The Causal Decision Theorist's Guide to Managing the News

J. Dmitri Gallow

Causal decision theory (CDT) says that you should do your best to improve the world in which you find yourself. A potential action is choiceworthy to the extent that you expect it to promote valuable ends. In contrast, evidential decision theory (EDT) says that you should act so as to give yourself good news about the world. An act is choiceworthy to the extent that it indicates valuable ends. Causal decision theorists complain that EDT prescribes an irrational policy of 'managing the news'—favoring acts that provide good news about the world, even when they do less to *improve* the world than the alternatives. CDT's solution is to ignore correlations between your acts and states of nature when it is evaluating the choiceworthiness of acts—that is, its solution is to ignore any 'news value' your acts may carry (§1).

Your acts may carry news about the good the world has provided, and such news is rightly disregarded when deciding how to act. But, by providing news about states of nature, your acts may also carry the news that they will do the most possible to improve things, and such news ought not be disregarded. In ignoring correlations between your acts and states of nature, CDT ignores important correlations between your acts and the goods they are in a position to bring about. CDT is therefore in need of revision; and I have a revision to offer (§4). On this theory, an act is choiceworthy to the extent that its performance indicates that it does more to improve the world than any alternative would. The theory deals with objections to CDT raised by Hunter & Richter, 1978 (§3), Ahmed, 2014 (§5), and Spencer & Wells, 2017 (§6).

1 Causal Decision Theory

Suppose you associate with each epistemically possible world w a *value*, $\mathcal{V}(w)$. As I'll use the term here, a world need not settle all truths, but it must settle all truths which are not evaluatively neutral. If we assume the number of worlds to be finite, then with this value function and your subjective probability function Pr, defined over sets of these worlds, we may define the value of a *proposition* (a

Draft of 12 October, 2018; Word count: 6,851
Comments appreciated. ✉: jdmitrigallow@pitt.edu

set of worlds), ϕ, as follows:[1]

$$\mathcal{V}(\phi) \stackrel{\text{def}}{=} \sum_{w} \Pr(w \mid \phi) \cdot \mathcal{V}(w)$$

Suppose you face a decision in which you have available the acts $A_1, A_2, \ldots, A_M$, and $\{S_1, S_2, \ldots, S_N\}$ is a partition of states of the world. Then, it follows from this definition that the value of performing act A is given by

$$\mathcal{V}(A) = \sum_{S} \Pr(S \mid A) \cdot \mathcal{V}(SA)$$

Evidential decision theory (EDT) advises you to choose whichever act maximizes this quantity, $\mathcal{V}(A)$. It advises you to choose the act which has the highest expected value, conditional on your having performed it. In a slogan, it tells you to act so as to give yourself the best news about the world. In most ordinary cases, this coincides with doing your best to improve the world. However, there is a class of decision problems in which acting so as to give yourself good news can lead you to make matters worse.[2] These are cases in which your choice is correlated with a state of the world over which you exercise no control—throughout, I'll call such a state a 'state of nature'. Consider Newcomb's Problem.[3]

> Newcomb's Problem
> On the celebrity version of a game show, you are given a choice of two doors. Behind door #1 is either £1,000,000 or nothing at all. Behind door #2, there is definitely £10,000. Normally, contestants have to choose between taking a chance on winning a million pounds with door #1 or walking away with the guaranteed £10,000 behind door #2. Since this is the celebrity version of the game, and you are playing for charity, they've made the game a bit easier: if you want, you are free to open *both* doors and take whatever money you find. Incidentally, before the show was taped, the producers analyzed your social media accounts with AI bots in an effort to predict how you would behave. If the bots predicted that you would open only one door, then the producers put £1,000,000 behind door #1. If, however, they predicted that you would open both doors, then they put nothing behind door #1. The predictions of these bots are 51% reliable.[4]

The decision you face in Newcomb's Problem is an easy one. You get to open both doors and take both prizes. You may find yourself in a good world in which £1,000,000 awaits behind door #1, or you may find yourself in an unfortunate

1 I abuse notation, writing 'Pr(w)' rather than 'Pr($\{w\}$)'.

2 There is some controversy about whether these are genuine decision problems at all. See Jeffrey (1993, 2004) for an argument that they are not, and see Joyce (2007) for a reply.

3 See Nozick (1969).

4 That is, the probability that the bots predicted you would open both doors, given that you do, is 51%. And the probability that the bots predicted you would only open one, given that you only opened one, is 51%.

world in which no money awaits behind door #1. If the world is good, taking both prizes does the most to improve it. If the world is unfortunate, even so, taking both prizes does the most to improve it. So, in either case, taking both prizes does the most good. So you should take both prizes.

EDT, however, advises you to leave a prize behind. Let 'O' (for *o*ne) be the proposition that you only open door #1, and let 'B' (for *b*oth) be the proposition that you open both doors. Let 'M' (for *m*illion) be the proposition that there are a million pounds behind door #1. Then, the expected value of O is higher than the expected value of B, since[5]

$$\begin{aligned}\mathcal{V}(O) &= \Pr(M \mid O) \cdot \mathcal{V}(MO) \; + \; \Pr(\neg M \mid O) \cdot \mathcal{V}(\neg MO) \\ &= 0.51 \cdot 1{,}000{,}000 \; + \; 0.49 \cdot 0 \\ &= 510{,}000\end{aligned}$$

whereas

$$\begin{aligned}\mathcal{V}(B) &= \Pr(M \mid B) \cdot \mathcal{V}(MB) \; + \; \Pr(\neg M \mid B) \cdot \mathcal{V}(\neg MB) \\ &= 0.49 \cdot 1{,}010{,}000 \; + \; 0.51 \cdot 10{,}000 \\ &= 500{,}000\end{aligned}$$

I say that EDT recommends leaving a prize behind, but not all defenders of EDT agree. Some say that, in my analysis of Newcomb's Problem, I neglected one crucial piece of evidence at your disposal. Prior to deliberation, you will know whether you are leaning towards selecting both doors or whether you are leaning towards selecting only the one—that is, you will have information about your pre-deliberation inclinations.[6] Let's call this additional information a 'tickle'. If you feel a tickle, then you should incorporate this information by conditionalizing on it. But, since any correlation between the bots' predictions and your choice must be due entirely to the factors which caused you to have the tickle, conditionalizing upon this information will screen off your action from the bots' prediction.[7] And once your action is screened off from the bots' prediction, EDT will recommend taking both prizes. More generally, since EDT says that your pre-deliberation probabilities and values should determine your choice, and since you have introspective access to your initial probabilities and values, these probabilities and values must always correlate with the bots' predictions to the same degree as your final act. And that means that your pre-deliberation probabilities and values should completely screen off your act from the bots' prediction. So EDT will recommend taking both prizes.

The causal decision theorist needn't, and shouldn't, deny that the tickles of our initial inclinations, probabilities, or values can teach us something about the

---

5 I suppose that your utilities are linear in sterling, that you are not risk-adverse, and so on and so forth.

6 See Eells (1981, 1982).

7 To say that the tickle, T, 'screens off' your decision, A, from the bots' prediction, P, is just to say that, conditional on T, A and P are probabilistically independent—*i.e.*, $\Pr(PA \mid T) = \Pr(P \mid T) \cdot \Pr(A \mid T)$.

likely consequences of our action. What they should deny is that tickles like these will *always* do so—for two reasons. Firstly, you may not have introspective access to your own probabilities and values.[8] Secondly, and relatedly, you need not be following the advice of EDT in order for EDT to *give* that advice. Suppose that you make this decision unreflectively, without bothering to think about it at all. In that case, even though you do not follow EDT, EDT may still be used to evaluate your act as rational or irrational. Then, because your deliberation provided you with no information to screen off your act from the bots' predictions, EDT will say that leaving a prize behind was the rational choice.[9]

Amongst those who urge you to open both doors, a common diagnosis of where EDT has gone wrong is this: when EDT is considering how valuable the act A is, it uses a probability function conditionalized on A. But conditionalizing upon A can provide you with two, importantly different, kinds of information. Firstly, it can provide information about states which are causally downstream of your act. In this way, conditionalizing on A provides you with information about the good your act stands to causally promote. Secondly, it can provide information about states which are correlated with, though not causally downstream from, your act. In this way, conditionalizing on A provides you with information, not about the good you stand to promote, but rather about the good the world has provided for you.

We may separate the states of the world, S, into those factors which are not causally downstream of your act—call these factors 'K'—and those factors which are causally downstream of your act—call those factors 'C'. Then,

$$\begin{aligned}
\mathcal{V}(A) &= \sum_S \Pr(S \mid A) \cdot \mathcal{V}(SA) \\
&= \sum_K \sum_C \Pr(KC \mid A) \cdot \mathcal{V}(KCA) \\
&= \sum_K \underbrace{\Pr(K \mid A)}_{\textit{evidential}} \cdot \sum_C \underbrace{\Pr(C \mid KA)}_{\textit{causal}} \cdot \mathcal{V}(KCA)
\end{aligned}$$

In the difference it makes to the terms $\Pr(C \mid KA)$, conditionalizing on the proposition A provides *causal* information about which states your act *promotes*. In the difference it makes to the terms $\Pr(K \mid A)$, conditionalizing on the proposition A provides *evidential* information about which states your act merely *indicates*, but does not cause.

The causal decision theorist thinks that EDT's irrational policy of managing of the news has arisen from EDT's evaluating acts not only by the good they *promote*, but additionally by the good they *indicate*. Evaluating acts in this way leads you to endorse perverse acts which promote bad outcomes but at the same

---

8 Perhaps a suitably idealized agent *would* have initrospective access to their own probabilities and values. Even so, you may not. *Cf.* Lewis (1981).

9 Here, I have presupposed that EDT is at least partly an *evaluative* theory which says whether a given act is rational or irrational for a given agent in a given decision scenario, even if that agent isn't being *guided by* that theory. There are other ways of understanding EDT; and these alternative understandings may escape the Newcomb objection.

time carry good news about factors outside of your control. The causal decision theorist therefore suggests removing A's evidential influence on $\mathcal{V}(A)$ by replacing the terms $\Pr(K \mid A)$ with $\Pr(K)$. Thereby, CDT does not consider the merely evidential value of the act A. Rather, it considers only its *causal* value. Let's call the resulting quantity the *expected efficacy* of an act, $\mathcal{E}(A)$,[10]

$$\begin{aligned}\mathcal{E}(A) &\overset{\text{def}}{=} \sum_K \Pr(K) \cdot \sum_C \Pr(C \mid KA) \cdot \mathcal{V}(KCA) \\ &= \sum_K \Pr(K) \cdot \mathcal{V}(KA)\end{aligned}$$

While EDT advises you to choose an act with maximal expected value, $\mathcal{V}(A)$, CDT advises you to choose an act with maximal expected efficacy, $\mathcal{E}(A)$.

Importantly, the causal decision theorist does not deny that your action can indicate good states of the world. Indeed, if they did not think that your actions could indicate good states of the world, then they would not think that Newcomb's Problem is a counterexample to EDT. And the act you have selected can continue to indicate things about the state of the world, even after it has been selected. Once you perform an act, B, you will learn that you have performed it. Since conditionalization is a rational way to learn from experience, your new probability function will be your old probability function conditionalized upon the proposition B, $\Pr(- \mid B)$. Just as you may evaluate the efficacy of an act A from your current perspective, so too may you evaluate A's expected efficacy from the perspective you will occupy after having performed B. Let's use '$\mathcal{E}(A \mid B)$' for your best estimate of A's efficacy, conditional on your having performed B,

$$\mathcal{E}(A \mid B) \overset{\text{def}}{=} \sum_K \Pr(K \mid B) \cdot \mathcal{V}(KA)$$

Say that the act B *recommends* the act A iff, conditional on the information that B has been performed, no other act has greater expected efficacy than A.

Recommendation The act B *recommends* the act A iff $\mathcal{E}(A \mid B) \geqslant \max_X \mathcal{E}(X \mid B)$.

Thus, an act *recommends itself* iff no other act has a higher expected efficacy than it, on the assumption that it is selected.

Given the quantities $\mathcal{E}(A \mid B)$, for every pair of acts A, B, we may construct what I will call the *recommendation structure* of a decision problem. The recommendation structure of a decision problem will tell us which acts every act recommends. For instance, the recommendation structure of Newcomb's Problem is shown in figure 1. On the assumption that you have taken both prizes, B, taking both prizes has a higher expected efficacy than leaving a prize behind, O. So B recommends itself. And, on the assumption that you have left a prize behind, selecting both prizes still has a higher expected efficacy than leaving a prize behind.

[10] This is Skyrms (1980, 1982)'s formulation of CDT. For alternatives, see Gibbard & Harper (1978), Lewis (1981), Joyce (1999), Meek & Glymour (1994), Hitchcock (2016), and Stern (2017). The differences between these formulations are not relevant to anything I will be doing here.

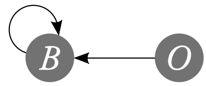

Figure 1: The recommendation structure of Newcomb's Problem

So O recommends B. So B is universally recommended.

If an act is universally recommended, then CDT will advise you to perform it. However, there need not be a universally recommended act. This can happen in three different ways. Firstly, it could turn out that no act recommends itself (§2). Secondly, it could turn that that multiple acts recommend themselves (§3). Thirdly, it could be that, even though there is a unique act which recommends itself, there are other acts which recommend against that self-recommending act (§7).

2 Death in Damascus

For a decision problem in which no act recommends itself, consider Death in Damascus.[11]

> Death in Damascus
> You must choose to travel to either Aleppo and Damascus. You know that, tomorrow, Death will look for you in one of these cities. If you and Death are in the same city, you will die; if you and Death are in different cities, you will live. Death is good at making predictions about where you will be, but he's not perfect. Moreover, he has a tendency to guess Damascus. The probability that Death is in Damascus, given that you are in Damascus, is 90%. Whereas the probability that Death is in Aleppo, given that you're in Aleppo, is only 60%.

Whether you live or die is the only factor relevant to your decision, and you prefer living to dying. If we let 'D' be the act of going to Damascus, 'A' be the act of going to Aleppo, and use 'K_D' and 'K_A' for Death's being in Damascus and Aleppo, respectively, then your values are as shown in table 1.

In Death in Damascus, no act recommends itself. For, on the assumption that you've gone to Aleppo, going to Damascus has a higher expected efficacy. And on the assumption that you've gone to Damascus, going to Aleppo has a

11 *Cf.* Gibbard & Harper (1978). The original Gibbard & Harper example had symmetric probabilities for Aleppo and Damascus. Egan (2007) uses asymmetric Death in Damascus cases to argue against CDT. However, in Egan's cases, there is an asymmetry in the values of performing the acts, rather than an asymmetry in the probabilities. In particular, in Egan's cases, one of the two acts—the *safe* act—will always yield the status quo, whereas the other act—the *risky* act—will be either very good or very bad, depending upon a factor outside of your control, but which is highly correlated with your act. If you select the safe act, you should expect the risky one to have very good consequences. But, if you select the risky act, then you should expect it to have very bad consequences. Here, too, no act recommends itself. Egan believes that you should choose the safe act.

	K_D: Death in Damascus	K_A: Death in Aleppo
D: You go to Damascus	0	10
A: You go to Aleppo	10	0

Table 1: Values for Death in Damascus

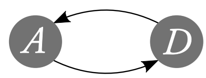

Figure 2: The recommendation structure of Death in Damascus

higher expected efficacy.

$$\mathcal{E}(A \mid A) = 4 \qquad \mathcal{E}(D \mid A) = 6$$
$$\mathcal{E}(A \mid D) = 9 \qquad \mathcal{E}(D \mid D) = 1$$

So, as soon as you start to follow CDT's advice, it will issue new recommendations. As soon as you find yourself going to Aleppo, and you therefore raise your probability in the proposition A, CDT will tell you to go to Damascus instead. And as soon as you find yourself going to Damascus, and therefore raise your probability in the proposition D, CDT will tell you to go to Aleppo. (See figure 2.) In cases where no act recommends itself, the verdicts of CDT are unstable.[12]

3 The Hunter Richter Game

For a case in which multiple acts recommend themselves, consider the Hunter Richter Game.[13]

> The Hunter Richter Game
> Your twin is very much like you. In similar decision scenarios, you tend to make the same choices. And you are given the opportunity to play a game with your twin. Both you and your twin can, if you decide to play, press either button A, button B, or button C. Or you might decide to not play and not press any button. If you and your twin both press the same button, then you will both win £100. If you and your twin press different buttons, then you will both lose £100. If either you or your twin decide not to play, you neither win nor lose any money.

[12] There are *deliberational* versions of CDT which solve this issue with instability by advising you to get yourself into a position in which you have probability a that you will go to Aleppo and probability $1 - a$ that you will go to Damascus, and, from this deliberative perspective, going to Aleppo and Damascus look equally efficacious. From this deliberative standpoint, you should either perform the mixed act of going to Aleppo with probability a and going to Damascus with probability $1 - a$, or else, perhaps, just *pick* either Aleppo or Damascus. See Skyrms (1990), Arntzenius (2008), and Joyce (2012).

[13] From Hunter & Richter (1978). See also the discussion in Lewis (1981).

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} K_A & K_B & K_C & K_D \end{array} \\ \begin{bmatrix} 100 & -100 & -100 & 0 \\ -100 & 100 & -100 & 0 \\ -100 & -100 & 100 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Table 2: Values for The Hunter Richter Game. K_A is the state in which your twin chooses A; and similarly for K_B, K_C, and K_D.

Your values in the Hunter Richter Game are shown in table 2. For the sake of concreteness, let's suppose that the probability that your twin chooses A, K_A, given that you choose A, is 70%. And similarly for B, C, and D. Suppose also that, given that you select an option and your twin does not select that option, none of the remaining options are any more likely to have been selected by your twin than any other.

Suppose that your act probabilities at the beginning of deliberation are uniform: $\Pr(A) = \Pr(B) = \Pr(C) = \Pr(D) = 1/4$.[14] Then, because of the correlation between your acts and those of your twin, your prior distribution over the states K_A, K_B, K_C, and K_D will be similarly uniform, and the expected efficacy of pressing A will be negative,

$$\begin{aligned} \mathcal{E}(A) &= \Pr(K_A) \cdot \mathcal{V}(AK_A) + \Pr(K_B) \cdot \mathcal{V}(AK_B) + \Pr(K_C) \cdot \mathcal{V}(AK_C) + \Pr(K_D) \cdot \mathcal{V}(AK_D) \\ &= 1/4(100) + 1/4(-100) + 1/4(-100) + 1/4(0) \\ &= -25 \end{aligned}$$

as will the expected efficacy of pressing B or C. However, the expected efficacy of not playing will be zero. So CDT will advise you to not play.

Your deliberative humility was crucial for CDT delivering this verdict. Suppose that, rather than starting out relatively unopinionated about which option you would end up selecting, you had instead started out 60% confident that you would select A. Then, you would have begun deliberation thinking that your twin was 46% likely to select A, and your pressing A would have a positive expected efficacy of 10, while B and C would both have a negative expected efficacy of -46, and D would still have an expected efficacy of 0. So CDT would advise you to press A. Similarly, if you had begun deliberation more confident than not that you would press B, CDT would advise you to press B; and, if you had begun deliberation more confident than not that you would press C, CDT would advise you to press C. For this reason, in The Hunter Richter Game, every available act recommends itself. The recommendation structure of The Hunter Richter Game is as shown in figure 3a.

Though every act recommends itself, CDT does not say that every act is permissible. Instead, it says that which acts are permissible depends upon how confi-

[14] Throughout, I'll refer to your unconditional probabilities that you will perform a particular act, $\Pr(A_i)$, as your *act probabilities*.

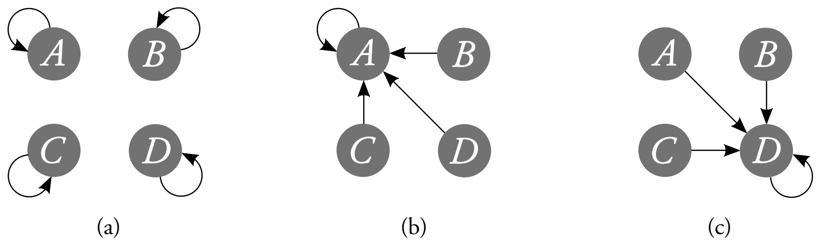

FIGURE 3: The recommendation structure of THE HUNTER RICHTER GAME—with (3b and 3c) and without (3a) tickes.

dent you are, at the beginning of deliberation, that you will end up selecting each act at deliberation's end. In particular, CDT says that you should press a button only if you begin deliberation more confident than not that you will end up deciding to press that button. Perhaps this is as it should be. After all, your initial probability that you will press A is evidence about whether your twin will press A. Information about which button your twin will press is information about the good your own pressing is in a position to accomplish. More generally, since CDT says that your pre-deliberation probabilities and values should determine your choice, and since you have introspective access to your initial probabilities and values, these probabilities must always provide relevant information about how you will decide, and therefore, they must always provide relevant information about how your twin will decide.

Perhaps. For perhaps your initial probabilities are 'tickles' which provide information about which button your twin will press (*cf.* §1). Learning about this tickle, you should conditionalize upon it. Doing so will render your act and your twin's act probabilistically independent. If your initial probability for A was high enough, then, after you conditionalize on the tickle, every act will recommend pressing A (figure 3b). If your initial probabilities for A, B, and C were not high enough, then, after conditionalizing on this tickle, every act will recommend not pressing (figure 3c).

We needn't, and shouldn't, deny that the tickles of our initial probabilities can teach us something about the likely consequences of our action. But, just as in our discussion of NEWCOMB'S PROBLEM, we should deny that they will *always* do so—and for the same two reasons. Firstly, you may not have introspective access to your own probabilities and values.[15] Secondly, you need not be following the advice of CDT in order for CDT to *give* that advice. Suppose that both you and your twin follow EDT. In that case, even though you do not follow CDT, CDT may still be used to evaluate your act as rational or irrational. In doing so, it will take your initial act probability for A to be relevant to your decision, even though this act probability does not give you any relevant information about which button either you or your twin will push.[16]

[15] Perhaps a suitably ideal agent *would* have introspective access to their own probabilities and values. Even so, you may not. *Cf.* LEWIS (1981).

[16] Here, I have presupposed that CDT is at least partly an *evaluative* theory which says whether

4 Manage the Improvement News

The Hunter Richter Game brings out something odd about the way CDT treats your act probabilities. Grant that, in the Hunter Richter Game, these probabilities provide evidence about how your twin will decide. Even so, you should not treat this probability as if it were a factor outside of your control—something you must passively observe in order to learn how best to act. Part of what it is to be a freely deliberating agent is to be in a position to revise your act probabilities at will. For, if you are free, then you are in a position to manufacture the evidence that you will press *A* by simply forming the intention to press *A*. If it is in your power to press any button, then it is in your power to rationally raise your act probability that you will press that button. Exactly this position is defended by arch-causal decision theorist Joyce,

> ...a rational agent, *while in the midst of her deliberations*, is in a position to legitimately ignore any evidence she might possess about what she is likely to do. She can readjust her probabilities for her currently available acts at will...A deliberating agent who regards herself as free need not proportion her beliefs about her own acts to the antecedent evidence that she has for thinking that she will perform them.[17]

In this, I think Joyce is exactly correct. But if, in the Hunter Richter Game, you are in a position to give yourself evidence that you will press *A* by forming the intention to press *A*, then you are in a position to give yourself the news that your twin will press *A*, and, thereby, to give yourself the news that your pressing *A* has maximal expected efficacy.

How should this influence your deliberation? What kind of news should you want to give yourself? EDT says that you should want to give yourself propitious evidence of the provisions of nature. This is irrational. The provisions of nature are out of your hands; they do not speak in favor of acting one way or another. In its attempt to avoid EDT's irrational managing of the news, the causal decision theorist says that you should never manage the news—that you should never choose so as to give yourself propitious evidence. But there are two, importantly different, kinds of news that your act is in a position to provide. Your act can provide evidence about the provisions of nature. So too can it provide evidence about the degree to which it will *improve upon* the provisions of nature. The former evidence is rightly disregarded, but the latter is not.

Note that, if CDT is correct, then we may measure the choiceworthiness of

a given act is rational or irrational for a given agent in a given decision scenario, even if that agent isn't being *guided by* that theory. There are other ways of understanding CDT; and these alternative understandings may escape the present objection.

17 Joyce (2007, p. 557, italics in original). Joyce (2018, §5) will want to qualify these remarks by adding that "your credences for acts should respond only to evidence about their *choiceworthiness*." On this, Joyce and I are in agreement. (Though we will disagree about what makes acts choiceworthy.)

an act A with $\mathscr{C}_{\text{CDT}}(A)$, where

$$\mathscr{C}_{\text{CDT}}(A) \stackrel{\text{def}}{=} \mathcal{E}(A) - \max_X \mathcal{E}(X)$$

If $\mathscr{C}_{\text{CDT}}(A) = 0$, then you expect A to do at least as much to improve things as any other available act, and A is *perfectly* choiceworthy. If $\mathscr{C}_{\text{CDT}}(A)$ is negative, then there is another available act which you expect to improve things more than A, and A is less than perfectly choiceworthy. The difference $\mathcal{E}(A) - \max_X \mathcal{E}(X)$ tells us how far short of doing the most possible to improve things A falls. Thus, $\mathscr{C}_{\text{CDT}}(A)$ tells us how far from perfect choiceworthiness A falls.[18] CDT tells you to prefer acts which you expect to improve things more to acts which you expect to improve things less.[19]

Causal Decision Theory (CDT) A is to be preferred to B iff $\mathscr{C}_{\text{CDT}}(A) \geqslant \mathscr{C}_{\text{CDT}}(B)$.

When CDT evaluates acts, it does so from the perspective you *currently* occupy—it takes expectations given your *current* act probabilities, and not the act probabilities you will come to have after selecting A. In contrast, EDT looks at the good you *will* expect the act to bring about, if the act is selected. If EDT is correct, then we may measure choiceworthiness of an act, A, with $\mathscr{C}_{\text{EDT}}(A)$, where

$$\mathscr{C}_{\text{EDT}}(A) \stackrel{\text{def}}{=} \mathcal{E}(A \mid A) - \max_X \mathcal{E}(X \mid X)$$

If $\mathscr{C}_{\text{EDT}}(A) = 0$, then A will be expected to bring about as much good, once it is selected, as any other act, X, will be expected to promote, once X is selected. In that case, performing A gives you the best possible news about the good your act will promote, and EDT says that A is perfectly choiceworthy. If $\mathscr{C}_{\text{EDT}}(A)$ is negative, then there will be some other act which gives you better news about the good you will promote, and EDT says that A is less than perfectly choiceworthy. The difference $\mathcal{E}(A \mid A) - \max_X \mathcal{E}(X \mid X)$ measures how far short of giving the best news possible A falls. Thus, $\mathscr{C}_{\text{EDT}}(A)$ tells us how far short of perfectly choiceworthiness A falls. EDT tells you to prefer acts which provide better news about the good they will promote to those which provide worse news about the good they will promote.

Evidential Decision Theory (EDT) A is to be preferred to B iff $\mathscr{C}_{\text{EDT}}(A) \geqslant \mathscr{C}_{\text{EDT}}(B)$.

I say that you should choose so as to give yourself the news that you are doing the most possible to improve things. In a slogan: you should manage the improvement news. CDT says that you should do the most possible to improve things, but because it uses unconditional expectations, it does not consider how performing the act will affect your opinions about which act does the most possible to improve things. EDT evaluates acts on the basis of the news they carry, but

[18] Thus, $\mathscr{C}_{\text{CDT}}$, unlike $\mathcal{E}$, measures choiceworthiness on a ratio scale, rather than an interval scale. This is as it should be, for there is a theoretically significant zero point, corresponding to perfect choiceworthiness.

[19] Throughout, 'prefer' means 'weakly prefer'.

it conflates news about factors outside of your control with news about what you yourself are in a position to accomplish. If we wish to manage the improvement news, we may measure the choiceworthiness of an act, A, with $\mathscr{C}_{\text{MIN}}(A)$, where

$$\mathscr{C}_{\text{MIN}}(A) \stackrel{\text{def}}{=} \mathcal{E}(A \mid A) - \max_X \mathcal{E}(X \mid A)$$

If $\mathscr{C}_{\text{MIN}}(A) = 0$, then $\mathcal{E}(A \mid A)$ is as high as $\mathcal{E}(X \mid A)$, for any other available act X. If that's so, then performing A gives you the news that A does at least as much to improve upon the provisions of nature as any other available act would. In that case, A is perfectly choiceworthy. If $\mathscr{C}_{\text{MIN}}(A)$ is negative, then A carries the news that some other act would improve things more than it. In that case, A is less than perfectly choiceworthy. I say that you should prefer acts on the basis of the news they provide about the degree to which they do a better job improving things than the alternatives would.[20]

Manage the Improvement News (MIN) A is to be preferred to B iff $\mathscr{C}_{\text{MIN}}(A) \geqslant \mathscr{C}_{\text{MIN}}(B)$.

Suppose that A recommends itself. Then, for any act X which maximizes $\mathcal{E}(X \mid A)$, $\mathcal{E}(X \mid A)$ will be equal to $\mathcal{E}(A \mid A)$. In that case, A will be perfectly choiceworthy. Performing a self-recommending act gives the news that you are doing the most you can to improve the world in which you find yourself. It is always permissible. Thus, in The Hunter-Richter Game, any act is permissible. Suppose that an act does not recommend itself and there is another act which does. Then, the self-recommending act will be perfectly choiceworthy, while the non-self-recommending act will be less than perfectly choiceworthy. So you should prefer the self-recommending act. In particular, if there is one and only one act which recommends itself, then that act will be the only permissible one. So, in Newcomb's Problem, opening both doors is the only permissible option.

Note that, given the measures $\mathscr{C}_{\text{CDT}}$ and $\mathscr{C}_{\text{EDT}}$, there will always be some available option which is perfectly choiceworthy. Not so for $\mathscr{C}_{\text{MIN}}$. Consider, for instance, Death in Damascus. Going to Damascus gives you the news that going to Aleppo would do more to improve things, so D is less than perfectly choiceworthy: $\mathscr{C}_{\text{MIN}}(D) = \mathcal{E}(D \mid D) - \mathcal{E}(A \mid D) = -8$. Similarly, going to Aleppo gives you the news that going to Damascus would do more to improve things, so A is less than perfectly choiceworthy: $\mathscr{C}_{\text{MIN}}(A) = \mathcal{E}(A \mid A) - \mathcal{E}(D \mid A) = -2$. No matter how you choose, you will end up giving yourself the bad news that you are doing less than you could to improve things. Even so, not all bad news is equally bad. By going to Aleppo, you give yourself the bad news that you've fallen short of doing the most possible to improve things. But, by going to Damascus, you

[20] A nearby theory is defended by Barnett (ms). Barnett says that A is to be preferred to B iff $\mathcal{E}(A \mid B) - \mathcal{E}(B \mid B) \geqslant \mathcal{E}(B \mid A) - \mathcal{E}(A \mid A)$—that is, roughly, iff A is a greater improvement on B (assuming that B is chosen) than B is an improvement on A (assuming that A is chosen). As Barnett demonstrates, his theory leads to cyclic preferences. In contrast, MIN will never lead to cyclic preference. There are also parallels between the motivations for MIN and the motivations cited in favor of Wedgwood (2013)'s *benchmark* theory—though problems for that theory do not carry over to MIN.

$$\begin{array}{c} \\ D \\ A \\ F \end{array} \begin{array}{c} \begin{array}{cccc} K_{DD} & K_{DA} & K_{AD} & K_{AA} \end{array} \\ \left[\begin{array}{cccc} 0 & 10 & 0 & 10 \\ 10 & 0 & 10 & 0 \\ -\Delta & 10-\Delta & 10-\Delta & -\Delta \end{array} \right] \end{array}$$

TABLE 3: Values for DICING WITH DEATH. K_{DD} is the state in which the die reads 'Damascus' and Death awaits in Damascus; K_{DA} is the state in which the die reads 'Damascus' and Death awaits in Aleppo; and similarly for K_{AD} and K_{AA}.

give yourself the *worse* news that you have fallen *even further* short of doing the most you could to improve things. So, according to MIN, while neither A nor D is *perfectly* choiceworthy, A is still *more* choiceworthy than D is, and you should prefer it on that basis.[21]

5 Dicing with Death

Consider the following decision problem, due to AHMED (2014):

> DICING WITH DEATH
> As in DEATH IN DAMASCUS, you must choose to either go to Aleppo or Damascus. Once again, Death is a very good predictor of which city you will choose, and if Death predicts correctly, then you will die. Suppose that the probability of Death getting his prediction correct is 90%, irrespective of the city you happen to select. A man makes you an offer: he has a die that you can use to make your decision about where to go—half of its faces say 'Aleppo', and half say 'Damascus'. He has rolled the die already, and he informs you that, while Death is very good at predicting your choice, Death is unable to predict the outcome of this die roll. Given that you select your city by following the die, the chance that Death will find you is only 50%. In exchange for using the die, the man asks for a pittance, Δ.

In DICING WITH DEATH, you face a choice between three potential acts: going to Aleppo, A, going to Damascus, D, and following the die, F. None of these acts recommend themselves. On the assumption that you go to Aleppo, following the die looks as though it would have better consequences, and going to Damascus looks as though it would have better consequences still. On the assumption that you've gone to Damascus, following the die looks to have better consequences,

[21] In EGAN's cases (see footnote 11), what MIN recommends will depend upon your values. The risky option could be either very bad or very good, depending upon a factor outside of your control. If it is just as bad in the bad state as it is good in the good state (that is, if $\mathcal{V}(G)-\mathcal{V}(Q) = \mathcal{V}(Q)-\mathcal{V}(B)$, where G is the good outcome, B the bad, and Q the status quo), then MIN says that both the risky and the safe options are permissible. If the risky option is more bad in the bad state than it is good in the good state, then MIN says to choose the safe option. And, if the risky option is less bad in the bad state than it is good in the good state, then MIN says to choose the risky option.

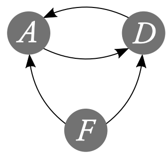

FIGURE 4: The recommendation structure of DICING WITH DEATH.

and going to Aleppo, better still. On the assumption that you follow the die, both going to Aleppo and going to Damascus look to have better consequences—since both will give you the same chance of survival as following the die, yet save you the pittance, Δ.

$$\mathcal{E}(A \mid A) = 1 \qquad \mathcal{E}(D \mid A) = 9 \qquad \mathcal{E}(F \mid A) = 5 - \Delta$$
$$\mathcal{E}(A \mid D) = 9 \qquad \mathcal{E}(D \mid D) = 1 \qquad \mathcal{E}(F \mid D) = 5 - \Delta$$
$$\mathcal{E}(A \mid F) = 5 \qquad \mathcal{E}(D \mid F) = 5 \qquad \mathcal{E}(F \mid F) = 5 - \Delta$$

So the recommendation structure of DICING WITH DEATH is as shown in figure 4. CDT will never recommend following the die, no matter your act probabilities.

Even though every available act gives you the news that you've fallen short of doing the most possible to improve the world—even though no available act is *perfectly* choiceworthy—both going to Aleppo and going to Damascus carry the news that you have fallen very far short of doing the most possible to improve things; whereas, following the die carries the news that you've only fallen short by a pittance, Δ.

$$\mathscr{C}_{\text{MIN}}(A) = -8 \qquad \mathscr{C}_{\text{MIN}}(D) = -8 \qquad \mathscr{C}_{\text{MIN}}(F) = -\Delta$$

So, unlike CDT, MIN advises you to follow the die.

6 THE FRUSTRATOR

If you manage the improvement news, you will violate the principle of **Causal Dominance**,

Causal Dominance If $\{K_i\}$ is a partition such that you do not have the ability to affect which K_i is true, and if there is a *dominating* act D such that $\mathcal{V}(DK_i) > \mathcal{V}(AK_i)$, for every K_i in this partition, then the *dominated* act A is not rational.

To see why, consider the following decision problem, due to SPENCER & WELLS (2017):

> THE FRUSTRATOR
> Before you are two buttons, a white button and a black button. The white button is connected to a white box; the black button is connected to a black box. You may reach out and press either button.

	K_B	K_W
B_R	0	100
W_R	100	0
B_L	$-\Delta$	$100-\Delta$
W_L	$100-\Delta$	$-\Delta$

TABLE 4: Values for THE FRUSTRATOR.

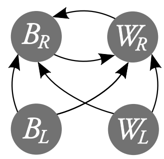

FIGURE 5: The recommendation structure of THE FRUSTRATOR

> When you do so, you will be given whatever money is in the associated box. Yesterday, *the Frustrator* made a prediction about which button you would press. If she predicted that you would press the black button, then she put £100 in the white box and left the black box empty. If she predicted that you would press the white button, then she put £100 in the black box and left the white box empty. She is rather good at these predictions—but only when you press the buttons with your right hand. (Perhaps because she bases the prediction on a brain scan of only your left hemisphere.) Given that you press a button with your right hand, she is 90% likely to have correctly predicted which button you would press. However, given that you press a button with your left hand, she is only 50% likely to have predicted correctly. You are free to press with either hand, but, if you use your left, then you must pay a pittance, Δ.

In THE FRUSTRATOR, you have four available acts: pressing black with your right hand, B_R, pressing black with your left hand, B_L, pressing white with your right hand, W_R, and pressing white with your left hand, W_L. And there are two states of nature: either the Frustrator predicted black, K_B, or the Frustrator predicted white, K_W. The values for the case are as shown in table 4.

No option in THE FRUSTRATOR recommends itself. On the assumption that you have pressed the black button with your right hand, you should expect that pressing the white button with your right will promote the most good. Similarly, on the assumption that you have pressed the white button with your right hand, you should expect that pressing the black button with your right will promote the most good. And on the assumption that you have pressed either button with your left hand, you should expect that pressing either button with your right would do better. So the recommendation structure of THE FRUSTRATOR is as shown in figure 5.

In THE FRUSTRATOR, you cannot give yourself the news that you are doing your best to improve the world. No matter what you do, you will expect yourself to have fallen short of doing the most you could to make things better. However, if you press a button with your right hand, you will expect to fall *very far* short of doing the most possible. Whereas, if you press a button with your left hand, you will only expect to fall short by a pittance, Δ.

$$\mathscr{C}_{\text{MIN}}(B_R) = -80 \quad \mathscr{C}_{\text{MIN}}(W_R) = -80 \quad \mathscr{C}_{\text{MIN}}(B_L) = -\Delta \quad \mathscr{C}_{\text{MIN}}(W_L) = -\Delta$$

So, if you manage the improvement news, then you will press a button with your left hand—either button will do.

But note that, in THE FRUSTRATOR, pressing any button with your left hand is *causally dominated* by pressing that button with your right. No matter whether the Frustrator has predicted black or white, pressing with your right hand is a pittance more valuable. Is this a problem for MIN? I do not believe so. Firstly, pressing a button with your left hand is, intuitively, more rational than pressing with your right. Indeed, SPENCER & WELLS presented THE FRUSTRATOR as a counterexample to the principle of **Causal Dominance**. Secondly, we should never have accepted the principle of **Causal Dominance** in the first place. It faces other, less controversial, counterexamples.

For instance, suppose that you are given the opportunity to play the following game: you name an integer, n, and God provides you with n days in heaven.[22] In this decision problem, every option is causally dominated. For any choice of n, $n+1$ does strictly better. So the principle of **Causal Dominance** tells us that none of your options are permissible. Assuming there are no rational dilemmas, the **Causal Dominance** principle must be false. What's gone wrong with the **Causal Dominance** principle in this case is that, though each option is dominated, the act that dominates it is *itself* rationally defective in some way. In particular, in this decision problem, every dominating option is itself dominated. Noticing this, we may wish to re-formulate the principle so that it says an option is impermissible if it is causally dominated by an option which is not itself causally dominated.

Causal Dominance (v2) If $\{K_i\}$ is a partition such that you do not have the ability to affect which K_i is true, and if there is an act D such that:

1. $\mathcal{V}(DK_i) > \mathcal{V}(AK_i)$, for every K_i, and
2. there is no act D^* such that $\mathcal{V}(D^*K_i) > \mathcal{V}(DK_i)$, for every K_i,

then A is not permissible.

Since neither B_R nor W_R is causally dominated, **Causal Dominance (v2)** would rule both B_L and W_L irrational. So MIN is inconsistent with **Causal Dominance (v2)**. But we may think that this reformulation is still too strong. One way for a dominating option to be rationally defective is for it to be dominated itself—but this is not the only way. Another way for an option to be rationally defective is for it to recommend against itself. Notice that, if an act is causally dominated, then it will recommend against itself, though not all acts which recommend against

[22] *Cf.* the *EverBetter wine* case from POLLOCK (1983).

themselves are causally dominated. So recommending against oneself is a strictly more general rational defect. Noticing this, we may favor the following, strictly weaker formulation of **Causal Dominance**:[23]

Causal Dominance (v3) If $\{K_i\}$ is a partition such that you do not have the ability to affect which K_i is true, and if there is an act D such that:

1. $\mathcal{V}(DK_i) > \mathcal{V}(AK_i)$, for every K_i, and
2. D recommends itself

then A is not permissible.

MIN will always abide by **Causal Dominance (v3)**. For no causally dominated act will recommend itself, and MIN will advise you to select an act which recommends against itself only if every other act likewise recommends against itself. But if every other act recommends against itself, then there can be no dominating act which recommends itself.

In sum: lest we tolerate rational dilemmas, all parties should agree that **Causal Dominance** is false. The principle is too strong, and must be weakened. The causal decision theorist thinks it need only be weakened to **Causal Dominance (v2)**. But those who manage the improvement news will insist it must be weakened further still. This disagreement makes a difference only with marginal cases like THE FRUSTRATOR. In decision problems like NEWCOMB'S PROBLEM, both causal dominance principles will speak with a single voice.

7 Three-Option Nicotine Gene

If a unique option recommends itself, then MIN will advise you to select it. Some have thought that this is a bad consequence of a decision rule. They believe that there are cases in which you should select an option even though it recommends against itself, and there is an alternative which recommends itself. For one such case, consider THREE-OPTION NICOTINE GENE.[24]

> THREE-OPTION NICOTINE GENE
> You are deciding whether to smoke, vape, or not use nicotine. Being a nicotine user is strongly correlated with cancer, because there is a common genetic cause of both nicotine use and cancer—call it 'the nicotine gene'. So using nicotine provides you with evidence that you will contract cancer, but it does nothing itself to promote cancer. If you have the nicotine gene, then you will enjoy using nicotine. On the other hand, if you lack the nicotine gene, then you will not enjoy using nicotine. Incidentally, there are two types of

[23] An analogous dominance principle (restricted to the case of epistemic decision theory) is defended in PETTIGREW (2016, §2.2).

[24] The case is due to Anil Gupta, and reported in EGAN (2007). I have slightly changed the case in superficial respects. See also the *Three Shells* case from SKYRMS (1984, p. 85) and the discussion of *Three Shells* in SPENCER & WELLS (2017).

$$\begin{array}{c} \\ S \\ V \\ N \end{array} \begin{array}{c} \begin{matrix} K_S & K_V & K_N \end{matrix} \\ \begin{bmatrix} 10 & 20 & 100 \\ 20 & 10 & 100 \\ 0 & 0 & 120 \end{bmatrix} \end{array}$$

TABLE 5: Values for THREE-OPTION NICOTINE GENE.

> nicotine gene—type S and type V. If you have the nicotine gene of type S, then you are likely to smoke, and though you would enjoy any kind of nicotine, you would most enjoy vaping. On the other hand, if you have the nicotine gene of type V, then you are likely to vape, and though you would enjoy any kind of nicotine, you would most enjoy smoking.

In this decision problem there are three available acts: smoking, 'S', vaping, 'V', and none of the above, 'N'. Let's suppose that there are three relevant factors beyond your control: either you have type S of the nicotine gene, 'K_S', you have type V of the nicotine gene, 'K_V', or you do not have the nicotine gene, 'K_N'. And let's suppose that, conditional on you choosing act A, the probability that state K_A obtains is 80%, and each other state has a probability of 10%. Your values for THREE-OPTION NICOTINE GENE are shown in table 5.

If you find yourself smoking, then you should expect vaping to do more good, and you should expect not using nicotine to do even less good than smoking. Similarly, if you find yourself vaping, you should expect smoking to do more good, and you should expect not using nicotine to do even less good than vaping. Finally, if you find yourself not using nicotine, then you should expect not using nicotine to do more good than any of your other options.

$$\begin{array}{lll} \mathcal{E}(S \mid S) = 20 & \mathcal{E}(V \mid S) = 27 & \mathcal{E}(N \mid S) = 12 \\ \mathcal{E}(S \mid V) = 27 & \mathcal{E}(V \mid V) = 20 & \mathcal{E}(N \mid V) = 12 \\ \mathcal{E}(S \mid N) = 83 & \mathcal{E}(V \mid N) = 83 & \mathcal{E}(N \mid N) = 96 \end{array}$$

So, while neither smoking nor vaping recommend themselves, not using nicotine at all does recommend itself (figure 6a). Therefore, while both smoking and vaping give the bad news that you'll fallen short of doing the most possible to improve things, foregoing nicotine does not.

$$\mathscr{C}_{\text{MIN}}(S) = -7 \qquad \mathscr{C}_{\text{MIN}}(V) = -7 \qquad \mathscr{C}_{\text{MIN}}(N) = 0$$

So MIN advises you to not use nicotine.

Why does EGAN think this is the wrong verdict? He writes:

> If you find yourself deciding [to smoke], one thing you know for sure is that [not using nicotine] is *not* the way to go...we can understand someone who finds herself deciding on [smoking or vaping] rethinking and doing some vacillating *between [smoking and vaping]*.

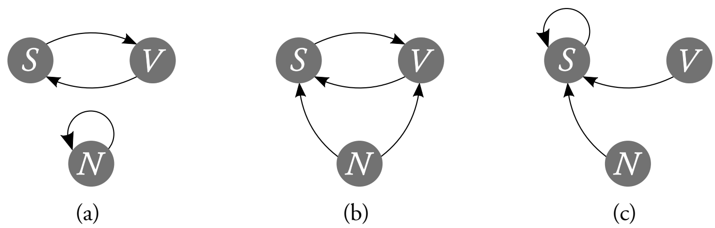

FIGURE 6: The recommendation structure of THREE-OPTION NICOTINE GENE—with (6b and 6c) and without (6a) tickles.

> What seems clearly *irrational* is for the person who finds herself deciding on either [smoking or vaping] to perform [N] on grounds of its [self-recommendation]. If she finds herself deciding on [smoking or vaping], she has excellent reason to think that [N] would be the *worst* thing to do.[25]

Here, EGAN is supposing that your initial inclinations towards smoking or vaping provide you with evidence that you have the nicotine gene. If so, then those initial inclinations are 'tickles' (recall §1). When you get the information provided by these tickles, you should condition upon it. When you do so, N will no longer recommend itself.

Suppose that, after having felt the tickle for nicotine, the probability that you have a nicotine gene is 90%, whether you have a nicotine gene is independent of whether or what kind of nicotine you use, the probability that you have the vaping (alternatively, smoking) type of gene, conditional on smoking (alternatively, vaping), is 80%, and, conditional on foregoing nicotine, you are equally likely to have either type of nicotine gene. Then, foregoing nicotine will no longer recommend itself (figure 6b). It will, moreover, give you worse news about the degree to which you are doing your best to improve things than either smoking or vaping will.

$$\mathscr{C}_{\text{MIN}}(S) = -7 \qquad \mathscr{C}_{\text{MIN}}(V) = -7 \qquad \mathscr{C}_{\text{MIN}}(N) = -10$$

So MIN will tell you to either smoke or vape. If you then begin to feel a tickle for vaping, and thereby learn that you have type V of the nicotine gene, then all acts will recommend smoking (figure 6c), and MIN will tell you to smoke.

In sum, I do not think that cases like THREE-OPTION NICOTINE GENE pose any serious trouble for MIN. They only appear to be counterexamples when you suppose that your initial inclinations provide information about the state of the world, but you forget to take that information into account before deciding what to do.

[25] EGAN (2007, p. 112, with notational changes). SKYRMS and SPENCER & WELLS make similar remarks when explaining why they think the only self-recommending act in SKYRMS's *Three Shells* case is irrational. They each assume that your initial probabilities provide information about which prediction has been made.

References

Ahmed, Arif. 2014. "Dicing with Death." *Analysis*, vol. 74 (4): 587–592. [1], [13]

Arntzenius, Frank. 2008. "No regrets, or: Edith Piaf revamps decision theory." *Erkenntnis*, vol. 68: 277–297. [7]

Barnett, David James. ms. "A Note on Egan's Counterexamples to Causal Decision Theory." [12]

Eells, Ellery. 1981. "Causality, Utility, and Decision." *Synthese*, vol. 38 (2): 295–329. [3]

—. 1982. *Rational Decision and Causality*. Cambridge University Press, Cambridge, UK. [3]

Egan, Andy. 2007. "Some Counterexamples to Causal Decision Theory." *Philosophical Review*, vol. 116 (1): 93–114. [6], [13], [17], [18], [19]

Gibbard, Allan & William L. Harper. 1978. "Counterfactuals and Two Kinds of Expected Utility." In *Foundations and Applications of Decision Theory*, A. Hooker, J.J. Leach & E.F. McClennan, editors, 125–162. D. Reidel, Dordrecht. [5], [6]

Hitchcock, Christopher. 2016. "Conditioning, Intervening, and Decision." *Synthese*, vol. 194 (4): 1157–1176. [5]

Hunter, Daniel & Reed Richter. 1978. "Counterfactuals and Newcomb's Paradox." *Synthese*, vol. 39 (2): 249–261. [1], [7]

Jeffrey, Richard. 1993. "Causality and the logic of decision." *Philosophical Topics*, vol. 21: 139–151. [2]

—. 2004. *Subjective Probability: the Real Thing*. Cambridge University Press, Cambridge, UK. [2]

Joyce, James M. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge. [5]

—. 2007. "Are Newcomb problems really decisions?" *Synthese*, vol. 156: 537–562. [2], [10]

—. 2012. "Regret and instability in causal decision theory." *Synthese*, vol. 187 (1): 123–145. [7]

—. 2018. "Deliberation and Stability in Newcomb Problems and Pseudo-Newcomb Problems." In *Newcomb's Problem*, Arif Ahmed, editor. Oxford University Press, Oxford. [10]

Lewis, David K. 1981. "Causal Decision Theory." *Australasian Journal of Philosophy*, vol. 59 (1): 5–30. [4], [5], [7], [9]

Meek, Christopher & Clark Glymour. 1994. "Conditioning and Intervening." *The British Journal for the Philosophy of Science*, vol. 45: 1001–1021. [5]

Nozick, Robert. 1969. "Newcomb's Problem and Two Principles of Choice." In *Essays in Honor of Carl G. Hempel*, Nicholas Rescher, editor, 114–146. D. Reidel, Dordrecht. [2]

Pettigrew, Richard. 2016. *Accuracy and the Laws of Credence*. Oxford University Press, Oxford. [17]

Pollock, John L. 1983. "How do you Maximize Expectation Value?" *Noûs*, vol. 17 (3): 409–421. [16]

Skyrms, Brian. 1980. *Causal Necessity*. Yale University Press, New Haven. [5]

—. 1982. "Causal Decision Theory." *Journal of Philosophy*, vol. 79 (11): 695–711. [5]

—. 1984. *Pragmatics and Empiricism*. Yale University Press, New Haven. [17], [19]

—. 1990. *The Dynamics of Rational Deliberation*. Harvard University Press, Cambridge, MA. [7]

Spencer, Jack & Ian Wells. 2017. "Why Take Both Boxes?" *Philosophy and Phenomenological Research*. [1], [14], [16], [17], [19]

Stern, Reuben. 2017. "Interventionist Decision Theory." *Synthese*, vol. 194: 4133–4153. [5]

Wedgwood, Ralph. 2013. "Gandalf's solution to the Newcomb Problem." *Synthese*, vol. 190 (14): 2643–2675. [12]