ii)  $P(E_1 | A_2) < P(E_1)$

---

12. Obviously, this construal of conditional probability requires that we follow Hájek (2003) in regarding the conditional probability distribution as a primitive object rather than being defined in terms of the ratio formula.

13. I thank a reviewer for emphasizing this to me.

14. I thank Malcolm Forster for emphasizing this to me.

However, as we will soon see, the EDT versions of Claim 1 and Claim 2 lead to a contradiction; hence, at least one of them must be false. Crucially, the argument does not depend on the assumption that agents assign probabilities to their own acts. Equally crucially, the result cannot be evaded by conditioning on any set of confounding factors, so the usual way that sophisticated EDT gets out of its problems is not going to work.

As it happens, it is straightforward to come up with probability distributions such that either claim is true, so proponents of EDT can consistently accept either Claim 1 or Claim 2; however, they cannot accept both at the same time. That is a strange consequence. It means that proponents of EDT are forced to accept either that going out to see a movie (or, indeed, performing any other intuitively irrelevant act we may come up with) is somehow going to influence our degrees of belief about how the causal system in Figure 4 will behave, or they must accept that interfering with the system is necessarily going to make us expect that the system will perform better.

Note that there are two separate problems here. First, as I mentioned earlier, it is hard to see how we could arrive at the correct verdict that  $A_3$  is the optimal act without vindicating both Claim 1 and Claim 2. But aside from the fact that it is hard to see how one could arrive at the correct verdict except via Claim 1 and Claim 2, both of these claims are plausible in their own right: our decision theory should be able to model both the fact that  $A_3$  is thought to be irrelevant to the desired outcome and that  $A_1$  and  $A_2$  are both going to influence the desired outcome in a negative way.

It is of course possible that EDT has the resources to consistently capture Claim 1 and Claim 2 in a way that differs from the proposals I have given in this section. In that case, the challenge I present is for the proponents of EDT to come up with such a proposal. In Section 6, I will discuss one possible proposal that I believe fails.

The aim of the rest of this section is to show in greater detail why it is that the EDT precisifications of Claim 1 and Claim 2 are inconsistent with each other. In fact, we will show something more general. Let  $\{A_1, A_2, \dots, A_n\}$  be a partition of acts and let  $\{S_1, S_2, \dots, S_m\}$  be a partition of states. Suppose the utility function  $U$  satisfies the following constraint: for each  $S_j$ , we have, for all  $A_i$  and  $A_k$ , that  $U(S_j, A_i) = U(S_j, A_k)$ . This assumption entails that the value of a state does not depend on the act performed at that state; only the state itself matters. Hence, for all  $i$  and  $j$ , we can write the utility function purely as a function of the states:  $U(S_j, A_i) = U(S_j)$ . The assumption is satisfied in the example in Section 2, since in that example we only care about the value of the variable  $E$ . More generally, the assumption should arguably not be controversial, since we can always ensure that it is satisfied simply by partitioning the states finely enough.<sup>15</sup>

---

15. In the philosophical decision theoretic literature, it is often implicitly assumed that this is possible.

Let  $\sum_j^m U(S_j)P(S_j)$  be the expected utility not conditional on any act. In the kind of example discussed in Section 2, we may consider this the “antecedent” expected utility—that is, the utility we would expect to derive from the system if we were to leave it alone (supposing this is possible). We can now formulate the following generalizations of Claim 1 and Claim 2:

**Generalized version of Claim 1:** There exists an act  $A_n$  in the partition  $\{A_1, A_2, \dots, A_n\}$  that is thought to be irrelevant to every state, that is, such that for all possible states  $S_j$ , we have  $P(S_j | A_n) = P(S_j)$ .

**Generalized version of Claim 2:** For every  $A_i$  such that  $i < n$ , performing  $A_i$  will have a lower expected utility (EEU-utility) than the antecedent expected utility. That is, the following holds:

$$\forall i < n, \quad \sum_j^m U(S_j)P(S_j | A_i) < \sum_j^m U(S_j)P(S_j). \quad (6)$$

Claim 1 guarantees that there exists an act whose expected utility equals the antecedent expected utility, and implies, together with Claim 2, that this act has a higher expected utility than any other act. To proceed further, we need to make an additional assumption about the assignment of probabilities to acts. There has been much debate in the decision theory literature over whether agents, in general, assign probabilities to their own acts (e.g., Spohn 1977; Rabinowicz 2002; Ledwig 2005; Hájek 2016; Liu & Price 2019), especially during deliberation. I do not want to take a stance on the question of whether agents can assign probabilities to their own acts, and nothing in my subsequent argument will depend on the assumption that they can or do. However, I maintain that all rational agents will have degrees of belief that at the very least are *consistent* with having a probability function over their own acts.

To illustrate what this claim amounts to, suppose you are considering whether to study for your test. Suppose, moreover, that your conditional degree of belief that you will pass the class given that you study is 0.2 and that your conditional degree of belief that you will pass the class given that you don’t study is also 0.2, but that you have a 0.9 credence that you will pass the class. This combination of degrees of belief is clearly irrational, even if you have no degrees of belief over your own acts. You know that you will either study or not, and in either case your degree of belief is low that you will pass the test. Given these facts, it is irrational to have a high unconditional degree of belief that you will pass the test. The reason why this combination of degrees of belief is irrational is because it is not consistent with any possible probability function over the acts available to you. Hence, the following formal assumption seems well-justified:

**Probabilistic Extendability:** Given a set of conditional probabilities  $P(S_j | A_i)$  and unconditional probabilities  $P(S_j)$  defined over partitions  $S_j$  and  $A_i$ , there exists at least one probability distribution  $Q$  defined on the partition  $A_i$  such that  $P(S_j) = \sum_i P(S_j | A_i)Q(A_i)$

Probabilistic extendability merely requires that it be *mathematically possible* for any rational agent to have a joint probability distribution over the conjunction of states and acts, that is, over  $S_j$  &  $A_i$ . Note that there will typically be many probability distributions that satisfy the requirement in Probabilistic Extendability.

Let  $Q$  be one such probability distribution. That is,  $Q$  is a probability distribution defined on  $\{A_1, A_2, \dots, A_n\}$  such that  $P(S_j) = \sum_i P(S_j | A_i)Q(A_i)$ . Again, we do not assume that  $Q$  represents any agent's degrees of belief, nor indeed do we assume that any agent has any degrees of belief about their own acts at all. For us,  $Q$  is just a mathematical object that must exist for any rational agent. Now, multiplying (6) by  $Q(A_i)$  yields:

$$\forall i < n, \quad \sum_j^m U(S_j)P(S_j | A_i)Q(A_i) < \sum_j^m U(S_j)P(S_j)Q(A_i) \tag{7}$$

Next, summing each side of (7) over all  $i < n$  gives:

$$\sum_i^{n-1} \sum_j^m U(S_j)P(S_j | A_i)Q(A_i) < \sum_i^{n-1} \sum_j^m U(S_j)P(S_j)Q(A_i) \tag{8}$$

Claim 1 implies that  $U(S_j)P(S_j | A_n)Q(A_n) = U(S_j)P(S_j)Q(A_n)$ . Hence, we can extend the outer sums in (8) to  $i = n$ , which yields:

$$\sum_i^n \sum_j^m U(S_j)P(S_j | A_i)Q(A_i) < \sum_i^n \sum_j^m U(S_j)P(S_j)Q(A_i) \tag{9}$$

Rearranging the double sums in (9) yields:

$$\sum_j^m U(S_j) \sum_i^n P(S_j | A_i)Q(A_i) < \sum_j^m U(S_j)P(S_j) \sum_i^n Q(A_i) \tag{10}$$

But  $\sum_i^n P(S_j | A_i)Q(A_i) = P(S_j)$  and  $\sum_i^n Q(A_i) = 1$ , so (10) entails:

$$\sum_j^m U(S_j)P(S_j) < \sum_j^m U(S_j)P(S_j) \tag{11}$$

Since (11) is a contradiction, the conjunction of Claim 1, Claim 2, and Probabilistic Extendability is false. Since Probabilistic Extendability is a reasonable rational requirement, the problem must lie with the conjunction of Claim 1 and Claim 2.

## 5. CDT Vindicates Both Claim 1 and Claim 2

On causal decision theory, the object of interest is a manipulated distribution  $P^A(X, U, E)$ . Note the different notation—as emphasized before, this is not a conditional probability distribution;  $P^A$  is a (possibly new) probability distribution that results from acting on the original DAG in Figure 4. In fact, for different values of  $A$ ,  $P^A$  may result in probability distributions over distinct DAGs.

Given the set-up in Section 2, we clearly have  $A_1 = do(X_1)$  and  $A_2 = do(X_2)$ . How about the act of not intervening, that is,  $A_3$ ? In CDT, not interfering with a causal system has a very simple meaning: it is just an act that leaves the DAG that represents the causal system, as well as the joint distribution over that system, unaffected. In other words, by definition:  $P^{A_3}(X, U, E) = P(X, U, E)$ . Note that in the interventionist framework, there is always implicitly such an act: given any DAG, one can always choose *not* to intervene on the DAG.<sup>16</sup> Note that the  $A$  variable is not itself part of the DAG. In the version of the interventionist framework that uses Pearl's *do*-operator it is not necessary (in this example) to represent the various acts as part of the DAG; instead, acts are understood as operations on the original DAG in Figure 4.

Given what we have just said, it is clear that Claim 1 receives the following precisification on CDT, and it's clear that it is correct:

**Claim 1 (CDT):** Our degree of belief that  $E_1$  happens is unaffected by the supposition that  $A_3$  is chosen:  $P^{A_3}(E) = P(E)$ .

Note, again, that the CDT version of Claim 1 is saying that performing  $A_3$  will have no influence on  $E$  in two respects: it will neither change the causal relationship between the variables in the system (i.e., the original DAG in Figure 4.) nor will it change the probability of  $E$ . As we will see more clearly in the next section, this notion of independence is distinct from—and cannot be reduced to—the notion that  $E$  and  $A_3$  are probabilistically independent, even given confounders.

Claim 2, on the other hand, clearly receives the following interpretation on CDT:

**Claim 2 (CDT):** Our degree of belief that  $E_1$  happens is lowered by the supposition that either  $A_1$  or  $A_2$  is chosen:

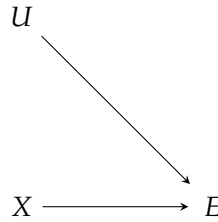
$$(iii) \quad P^{A_1}(E_1) < P(E_1)$$

$$(iv) \quad P^{A_2}(E_1) < P(E_1)$$

---

16. Zhang et al. (2019) use the notation  $do(x) = \emptyset$  to represent the idea that  $do(x)$  does nothing.

To verify that the CDT version of Claim 2 holds, we can use the following DAG, which represents the graphical effect that both  $A_1$  and  $A_2$  have on the relationship between  $X$ ,  $U$ , and  $E$ :



**Figure 5:** Effect of intervening on the causal system.

First, note that  $A_1$  and  $A_2$  (i.e.,  $do(X_1)$  and  $do(X_2)$ ) both break the causal relationship between  $X$  and  $U$ . This means that  $A_1$  and  $A_2$  change the antecedent joint distribution over  $X$ ,  $U$ , and  $E$ . For example,  $A_1$  changes the joint distribution over  $X$ ,  $U$ , and  $E$  as follows:

$$P^{A_1}(X,U,E) = P^{A_1}(E | X,U) \times P^{A_1}(U) \times P^{A_1}(X) \tag{12}$$

Since, by assumption,  $A_1$  forces  $X$  to take the value  $X_1$ ,  $P^{A_1}(X_2) = 0$ . Hence, we can sum both sides over the possible values of  $X$  and get:

$$P^{A_1}(U,E) = P^{A_1}(E | X_1,U) \times P^{A_1}(U) \tag{13}$$

It is clear that  $P^{A_1}(E | X_1,U) = P(E | X_1,U)$  and  $P^{A_1}(U) = P(U)$ , so we can rewrite (13) as follows:

$$P^{A_1}(U,E) = P(E | X_1,U) \times P(U) \tag{14}$$

Finally, setting  $E$  to  $E_1$ , summing over the possible values of  $U$ , and using the fact that  $P(E_1 | X_1,U_2) = 0$  (since  $E_1$  if and only if  $X$  and  $U$  match) yields:

$$P^{A_1}(E_1) = P(U_2) \tag{15}$$

In a precisely analogous way, we can show that:

$$P^{A_2}(E_1) = P(U_1) \tag{16}$$

Claim 2 now follows immediately, since, by assumption, both  $P(U_2)$  and  $P(U_1)$  are moderate whereas  $p(E)$  is high. Thus, both  $A_1$  and  $A_2$  will have a detrimental effect on the system. It follows that both Claim 1 and Claim 2 are true



and that CDT recommends that we do not interfere with the causal system, but instead perform the irrelevant act  $A_3$ , which is the intuitively correct result.

## 6. Three Objections and Their Rebuttals

In the following three subsections, I present and respond to three objections to the analysis that has been presented in the preceding sections.

### 6.1. Objection 1: EDT Gets the Right Verdict When We Include Our Act in the Causal Diagram

In previous sections, I have emphasized that  $do(X)$  is conceptually distinct from conditioning. This is true, but readers familiar with Meek and Glymour (1994) will know that there is a way of imitating what  $do(X)$  does in terms of conditioning, and may therefore suspect that it is possible to imitate what CDT does in an EDT framework that uses only conditional probability. I will argue that it is not. In order to model the decision problem in the alternative way, we extend the DAG in Figure 4 as follows:

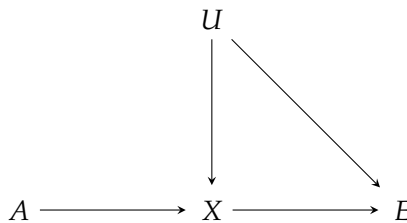


Figure 6: Extended DAG.

If we choose to represent the situation in terms of the DAG in Figure 6, then the next step is to place a conditional distribution  $Q(X,U,E|A)$  over this DAG that will mimic the effect of intervening on  $P$ . To mimic the effects of  $do(X_1)$  and  $do(X_2)$ , we define:

$$Q(X,U,E|A_1) = P^{do(X_1)}(X,U,E) \quad (17)$$

$$Q(X,U,E|A_2) = P^{do(X_2)}(X,U,E) \quad (18)$$

Given what we have found in Section 5, (17)–(18) imply a version of Claim 2 in that we will have:

$$Q(E_1|A_1) < P(E_1) \quad (19)$$

$$Q(E_1 | A_2) < P(E_1) \tag{20}$$

Furthermore, to represent the fact that  $A_3$  does not affect the behavior of the system (i.e., Claim 1), we define:

$$Q(X, U, E | A_3) = P(X, U, E) \tag{21}$$

That is, conditioning on  $A_3$  in the  $Q$ -distribution returns the original probability distribution  $P$ . Thus, if we have both the  $P$  and  $Q$  distributions, we can again vindicate (some version of) both Claim 1 and Claim 2. Furthermore, (19)–(21) will jointly imply that  $A_3$  is the optimal act, that is, we have:

$$Q(E_1 | A_1) < Q(E_1 | A_3) \tag{22}$$

$$Q(E_1 | A_2) < Q(E_1 | A_3) \tag{23}$$

An EDT theorist might now say that since we can represent the claim that  $A_3$  is better than  $A_1$  and  $A_2$  using the (normal) probability function  $Q$  alone, that means EDT can get the correct verdict that  $A_3$  is the optimal act.

I think this objection fails. As I have been emphasizing throughout the paper, the fact that  $A_3$  is optimal (i.e., (22)–(23) are true) is arrived at via an *inference* from (17)–(21), and (given the formalization used in this section), that inference involves both the  $P$  and  $Q$  distributions. Of course, there is nothing mathematically that prevents someone in the EDT framework from keeping around two distinct probability distributions, but it is hard to see the motivation for doing so from within the framework, because the difference between  $P$  and  $Q$  is naturally understood in causal terms:  $P$  represents the undisturbed “preintervention” probability distribution that represents our degrees of belief about the system when we regard it in isolation from our possible choices and actions, whereas  $Q$  represents an “entangled” distribution that represents our degrees of belief about the system when we view ourselves as causally interacting with it.

Proponents of EDT may reply that once we have arrived at (22)–(23), it does not matter how we got there: EDT is simply the view that you should do whatever has the highest conditional expected utility relative to your evidential probability function. How you arrive at your evidential probability distribution is your own business, and in principle you can adopt whichever distribution you like.<sup>17</sup> I think that if we adopt this point of view, then the distinction between EDT and CDT all but disappears, because it is *always* possible to imitate whatever verdict we get with CDT if we are free to choose our evidential probability distribution however we

---

17. Several reviewers and commentators have raised this objection.

like. For example, if we want to get the “right” (CDT) verdict in the medical Newcomb example discussed in Section 1, we can simply stipulate that our evidential probability function  $Q$  is such that  $Q(\text{Cancer} \mid \text{Smoking}) = P(\text{Cancer} \mid \text{do}(\text{Smoking}))$  and  $Q(\text{Cancer} \mid \neg \text{Smoking}) = P(\text{Cancer} \mid \text{do}(\neg \text{Smoking}))$ .  $Q$  will now be an “evidential” probability distribution<sup>18</sup> that gives the verdict that smoking is better than not smoking—we do not need to condition on any common causes and we certainly do not need to condition on any tickles.

Presumably, no proponent of EDT would accept this “solution” to the medical Newcomb problem. At least historically, a major motivation for EDT has been to reduce (or eliminate) explicitly causal presuppositions from decision making, because causality is supposedly metaphysically suspect. This is presumably why some proponents of EDT have taken the trouble to come up with elaborate tickle and metatickle justifications for why it is (in effect) legitimate to condition on causal confounders. In any case, regardless of what one may think about the metaphysics of causality, if the debate between proponents of EDT and proponents of CDT is to be of any substance, there need to be restrictions on what is considered a legitimate way of justifying an evidential probability. If all kinds of causal reasoning are allowed, then EDT can trivially emulate CDT in the way I have suggested.

As I said in the introduction, I take the debate between EDT and CDT to concern the extent to which causal reasoning plays a role in rational decision making. “Naive” EDT holds that it plays no role—if  $A_i$  maximizes the probability of  $E_j$  and  $E_j$  is the desired outcome, then you should do  $A_i$ , regardless of the causal relationship between  $A_i$  and  $E_j$ . “Sophisticated” EDT, on the other hand, holds (in effect) that the agent’s probability distribution needs to be appropriately conditioned on confounders. The argument I have presented shows that this recipe is not enough to successfully arrive at the correct verdict in the kind of example considered in this paper. On the other hand, the discussion in this section shows that a version of EDT that avails itself of two probability distributions (the  $P$  and  $Q$  distributions) could be successful. However—as was mentioned already—it is hard to see what might be an evidentialist motivation for making inferences on the basis of two distinct probability distributions.

## 6.2. Objection 2: Proponents of EDT Can and Should Reject Claim 1

I have claimed that in the kind of example introduced in Section 3, EDT will give the correct verdict that not interfering is best if and only if it is able to justify both Claim 1 and Claim 2. The basis for this claim is twofold: (1) these two claims are

---

18. It will be an evidential probability distribution in the formal sense that it is just a regular conditional probability distribution, where the event conditioned on is an act rather than an intervention.

independently plausible and therefore ought to be vindicated by any adequate theory of rational action, and (2) the two claims jointly entail the correct verdict. However, a possible response is to reject either Claim 1 or Claim 2 and then maintain that the correct verdict can be justified in some alternative way. In particular, if we replace Claim 1 with the following alternative, we will still get the correct entailment that not interfering is the optimal act:<sup>19</sup>

**Claim 1\* (EDT):** Our degree of belief that  $E_1$  happens is *increased* by the supposition that  $A_3$  is chosen:  $P(E_1 | A_3) > P(E_1)$ .

Moreover, Claim 1\* and Claim 2 *can* indeed both be true on EDT. In some ways, the fact that CDT vindicates Claim 1 while EDT vindicates Claim 1\* is precisely the difference between the two theories that I am trying to highlight in this paper. My goal in this subsection is to argue that replacing Claim 1 with Claim 1\* involves real costs.

To gain a more concrete understanding of the relevant issues, let us look at the example of the malfunctioning computer a bit more closely and see how (the interventionist version of) CDT handles the case as compared to a version of EDT that rejects Claim 1 and endorses Claim 1\*. Let  $C$  be a dichotomous variable where  $C_2$  is the event that your computer does not malfunction (in the next week, say) and let  $F$  be a three-valued variable where  $F_3$  is the event that you do not try to fix your computer. On the interventionist version of CDT, your prior (unconditional) degree of belief that the computer will not malfunction sometime in the next week can be—and plausibly should be—determined by your experience with the previous behavior of the computer. Thus, for example, if your experience is that your computer malfunctions every three weeks, on average, it is reasonable for you to have a degree of belief of roughly  $2/3$  that your computer will not malfunction in the next week. Thus,  $P(C_2) = 2/3$ . From a causal decision theoretic point of view, if you do not try to fix the computer, the computer will most likely keep behaving the way it has been behaving, and so  $P^{F_3}(C) = 2/3$ , that is, your degree of belief that the computer will not malfunction given that you do not try to fix it is just the same as your prior degree of belief that the computer will not malfunction, and both can be calibrated to match the observed frequency with which the computer has not malfunctioned in the past. On the other hand, since you don't know how to fix your computer, your degree of belief that the computer will not malfunction given that you try to fix your computer will be lower than your prior degree of belief that the computer will not malfunction.

All of the above makes intuitive sense, but the upshot of Section 4 is that—in contrast to interventionist CDT—EDT cannot vindicate the above picture. The

---

19. I am grateful to a reviewer for suggesting this objection.

reason is that, on EDT, the influence that acts have on events is formalized in terms of conditioning inside a single probability distribution. What does denying Claim 1 amount to in this context? Clearly it means that your degree of belief that the computer will not malfunction given that you do not try to fix it differs from your prior degree of belief that your computer will not malfunction. But this, in turn, can only mean one of two things: either your prior degree of belief that your computer will not malfunction is *not* determined by your prior experience with the computer, or you believe that your not interfering with your computer will make your computer's behavior deviate from its established norm. The latter is implausible, so it is the former option that is the EDT proponent's best bet. But the upshot is that you cannot just calibrate your prior degree of belief that your computer will malfunction to match the observed frequency with which your computer has malfunctioned in the past.

There are independent grounds for thinking that this is, indeed, how things must be on EDT. Suppose you assign probabilities to your own acts. Then—on EDT—your prior degree of belief that your computer does not malfunction can be written in the following form:

$$P(C_2) = P(C_2 | F_1)p(F_1) + P(C_2 | F_2)P(F_2) + P(C_2 | F_3)P(F_3) \quad (24)$$

In other words, your prior degree of belief that your computer will not malfunction is partly determined by the probabilities you assign to your own acts. This is a strange state of affairs, since it means that in an EDT framework (at least a version of the framework in which we are allowed to assign probabilities to our own acts), you will have to assign probabilities to your own acts before you can assign prior probabilities to any other event that you could potentially influence!

Suppose, for example, that you aren't even considering whether to repair your computer; you just want to assign a prior probability to  $C_2$  because—at some future point—you want to be able to take into account evidence relevant to  $C_2$  by using Bayes's formula. The implication of (24) is that the prior probability you assign to  $C_2$  is entangled with your degrees of belief about your own acts; it is simply not possible for you to have a degree of belief about the behavior of the computer that is separate from your beliefs about your own acts.

Of course, as was mentioned earlier, EDT is not committed to the thesis that agents assign probabilities to their own acts. Even so, the above considerations imply that if you want to assign a prior probability to  $C_2$  *and* you want to get the correct verdict that you should not try to fix your computer, you cannot do what is arguably most natural—you cannot just calibrate your prior probability to match the frequency with which  $C_2$  has happened in the past, and this is arguably bad enough.

Crucially, the interventionist version of CDT does not have the same issue. On the interventionist version of CDT, we always have two different ways of

forming degrees of belief about the behavior of some causal system: we can think about the causal system itself—that is, apart from our own choices and acts (this is equivalent to using the pre-intervention  $P$  distribution from the previous subsection); or, alternatively, we can think about ourselves and the system as forming an entangled causal system (this is equivalent to using the  $Q$  distribution from the preceding subsection). Moreover, we can compare our degrees of belief formed in either of these ways and on that basis form a decision about whether it is a good idea to intervene on the system at all. On EDT, on the other hand, our choices and acts are *always* entangled with the systems about which we are deliberating. Again, this is simply a consequence of using a single probability distribution to represent our degrees of belief and of representing the influence that our acts have on events in terms of conditional probabilities. And I maintain that this is a significant drawback of EDT as compared to interventionist CDT.

### ***6.3. Objection 3: Proponents of EDT Don't Need to Take Any Stance on Either Claim 1 or Claim 2***

A final response I have often received to my argument is that proponents of EDT do not need to take any stance on either Claim 1 or Claim 2, because EDT is just a theory of rational action and Claim 1 and Claim 2 concern the way in which acts relate to events rather than which acts are rational. The *only* commitment of EDT—the objection continues—is that conditional rather than causal probabilities should be used in decision making. In fact, this is an objection I have been trying to address at several points in the paper, but it is worth addressing it head on here. In doing so, I will in part be summarizing some things that I have already said.

First, as I have been emphasizing at several points, EEU-utility is not always explicitly calculable. For example, in cases where the needed conditional probabilities are not all available, one may need to rely on inferences of the sort that I have been discussing in this paper (i.e., from Claim 1 and Claim 2 to the conclusion that not interfering is the optimal act). In any case, I contend that agents often *do* decide how to act on the basis of these sorts of inferences rather than explicit expected utility calculations, and surely a complete theory of rational action should be able to capture the reasons why agents choose the way they do. The question, therefore, becomes whether EDT has the resources to justify the required inference. I have argued that it does not.

Second, one cannot characterize EDT in a merely formal way as the view that conditional probabilities rather than causal probabilities should be used in decision making. From a purely mathematical point of view, conditional probabilities



can be whatever you want them to be, so if there is to be any interesting difference between EDT and CDT at all, proponents of EDT need to say something about which conditional probabilities can be justified from a purely evidential point of view. I have argued that the only way to arrive at the “correct” conditional probabilities in the type of example introduced in this paper involves an inference that can only be vindicated with causal reasoning that cannot be modelled in the standard EDT framework.

Third, I reject the claim that EDT is simply a view about rational action. Implicitly, it is also a view about how one ought to understand the way one’s acts influence events. As I said in the previous subsection, EDT is implicitly committed to a view according to which agents are forced to regard their own acts and choices as being “entangled” with outside events in a way that is simply not the case on the interventionist version of causal decision theory. This, too, is a drawback of EDT.

## 7. Concluding Thoughts

The discussion in this paper highlights a distinction between CDT and EDT that has so far been overlooked in the literature. In my view the distinction points to a clear advantage in favor of CDT in general and the interventionist formulation of CDT in particular. Why has this distinction been missed?

I believe the main reason why interventionist decision theorists have missed it is because many of them tend to view decisions as interventions,<sup>20</sup> but conceiving of decisions in this way neglects the fact that *not intervening* is also a decision that can be modeled in the interventionist framework.

Another reason why the distinction between CDT and EDT discussed in this paper has been missed is that most of the discussion in the literature has focused on Newcomb-type problems, and in those types of problems all the acts have the same causal/graphical effect on the DAG. Consider, for example, the medical Newcomb example discussed in Section 1. If we perform the intervention *do(Smoking)*, then we will get exactly the same DAG as we will get if we perform the alternative intervention *do(¬Smoking)*; that is, intervening to either smoke or not smoke will result in the exact same causal diagram, namely the one shown in Figure 3. By contrast, in the type of examples considered in this paper, some of the acts change the graphical relationships between variables, while others leave those relationships intact. To adequately model what is going on in these situations, we cannot just use conditioning within a single probability distribution.

---

20. Meek and Glymour (1994) is a notable exception.



## Acknowledgments

I submitted the first version of this paper to *Analysis* in 2013. Since then, it has gone through countless expansions and revisions. I am grateful to reviewers at many journals (most of whom have disliked the paper) and to Malcolm Forster and Reuben Stern for reading several drafts. I am especially grateful to Reuben Stern for much discussion and encouragement over the past ten years.

## References

- Ahmed, Arif (2012). Push the Button. *Philosophy of Science*, 79(3), 386–95.
- Eells, Ellery (1982). *Rational Decision and Causality*. Cambridge University Press.
- Egan, Andy (2007). Some Counterexamples to Causal Decision Theory. *Philosophical Review*, 116(1), 93–114.
- Hájek, Alan (2003). What Conditional Probability Could Not Be. *Synthese*, 137(3), 273–323.
- Hájek, Alan (2016). Deliberation Welcomes Predictions. *Episteme*, 13(4), 507–28.
- Hitchcock, Christopher (2016). Conditioning, Intervening, and Decision. *Synthese*, 193, 1157–76.
- Hitchcock, Christopher (2022). Causal Models. In Edward N Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2022/entries/causal-models/>
- Horwich, Paul (1987). *Asymmetries in Time*. MIT Press.
- Jeffrey, Richard (1983). *The Logic of Decision* (2nd ed.). Cambridge University Press.
- Joyce, James (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.
- Ledwig, Marion (2005). The No Probabilities for Acts-Principle. *Synthese*, 144(2), 171–80.
- Lewis, David (1981). Causal Decision Theory. *Australasian Journal of Philosophy*, 59(1), 5–30.
- Liu, Yang and Huw Price (2019). Heart of DARCness. *Australasian Journal of Philosophy*, 97(1), 136–50.
- Meek, Christopher and Clark Glymour (1994). Conditioning and Intervening. *British Journal for the Philosophy of Science*, 45(4), 1001–21.
- Pearl, Judea (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
- Price, Huw (2012). Causation, Chance, and the Rational Significance of Supernatural Evidence. *Philosophical Review*, 121(4), 483–538.
- Rabinowicz, Wlodek (2002). Does Practical Deliberation Crowd Out Self-Prediction? *Erkenntnis*, 57(1), 91–122.
- Spirtes, Peter, Clark Glymour, and Richard Scheines (2000). *Causation, Prediction, and Search* (Second ed.). MIT Press.
- Spohn, Wolfgang (1977). Where Luce and Krantz Do Really Generalize Savage’s Decision Model. *Erkenntnis*, 11(1), 113–34.
- Stern, Reuben (2017). Interventionist Decision Theory. *Synthese*, 194, 4133–53.
- Stern, Reuben (2019). Decision and Intervention. *Erkenntnis*, 84, 783–804.
- Stern, Reuben (2021). An Interventionist Guide to Exotic Choice. *Mind*, 130(518), 537–66.
- Zhang, Jiji, Teddy Seidenfeld, and Hailin Liu (2019). Subjective Causal Networks and Indeterminate Suppositional Credences. *Synthese*, 198(27), 6571–97.