DOI: 10.1111/nous.12459

ARTICLE

NOÛS

Absolution of a Causal Decision Theorist

Melissa Fusco

Department of Philosophy, Columbia University

Correspondence
Melissa Fusco, Department of Philosophy, Columbia University.
Email: mf3095@columbia.edu

Abstract

I respond to a dilemma for Causal Decision Theory (CDT) under determinism, posed in Adam Elga's paper "Confessions of a Causal Decision Theorist". The treatment I present highlights (i) the status of laws as predictors, and (ii) the consequences of decision dependence, which arises natively out of Jeffrey Conditioning and CDT's characteristic equation.

My argument leverages decision dependence to work around a key assumption of Elga's proof: to wit, that in the two problems he presents, the CDTer must employ subjunctive-suppositional (rather than evidential) transformations of a shared prior.

"Chronic remorse, as all the moralists are agreed, is a most undesirable sentiment."
—Aldous Huxley

In Ahmed (2014) and Elga (2022), Arif Ahmed and Adam Elga present a dilemma for Causal Decision Theory (CDT) which features deterministic laws. My purpose here is to respond to that challenge on behalf of CDT. I focus on Elga's paper, "Confessions of a Causal Decision Theorist", which features a formal proof, and I aim for absolution. The treatment I present highlights (i) the status of laws as *predictors* and (ii) the consequences of *decision dependence* (Gibbard & Harper, 1978; Skyrms, 1990).

1 | THE SETUP

For expository purposes, it will be helpful to lay out both CDT and its main rival, Evidential Decision Theory (EDT). Where $\{A\}$ is a set of available actions, $\{S\}$ a set of states, and $V(\cdot)$ a

value function on possible worlds, lifted to propositions in the usual way,[1] the characterizing equations of causal and evidential expected utility are:

$$CEU(A) = \sum_S P^A(S)V(AS)$$
$$EEU(A) = \sum_S P(S \mid A)V(AS) \tag{1}$$

In Elga's terminology, the two probability functions (1) features, $P^A(\cdot)$ and $P(\cdot \mid A)$, are each the result of starting with a prior $P(\cdot)$ and *supposing* that (one chooses) A. Supposing A as such, which I'll notate $P^{[+A]}(\cdot)$, entails that $P^{[+A]}(\cdot)$ is a probability function—obeying the relevant axioms—and that $P^{[+A]}(A) = 1$. While both *imaging* (in the CEU equation) and *conditioning* (in the EEU equation) satisfy these constraints, each otherwise carries out supposition in a different way.

It is widely accepted that there is an important connection between EDT and learning. This comes out in the EDT-CDT dialectic in two steps. The first is the norm of Conditionalization, interpreted diachronically: where $P(\cdot)$ is an agent's credence function at t and $P^+(\cdot)$ is her credence function at a later time t^+, a host of arguments support the claim that an agent should use conditional probabilities, which in turn are defined by the Ratio Formula,[2] as a guide to belief revision:

> **Conditionalization**. If an agent learns exactly E between times t and t^+, she should adopt $P^+(\cdot) = P(\cdot \mid E)$.

The second step is immediate from equation (1): the EDTer's method of supposition is *also* guided by conditional probabilities. In this sense, an EDTer recommends estimating the utility of $A \in \{A\}$ as if you were to learn (you did) A.

In the standard dialectic, the CDTer denies this (Table 1). (S)he estimates utility under a *sui generis* attitude, *subjunctive supposition*, whose formal analogue at the level of credence of is imaging.

TABLE 1 Two Types of Supposition, and their entourages.

	conditioning	imaging
transformation of prior $P(S)$	$P(S \mid A)$	$P^A(S)$
attitude	learning	bringing about
expected utility expression	$\sum_S P(S \mid A)V(AS)$	$\sum_S P^A(S)V(AS)$
accompanying decision theory	evidential	causal

Loosely speaking, imaging separates *doing A* into evidential and causal components, and shifts probability only according to the latter. For example, if I *learn* that I have a Rolex instead of a Timex, that's good evidence that I'm rich, rather than poor (since typically, only rich people buy

[1] In this paper I will follow Jeffrey (1983) and subsequent literature in treating acts and states as propositions (sets of possible worlds) that are closed under the usual Boolean operations. Hence $V(B)$ is defined just in case, $\forall w, w' \in B, V(w) = V(w')$. In that case, $V(B) = V(w)$ for arbitrary $w \in B$.

[2] viz., the formula $P(B \mid A) := \frac{P(BA)}{P(A)}$. NB this formulation entails that conditional probabilities are undefined when $P(A) = 0$. Treatments that extend the definition to the $P(A) = 0$ case exist, but will not be relevant to this paper.

Rolexes.) But if I *bring it about* that I have a Rolex, rather than a Timex, that *causes* me to be (more) poor—because Rolexes cost a lot more than Timexes. So from a position of ignorance about my own wealth, $P^{\text{Rolex}}(\text{rich})$ is lower than the prior $P(\text{rich})$, but $P(\text{rich} \mid \text{Rolex})$ is higher.

Only one formal feature of imaging—its key feature—will be of interest to us in what follows.[3] It is that imaging your credence function P on A doesn't change the value assigned to $S \in \{S\}$ if your evidence also incorporates the claim that $S \in \{S\}$ are causally independent of $A \in \{A\}$.

> **Constraint on Imaging**. If the agent is certain that S is causally independent of A for any $A \in \{A\}$, then $P^A(S) = P(S)$.

2 | TWO BETS

Elga's challenge to CDT is very general. He argues that *no* suppositional operation can both (i) give us the intuitively correct answer in a betting problem about natural laws, and (ii) two-box in a Newcomb Problem whose payoffs are tied to the truth of the same set of laws. Of course, this is a particular problem for the causal decision theorist, since she is committed to two-boxing.

2.1 | The first bet, and a (possibly deterministic) Predictor

Elga's first betting problem (following closely on Ahmed, 2013, 2014) unfolds as follows. Consider a batch of deterministic laws—a scientific theory—abbreviated D, of which we suppose you to be quite confident. In the first betting problem, you are simply asked to bet either for or against D.

> **Problem 1**. Where P is you subjective credence function, $P(D) \gg 1/2$. Your utility is linear in dollars. Your value function v_1 is:

v_1	D	$\overline{D}$
A_1	k	0
A_2	0	k

Bet 1.

> A_1 is raising your hand. A_2 is not raising your hand. □

Elga's intuition is that A_1 is the right choice in Problem 1.

While I invite you to share this intuition, it does not follow straightaway from either version of Equation (1). It does not follow straightaway because $P(D) \gg 1/2$ does not entail that $P^A(D) \gg 1/2$ or that $P(D \mid A) \gg 1/2$ for $A \in \{A_1, A_2\}$. That is: you prior confidence that D is more likely

[3] There are many formal treatments of imaging in the literature, and I think it is fair to say that it is an open question—despite the case for (affirmative) closure made in Lewis (1981)—whether they are all equivalent. A classic account comes from Gärdenfors (1982) by way of Lewis (1976); this in turn owes a debt to Stalnaker (1968)'s semantics for conditionals. Another treatment of $P^A(\cdot)$, via a reduction to act-conditional chance, is due to Skyrms (1981). More recently, Pearl (2000) has analyzed imaging on A as conditioning on a related argument $\mathsf{do}(A)$, which is in turn understood in terms of causal models.

than $\overline{D}$ doesn't entail you hold D to be more likely than $\overline{D}$ *on the supposition that* A_1 (or A_2). It is the latter question(s) that—according to either form of suppositional decision theory—matter for calculating (and thus comparing) the expected utility of A_1 and A_2.[4] If you are a CDTer who nevertheless favors A_1 over A_2 in Problem 1, this suggests that you consider your available actions to make no subjunctive suppositional difference to whether D is true.

Since $D/\overline{D}$ is, by stipulation, the only outcome your value function v_1 is sensitive to in Problem 1, we note that the two-column payoff table in Bet 1 is equivalent to the following four-column payoff table, where the $D/\overline{D}$ outcomes are split along any further distinction, $C/\overline{C}$ (Table 2).

TABLE 2 Bet 1 extended across $\{C, \overline{C}\}$.

v_1	$(C \wedge D)$	$(C \wedge \overline{D})$	$(\overline{C} \wedge D)$	$(\overline{C} \wedge \overline{D})$
A_1	k	0	k	0
A_2	0	k	0	k

Thinking of Problem 1 this way will become important later, when we "splice" the bet on D with a bet on (a particular) C.

We said above that you are highly confident in D. As such, D is your leading candidate for a certain (contingent) role—the *true law* role. What, exactly, is this? Perhaps—in addition to issuing true predictions given appropriately specified initial conditions—the laws must support counterfactuals (Goodman, 1965; Maudlin, 2004) and interface in a certain way with our inductive practices (Goodman *op. cit.*); perhaps they also must in some way *necessitate* their outcomes (Armstrong 1983).

I will make a significant—but, I think, reasonable—assumption in this mode. I assume that you are sure that *something* (even if it turns out *not* to be D) plays the true law role with respect to the world you occupy. Your relationship to the law role is thus akin to an ideally rational agent's relationship to the role of objective chance (c.f. Lewis, 1971). A rational agent's uncertainty in respect of Lewis's Principal Principle concerns, not whether chance exists, but *which* probability function *plays* the chance role at the world she inhabits. In parallel fashion, you are sure the laws exist, but you are not sure which package of functions *plays* that role. D is your leading candidate.

Moreover, it is a natural(istic) thought that the laws' predictions include your own, presently available actions $A \in \{A\}$. I will write the prediction that you will A as

$$\triangle A$$

In keeping with the stipulation that D, along with its confirmation-theoretic rivals, *could* operate like a deterministic physical theory, I assume that these predictions depend on some factive input—a specification of *initial conditions*.[5]

We will dramatize the situation as follows.

> **Add-on: a Predictor**. You are certain, in Problem 1 and the Problems to follow, that you are being observed by a peerlessly accurate predictor who uses the laws to

[4] To see this, consider a parallel case in which A_1 is raising your hand *faster than the speed of light*.

[5] Given what we know about physical theories, it seems that these would have to be specific to an extreme degree, as many authors have noted (Lewis, 1971; Latham, 1987; Albert, 2000; Dorr, 2016).

make a prediction about your choice. Since you are highly confident that the laws are D, your high $P(D)$ entails high confidence that the predictor's predictions *just are* D's predictions.

We make this assumption because we want to separate two dimensions of your subjective ignorance regarding the laws:

(a) you don't know *whether D is true* at the world you occupy. (though you're highly confident it is.)
(b) you don't know *what D predicts* about $\{A\}$, when D is fed the initial conditions—the unfathomably specific historical facts *about* the world you occupy.
(you aren't highly confident about *anything* in the vicinity of this, because you haven't a clue as to what the actual initial conditions are.)

(a) and (b) are distinct aspects of your prior uncertainty.[6]

Of course, other entities make predictions too. Consider your former band-camp roommate *Dolly*, a mere human but someone who is nonetheless a highly accurate, opinionated predictor of your choices. Characters like Dolly are familiar from the literature on standard Newcomb problems, and we will occasionally appeal to (intuitions about) Dolly as a stand-in for (intuitions about) D in what is to follow. Where matters like laws of nature are concerned, we can read the definite description "the Predictor" in **Add-on** as "the predictive Laws". In other cases with the same structure (like ones that might involve Dolly), we can read "the Predictor" more generally.

We can now make a (Newcomb-)familiar assumption about the predictor's accuracy, which is that, conditional on any action A you perform, you have a high fixed credence—which we can write as $(1-\delta)$—that the predictor predicted you'd do A.

The Predictor's Strike Rate. For any $A \in \{A\}$ and predictor p:

$$P(\bigtriangleup_p A \mid A) = (1-\delta) \tag{2}$$

…where δ is a very small number, perhaps zero (we will focus on the zero case below). This formulation of the high strike rate entails that your faith in the predictor is resilient. No present change in $\{P(A) : A \in \{A\}\}$—say, via a Jeffrey shift that makes you more confident you will do

[6] The analogy with chance can continue to help us here. Suppose you are pretty sure that a particular probability function π plays the chance role: $P(Ch = \pi)$ is high. Nonetheless, you are confident as a matter of PP policy that it's *chance* that gets things right. In improbable worlds w where $\pi(\cdot)$ fails to play the chance role, $P_w(q) = \mathbb{E}(Ch(q))$ is still true, while $P_w(q) = \mathbb{E}(\pi(q)) = \pi(q)$ is false.
Indeed, we can say a bit more. Suppose, in the spirit of Lewis, that you believe that D *is* the history-relative objective chance function. That is not exactly a radical thought, since D would appear to provide a specification of the (extremal) objective chances once it is fed the initial conditions. Then $\bigtriangleup\phi$ is equivalent, at t, to $Ch(\phi \mid H_t) = 1$. Hence fealty to PP will entail much of what I say here. Simplifying somewhat:

$P_t(\neg A \mid Ch(\neg A \mid H_t) = 1) = 1$	Principal Principle
$P_t(\neg A \mid \bigtriangleup\neg A) = 1$	The proposed equivalence
$P_t(\neg A \mid \neg \bigtriangleup A) = 1$	D is deterministic w.r.t. history propositions H: $(\bigtriangleup\neg\phi) \equiv (\neg \bigtriangleup \phi)$
$P_t(\bigtriangleup A \mid A) = 1$	Probabilistic contraposition

…where the last equation states the predictor's strike rate in the main text, for the case where $\delta = 0$.

A_1 rather than A_2—would, from your perspective, serve to disconfirm your conviction that the predictor got it right.[7]

This concludes our detour through the laws. So far as Elga's intuitions are concerned, nothing interesting has yet happened. To recap: you are very confident that the laws of nature are D ($P(D) \gg 1/2$), and you are asked to either accept (by choosing A_1) a bet that pays \$$k$ iff they *are* D, or (by choosing A_2) a bet that pays \$$k$ iff they *aren't* D. It seems you should choose A_1. Moreover: you're very confident there is an excellent predictor around, one who very likely uses D itself to predict your actions. It follows that if you *do* do A_1 ($/A_2$), you should be pretty confident that the predictor predicted you would.

2.2 | The Second Bet

Elga's second decision problem is a Newcomb Problem.

> **Problem 2**. Once again, A_1 is raising your hand, and A_2 is not raising your hand. H is a proposition about the past, over which you are certain you have no causal powers. Your value function v_2, where m (think: *a million*) and t (think: *a thousand*) are positive, is:

v_2	H	$\overline{H}$
A_1	m	0
A_2	$m+t$	$0+t$

Bet 2

> As the payoff matrix illustrates, doing A_2 gains you t whether or not H is true.[8] But hark: H is a special proposition. It is the maximally inclusive specification of the initial conditions which, under D, entail A_1 (viz., $H = \bigtriangleup_D A_1$). (Ahmed, 2014, pg. 120; Elga, 2022, pg. 206) □

[7] This formulation of the high strike rate entails, for example, that you *aren't* highly confident in the predictor simply because: (i) you are sure you will do (some particular) A and thus (ii) sure that the predictor predicted A. We see the latter situation in Robert Stalnaker's discussion of his first brush with the Newcomb Problem:

> …let me report on my reaction to the Newcomb Problem when I first encountered it more than forty years ago. I thought, "It's obvious that one should take both boxes, and **there is no puzzle about how a predictor could get it right 90% of the time, since surely almost everyone will make this choice, so the predictor can safely predict that everyone will**, and be about 90% accurate" (2018, emphasis added)

…Equation (2) rules out this extensional construal of the predictor's accuracy, because it requires you to be confident that the predictor gets it right conditional on *any* act performed. In context, this would have required Young Stalnaker to be confident that the predictor was correct about what, say, Fred did, *even if* he [Young Stalnaker] conditioned on the fact that Fred *one-boxed*.

[8] To make the contrast with Evidential Decision Theory (EDT), which treats supposition as conditionalization, we can stipulate that $P(H \mid A_1) - P(H \mid A_2) > \frac{[1-P(H|A_2)]t}{m}$, and hence that EDT would endorse A_1 *instead* of A_2 in Problem 2. I will presuppose as much in what follows, though nothing important will hinge on it. Careful readers of Ahmed will note that I have followed Elga in inverting the identities of A_1 and A_2 (for ease of a combined exposition with Newcomb's Problem).

Causal decision theorists generally accept the verdict that A_2 is the right choice in Problem 2: two-boxing in a Newcomb Problem is, after all, a *sine qua non* of CDT.[9]

In **Problem 2** there is, by stipulation, only one distinction to which the value function v_2 is sensitive: H vs. $\overline{H}$. So once again, we note that the payoff matrix in Bet 2 is equivalent to the one below, for schematic $\{E, \overline{E}\}$ (Table 3):

TABLE 3 Bet 2, extended across $\{E, \overline{E}\}$.

v_2	$(H \wedge E)$	$(H \wedge \overline{E})$	$(\overline{H} \wedge E)$	$(\overline{H} \wedge \overline{E})$
A_1	m	m	0	0
A_2	$m+t$	$m+t$	$0+t$	$0+t$

E could be anything—so, in particular, it could be D itself. And indeed, this is the final form of the problem we will consider (c.f. Elga, pg. 208):

TABLE 4 Bet 2 (final), with $E = D$.

v_2	$(H \wedge D)$	$(H \wedge \overline{D})$	$(\overline{H} \wedge D)$	$(\overline{H} \wedge \overline{D})$
A_1	m	m	0	0
A_2	$m+t$	$m+t$	$0+t$	$0+t$

Elga (and Ahmed's) specification of H plugs nicely into our view of laws as predictors. In the vignette, your commitment to the predictor's accuracy commits you (we will assume) to $P(\bigtriangleup A_1 \mid A_1) = 1$, which entails $P(\bigtriangleup A_2 \wedge A_1) = 0$. These extreme values continue to hold when the predictor is D, so $P_D(\bigtriangleup A_2 \wedge A_1) = 0$. This entails $P(D \wedge \overline{H} \wedge A_1) = 0$. A similar argument shows $P(D \wedge H \wedge A_2) = 0$. These "$P$-zero" propositions are key to the dialectic to follow.

2.3 | Elga's Dilemma

Tables 2 and 4 above delineate two value functions, v_1 and v_2. We implicitly assumed these value functions could be paired with any prior—a fortiori, they could both be paired with the aforementioned prior $P(\cdot)$ to calculate the expected utility of A_1 and A_2.

Recall that $P(D)$ is very high. In addition, H is playing its special role, which underwrites credence zero in the propositions $(D \wedge H \wedge A_2)$ and $(D \wedge \overline{H} \wedge A_1)$. As we noted in §1, however, to connect this all to (suppositional) expected utility, we must look at suppositional transformations of $P(\cdot)$, rather than $P(\cdot)$ itself. To set up his dilemma for CDT, Elga uses a table with variables for the values of $\{H, \overline{H}\} \times \{D, \overline{D}\}$ under the suppositionally transformed probability functions $P^{A_1}(\cdot)$ and $P^{A_2}(\cdot)$:

[9] I nonetheless say CDTers "generally" endorse A_2 in Problem 2 because David Lewis—a CDTer—appears to deny in Lewis (1981, §5) that this type of Newcomb problem is possible on epistemic grounds.

TABLE 5 Credence functions $P^{A_1}(\cdot)$ and $P^{A_2}(\cdot)$.

	$(H \wedge D)$	$(H \wedge \overline{D})$	$(\overline{H} \wedge D)$	$(\overline{H} \wedge \overline{D})$
P^{A_1}	a	b	x	c
P^{A_2}	y	d	e	f

We saw in §1 that $P^X(X) = 1$ for any X, and that $P^X(\cdot)$ is a probability function. Two things follow: (i) the sum of the numbers across each row of Table 5 is 1; (ii), $P^{A_1}(X) = P^{A_1}(XA_1)$ and $P^{A_2}(X) = P^{A_2}(XA_2)$ for any X.

Since Table 5 lists values for P^{A_1} and P^{A_2}, rather than P, (i) and (ii) do not directly constrain the values in the table. Elga does, however, assume an indirect constraint: that, while a, b, c, d, e, and f are all positive, the highlighted values in the table, x and y, are zero. Gloss: the propositions $(D \wedge H \wedge A_2)$ and $(D \wedge \overline{H} \wedge A_1)$ not only receive credence zero in the prior P, but they continue to receive credence zero under *any* suppositional transformation of the prior—even subjunctive supposition.

And here we arrive at last at Elga's antinomy, which I will present for CEU as set out in Equation 1.

Fact 1. *(Elga, 2022, §5). When $x = y = 0$, there is no probability function P and value functions v_1, v_2 such that, in Problem 1, $CEU(A_2) < CEU(A_1)$ and, in Problem 2, $CEU(A_1) < CEU(A_2)$.*

This means it is impossible for any P to justify A_1 in **Problem 1** and A_2 in **Problem 2**.

Proof. By the assumption that in Problem 1, $CEU(A_2) < CEU(A_1)$ (Table 2):

$k(d + f) < k(a)$	definition of CEU
$(d + f) < a$	k is a positive number

By the assumption that in Problem 2, $CEU(A_1) < CEU(A_2)$ (Table 4):

$(ma + mb) < [(m + t)(d) + te + tf]$	definition of CEU
$m(a + b - d) < t(d + e + f)$	simplifying
$m(a + b - d) < t$	because $(d + e + f) = 1$
$(a + b - d) < t/m$	m is a positive number
$(a + b - d) \leq 0$	arbitrariness of t, m ($t < m$ and positive)
$a \leq (d - b)$	arithmetic

Chaining together these two inequalities concerning a, we have: □

$(d + f) < (d - b)$	constraints on a (above)
$f < -b$	arithmetic
$f + b < 0$	arithmetic
$\perp$	f, b are positive numbers (by assumption)

3 | Elga's Dilemma and Decision Dependence

The *reductio* above assumes that P^{A_1} and P^{A_2} are (subjunctive) suppositional transformations of a shared prior, P. In the next few sections, I will argue that a well-studied version of CDT aimed at treating *decision dependence* (Gibbard & Harper, 1978; Hare & Hedden, 2016)—to wit, the *deliberational dynamics* approach of Skyrms (1990)—has the dialectical resources to block this assumption in the relevant context.

First, let's be clear on how Elga's proof leverages the shared prior assumption Figure 1. Recall that the dialectic of the paper begins with a credence function P which is subject to various constraints. You confront each of **Problem 1** and **Problem 2** with P. The two problems are linked because the same pair of acts—or at least, act-types—are available in each: A_1 and A_2 (raising/lowering your hand). These acts are the inputs to suppositional transformations of P, P^{A_1} and P^{A_2}, in terms of which the expected utility of each act in each problem is calculated. Then Elga's argument shows that opting for A_1 in Problem 1 and A_2 in Problem 2 would put jointly unsatisfiable constraints on $P^{A_1}(\cdot)$ and $P^{A_2}(\cdot)$—and thereby, working backwards, onto P itself.

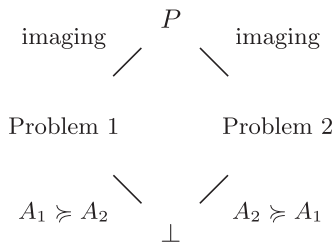


FIGURE 1 The Shared Prior.

But although you confront **Problem 1** and **Problem 2** with the same probability function, your utility functions v_1 and v_2 in these respective problems are *not* the same, nor is one a refinement of the other. They are *incompatible*. The utility function v_2 from **Problem 2** cares nothing for the $D/\overline{D}$ distinction, whereas the utility function v_1 from **Problem 1** cares *only* about this distinction.[10] That means there would be a weak link in the argument if the following were true: sometimes, when you are confronted with a decision problem which presents you with prizes, your prior P must, *before you act*, update in some way to reflect that confrontation. If that were the case, then the fact that $v_1 \neq v_2$ would *entail* that the credence function relevant to the calculation of expected utility in **Problem 1** is not the same as the credence function ultimately relevant to the expected utility calculation in **Problem 2**. And if that's right, Elga's algebraic argument, while valid, does not show what it needs to show.

[10] In more detail: according to v_1, for any $w, w' \in \overline{D}$, $v_1(w) = v_1(w')$. But Problem 2 requires that $v_2(w) \neq v_2(w')$ if w and w' are not members of the same cell of $\{H, \overline{H}\}$. So v_1 and v_2 cannot be the same function.

3.1 | Decision Dependence: A Primer

One might find the "weak link" line of argument I sketched above unattractive, for Humean reasons: it apparently denies that our credences are independent of our value functions. But I think the prospects for such an argument are quite good. Decision dependence, which falls out of the way causal decision theory handles predictor-style cases, registers a modest exception to the Humean picture: it opens up space for information about prizes to rationally alter one's credences *in one's own presently available acts*, represented by the probabilities assigned across the act-partition $\{A\}$. When these probabilities change in response to information about prizes, Jeffrey Conditioning then propagates this change across the probabilities of states, thus bringing about a corresponding change in the (causal) expected utility of $A \in \{A\}$.

Here is a classic case—perhaps *the* classic case—of decision dependence, from Gibbard and Harper (1978).

> **Death in Damascus**. You are in the desert between Damascus and Aleppo at t, when you learn that Death plans to harvest your soul. Death always follows a fixed schedule—"work[ing] from an appointment book…made up weeks in advance" (*op. cit.*, pg. 157). Your available acts are to go to Damascus (Dam) or to go to Aleppo (Alep). Your credence $P(\text{BookAlep} \mid \text{Alep})$—that Death's book says "Aleppo", given that you go there—is very high (functionally, 1), as is your subjective probability $P(\text{BookDam} \mid \text{Damascus})$—that the book says "Damascus", given that you go to Damascus. You also know that you have no causal influence over the book's contents at the time t of your deliberation.

	BookDam	BookAlep
Dam	0	10
Alep	10	0

Death in Damascus.

We stipulate that 10 is the value of surviving in either location, and 0 the value of dying. We assume the journeys themselves are costless and that, at the start of deliberation, $P(\text{BookDam}) = 0.6$ and $P(\text{BookAlep}) = 0.4$. CDT thus assigns Dam an expected value of 4 and Alep an expected value of 6.

However, in **Death in Damascus**, various events might cause you to become more confident that you are going to do one thing rather than another. Suppose you begin to walk towards Aleppo. This does not result in your (yet) becoming *certain* of any proposition $A \in \{\text{Dam}, \text{Alep}\}$, but does directly raise your confidence in Alep at the expense of Dam. The standard belief revision tool for tracking how such a shift influences the rest of your credence function is Jeffrey Conditionalization (Jeffrey, 1983).

> **Jeffrey Conditionalization**. If a learning experience directly alters an agent's prior over partition $\{E\}$ between times t and t^+, then she should adopt as a posterior
>
> $$P^+(\cdot) = \sum_E P^+(E)P(\cdot \mid E) \tag{3}$$

According to Jeffrey Conditionalization, as a self-aware agent executes a particular act A_i in **Death in Damascus**, (i) $P(\text{Book}A_i \mid A_i) \approx 1$ and (ii) $P(\text{Book}A_i \mid A_j) \approx 0$ for $A_i, A_{j\neq i} \in \{\text{Dam},$

Alep} guide revision. It follows that the causal expected utility of each option at time t is inversely proportional to its probability at t. By equation (1):

$$\begin{aligned} CEU_t(A_i) &= P_t^{A_i}(\mathsf{Book}A_i)V(\mathsf{Book}A_i \wedge A_i) + P_t^{A_i}(\mathsf{Book}A_j)V(\mathsf{Book}A_j \wedge A_i) \\ &= P_t^{A_i}(\mathsf{Book}A_i) \times 0 + P_t^{A_i}(\mathsf{Book}A_j) \times 10 \\ &= P_t^{A_i}(\mathsf{Book}A_j) \times 10 \end{aligned}$$

Since $\mathsf{Book}A_i$ is causally independent of $\{A_i, A_j\}$, $P_t^{A_i}(\mathsf{Book}A_i) = P_t(\mathsf{Book}A_i)$ by the **Constraint on Imaging** we saw in §1. Hence, applying the Law of Total Probability:

$$\begin{aligned} CEU_t(A_i) &= P_t(\mathsf{Book}A_j) \times 10 \\ &= [P_t(\mathsf{Book}A_j \mid A_i)P_t(A_i) + P_t(\mathsf{Book}A_j \mid A_j)P_t(A_j)] \times 10 \\ &= [(\approx 0)P_t(A_i) + (\approx 1)P_t(A_j)] \times 10 \\ &\approx P_t(A_j) \times 10 \\ &\approx [1 - P_t(A_i)] \times 10 \end{aligned}$$

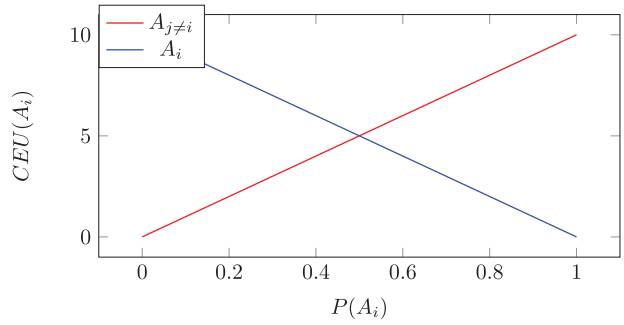


FIGURE 2 Causal Expected Utility of A_i as a function of $P(A_i)$ in **Death in Damascus**. [Color figure can be viewed at wileyonlinelibrary.com]

Adapting Hare & Hedden's slogan: in **Death in Damascus**, what you (think you) *ought* to do depends (negatively) on what you *anticipate* you'll do. No matter which destination you pick, when you get there, you will believe with certainty that you did worse than you otherwise would have (Figure 2).

The dependence of $CEU(A)$ on $P(A)$ illustrated by **Death in Damascus** is widespread for CDT. It holds, for example, in the traditional Newcomb Problem: if one "credally journeys" towards two-boxing via a Jeffrey shift, one becomes increasingly pessimistic about millionaire. However, the *comparative* facts about one-boxing and two-boxing never change: it is still always the case that the causal expected utility of two-boxing (A_2) exceeds the causal expected utility of one-boxing (A_1) by t, the (scaled) value of one thousand dollars (Figure 3). As Joyce (2012, pg. 130) emphasizes, Newcomb is unusual in this respect: although the two choices' causal expected utilities do exhibit decision dependence, the dominance of two-boxing over one-boxing at all values of P(one-box) ensures CDT's *recommendation* never depends on one's act-probabilities.

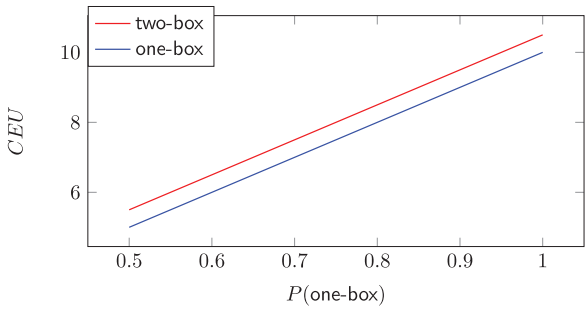


FIGURE 3 Causal Expected Utility as a function of $P(\text{one-box})$ in a Newcomb Problem. [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | A Nice Time

Decision dependence is a fact of life for CDT. Much of the literature on it is devoted to cases, like **Death in Damascus**, which are *nasty*, or unstable.[11] Our purposes, however, will be better served by focusing on the obverse phenomenon: *nice*, or self-reinforcing cases. As I will use the terminology, you are in a nice case if you have multiple (say, *two*) choices, and each is such that, if you perform it, you will believe with certainty that you did better than you otherwise would have. Once again, decision dependence makes this possible. A *nice* choice, like a *nasty* one, is in this sense a technical concept of CDT: EDT lacks the apparatus to model its distinctive subjunctive profile.

Here is an example.

> **Nice Choices in New Jersey**. You live in Hoboken, New Jersey. You are asked to choose between A_1, remaining in Hoboken, and A_2, riding a (free) bus to Secaucus. Yesterday, a very good predictor predicted where you would end up tonight in order to leave you a small sum of money in that location, and that location alone. She deposited her money like this:

	In Hoboken	in Secaucus
if she predicted Hob, she put	\$10	\$0
if she predicted Sea, she put	\$0	\$15

Your value function for this choice, v_{1a}, is thus:

v_{1a}	Pred Hoboken	Pred Secaucus
A_1 (viz., Hob)	\$10	\$0
A_2 (viz., Sea)	\$0	\$15

[11] Richter (1984), Egan (2007), Meacham (2010), Joyce (2012), Levinstein and Soares (2020), *multa inter alia*.

In **Nice Choices in New Jersey**, what you ought to do is positively correlated with what you anticipate you'll do. It is also asymmetric.[12] While (Sea ∧ \$15) is the bigger prize in **Nice Choices in New Jersey**, some priors heavily skewed towards Pred Hoboken will nonetheless assign a higher expected utility to Hob than to Sea; one example is a prior according to which P(Hoboken) = P(Pred Hoboken) = .75 (Figure 4). Since the x-axis in Figure 4 plots P(Sea) as the latter increases towards certainty, we can see the graph as showing us what would happen to the subjective expected utilities of each option if the agent were to undergo a Jeffrey Shift that pushes her credence in Sea from skepticism towards certainty—for example, what would happen if, with self-awareness, she journeyed towards Secaucus.

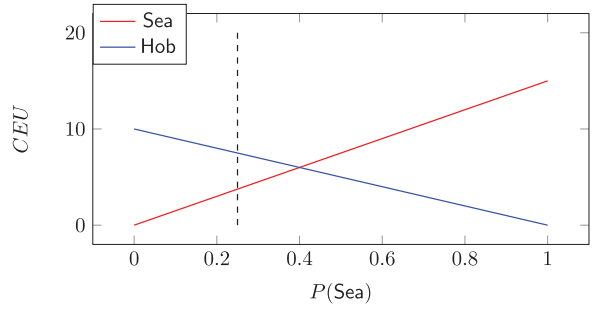


FIGURE 4 Decision Dependence in **Nice Choices in New Jersey**. [Color figure can be viewed at wileyonlinelibrary.com]

The analysis I've so far sketched of **Nice Choices in New Jersey** is standard in the CDT literature. Before proceeding, though, we should do a sanity check: ensuring that there isn't any hidden inconsistency in having a prior which skews towards Hoboken and Pred Hoboken, and yet is presented with the value function v_{1a} we find in **Nice Choices in New Jersey**. Why *would* one ever initially be more confident in Pred Hoboken than Pred Secaucus in such a case? And second: even if one *was* initially in such a state, shouldn't one's credences evolve as a result of what's been learned about v_{1a}?

Forestalling the question of evolution for a moment, it does *seem* possible to be in this situation. Your priors over your own acts could be anything. For example, in **Nice Choices in New Jersey** your priors might skew towards Pred Hoboken from the sheer weight of past experience. Suppose, for example, that Dolly helps you with the rent each month by leaving cash envelopes for you under back tables at Dunkin Donuts. You've been collecting \$10/month from her this way *for years*—and specifically, you've been accomplishing this by going to the Dunkin Donuts in *Hoboken* for years. You know Dolly to be an astonishingly accurate predictor of your movements: indeed, she uses a Laplace-o-meter to decide where to leave the money. Hence, you have very high *initial* credence that the money is in Hoboken as usual this month.

So much for the prior. However, today you suddenly learn something new, which effectively places you in the decision problem **Nice Choices in New Jersey** describes. For example, you could simply learn that, *all this time*, Dolly has been following the algorithm listed above (repeated):

[12] These cases are mashups of two of Hare & Hedden (2016)'s cases, the *Asymmetrically Nasty Demon* (pg 614) and the *Nice Demon* (pg 606; preceded by an identical case with the same name in Skyrms, 1982, pg. 706). See also Lewis (1981, §11)'s *Hunter-Richter Problem*. Cases with a similar "asymmetrically nice" structure can also be found in the ethics literature: see, for example, Harman (2009).

	In Hoboken	in Secaucus
if she predicted Hob, she put	\$10	\$0
if she predicted Sea, she put	\$0	\$15

.

The algorithm entails that if Dolly's device predicted Sea instead of Hob this time around, there's currently \$15 for you in Secaucus, and *nothing* in Hoboken! (Naturally, Dolly's device also makes predictions about whether you made the discovery about her algorithm that you just in fact made.) So the question now is: do you "stick" with your prior, which still—recall Figure 4—indicates that the expected utility of Hob exceeds the expected utility of Sea, or not? And if you do not, how should the prior evolve: what *credal dynamics* does rationality require, given the information you now possess?

4 | DELIBERATIONAL DYNAMICS

Skyrms (1990) provides a family of standard answers to the dynamical question, which are widely endorsed in the literature on decision dependence.[13] To understand his view, it will help to extend the notion of causal expected utility to the *status quo* (henceforth '*SQ*'). This is your expected value of *CEU* over the partition $\{A\}$—or, in Skyrms's terminology, the expected utility of the *vector* consisting of available acts A and their current probabilities according to P.

$$CEU(SQ) = \sum_{A} P(A)CEU(A) \tag{4}$$

Figure 3 shows $CEU(SQ)$ for **Nice Choices in New Jersey** as a function of $P(\mathsf{Sea})$. A bit of calculation shows this function is quadratic, reflecting a contour by which causal expected utility dips before rising as $P(\mathsf{Sea})$ increases.[14] In a game theoretic context, $CEU(SQ)$ also represents the causal expected utility of the *mixed act* consisting of performing each $A \in \{A\}$ with chance $P(A)$—a point we will return to below.

[13] See, for example, Joyce (2007, 2012), Lauro & Huttegger (2022), Harper (2022).

[14] By Equation (4):

$$CEU_t(SQ) = P_t(Sea)CEU_t(Sea) + P_t(Hob)CEU_t(Hob)$$

We know from calculations analogous to those in **Death in Damascus** that:

$$CEU_t(Sea) = P_t(Sea) \cdot 15$$

$$CEU_t(Hob) = P_t(Hob) \cdot 10$$

Hence:

$$CEU_t(SQ) = P_t(Sea)^2 \cdot 15 + P_t(Hob)^2 \cdot 10$$

$$= P_t(Sea)^2 \cdot 15 + [1 - P_t(Sea)]^2 \cdot 10$$

this is the equation graphed in Figure 5.

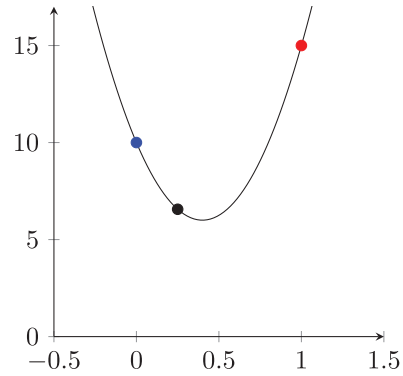


FIGURE 5 $CEU(SQ)$ as a function of $P(\mathsf{Sea})$. [Color figure can be viewed at wileyonlinelibrary.com]

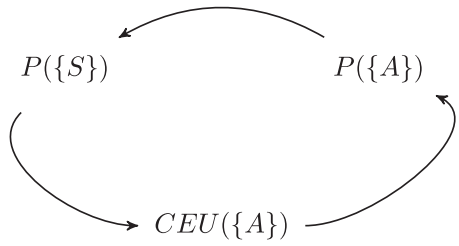


FIGURE 6 Feedback loop.

Skyrms's dynamics uses a preliminary (viz., prior-relative) calculation of $CEU(SQ)$ as input to a dynamical rule that incrementally *increases* the probability of acts A_i such that $CEU(A_i) > CEU(SQ)$ and *decreases* the probability of acts A_j such that $CEU(A_j) < CEU(SQ)$ (see the Appendix for the particular rules he considers.) The theory thus introduces a feedback loop between causal expected utility, probabilities over acts, and—via Jeffrey Conditioning—probabilities over dependency hypotheses (Figure 6). A series of theorems (for example, Skyrms, 2022) show that these processes terminate at some P_{final}, such that additional cycling through the dynamics does not alter act-probabilities or act-utilities any further.

Skyrms CDT sees asymmetric nice cases—and other instances of decision dependence—as cases where your act-credences are constrained to seek the *local* good; you do, in that sense, remain in the grip of the starting point provided by your prior, even if you do not remain *at* your prior. In the context of **Nice Choices in New Jersey**, this means Skyrmsian dynamics will move the agent's $P(\mathsf{Sea})$ to the *left* from the black dot in Figure 5—that is, towards the local maximum 10 at the posterior $P_{\text{final}}(\mathsf{Hob}) = 1$.

As I discuss below, there is some appeal to a version of CDT that differs from Skyrms on precisely this point. Such an view follows "if you would have peace, prepare for war"-type reasoning: if you would maximize (causal) expected utility in the long run, be prepared to minimize it in the short run. In the context of our example in Figure 5, this is the difference between moving towards the left equilibrium point (in blue) and the right equilibrium point (in red).

4.1 | Stacked and Sequential Problems

For the moment, though, I want to put the differences between Skyrms 1990 and myself to one side. Our mutual interest is in blocking Elga's argument. To see how, consider what happens when we add a second decision problem—a Newcomb problem—into the mix.

> **Newcomb in New Jersey**. As before, you live in Hoboken and must choose between A_1, remaining in Hoboken, and A_2, riding the bus to Secaucus. This time, though, a second, wealthier predictor (who also has a Laplace-o-meter) has decided to offer you some money! Her plan was as follows: last night, she made a prediction about whether you would go to Secaucus or Hoboken, and she put $20 in your bank account iff she predicted you would remain in Hoboken. Additionally, however, she ensured that there is $5 in a transparent box for you at the Dunkin' Donuts in Secaucus, which you can have just in case you travel there instead.

Your payoff matrix *for this predictor alone* is thus the following value function, v_{2a}, which exhibits the familiar Newcomb asymmetries:

v_{2a}	Pred Hoboken	Pred Secaucus
A_1 (viz., Hob)	\$20	\$0
A_2 (viz., Sea)	\$25	\$5

Skyrmsian decision theory—like any form of CDT—is a two-boxing decision theory. When **Newcomb in New Jersey** is considered *on its own*, then, Sea (viz., A_2) is the right choice relative to any prior P (recall Figure 3) for any of the dynamical rules Skyrms considers. If you are rational and follow through, then as you execute A_2—the equivalent of riding the bus to Secaucus, dutifully updating along the way—your credence in Pred Sea rises.

Thus Skyrmsian CDT, when it considers each of the two New Jersey problems in isolation, recommends different actions: A_1 (viz., Hob) in the first case and A_2 (viz., Sea) in the second. So what? The example shows how Skyrms can reject the following inference:

P1. No rational agent can choose A_2 in **Newcomb in New Jersey** and A_1 in **Nice Choices in New Jersey** while calculating causal expected utilities with respect to the same probability function P.

to

C1. No agent with prior P can choose A_1 in **Nice Choices in New Jersey** and A_2 in **Newcomb in New Jersey**.

P1 is true, because of decision dynamics: *once you have cycled through* Skyrms's feedback loop, $P_{\text{final}}(\text{Pred Hob})$ in **Newcomb in New Jersey** $\neq P_{\text{final}}(\text{Pred Hob})$ in **Nice Choices in New Jersey**. This *must* be the case, because in **Newcomb in New Jersey** $P(\text{Pred Hob})$ approaches zero,

and relative to that distribution, A_1 does *not* maximize *CEU* in **Nice Choices in New Jersey**. But C1 is false, because it concerns the *prior*, not the (Skyrmsian) posterior: as we just saw, an agent with prior $P(\text{Pred Hoboken}) = .75$ and $P(\text{Pred Secaucus}) = .25$ will indeed, on Skyrms's theory, choose A_1 in **Nice Choices in New Jersey** and A_2 in **Newcomb in New Jersey**. So here is the diagnosis of Elga's argument: P1 is the equivalent of what his formal proof shows. But to indict CDT, it is C1 that is needed.

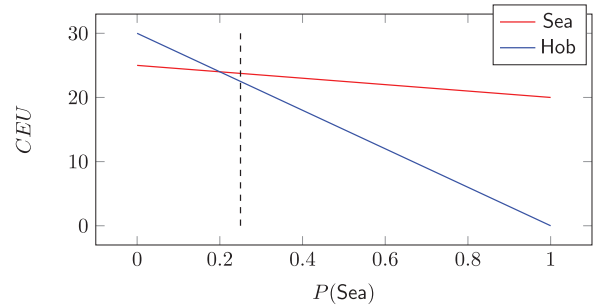
For completeness—and before circling back to my disagreement with Skyrms—it is worth checking that the reasoning we just looked at is consistent with every interpretation of Elga's argument. What we just considered was a case where an agent with prior P can either evolve towards greater certainty in Hob (in **Nice Choices in New Jersey**) or a greater certainty in Sea (in **Newcomb in New Jersey**) as a rational response, respectively, to (i) facing the first problem alone, or (ii) facing the second problem alone. There is, of course, another possible way of considering the dialectical force of the two problems—one could stipulate that you are facing them *simultaneously* at the same world. This way of interpreting the problem goes with taking Elga's talk of the *same* two acts, A_1 and A_2, being available in **Problem 1** and **Problem 2** as *token* identity, rather than type identity. Then our response does not really work: surely I cannot have one "Skyrms-evolved" credence function P (according to which Hob is highly likely) and some *other* sort of "Skyrms-evolved" credence function $P^\star$ (according to which Sea is highly likely) in the same world at the *same time*!

…This is certainly true, though I do not think it will remedy Elga's argument. If you confront **Nice Choices in New Jersey** and **Newcomb in New Jersey** simultaneously, you face a different problem, with a single, novel utility function, $v^*(\cdot) = v_{1a}(\cdot) + v_{2a}(\cdot)$. This is illustrated in Figure 7. (Naturally, you expect both predictors to be correct, which is why only two columns are needed to characterize the value function v^*.)

Stacked Predictors in New Jersey.

v^*	Pred Hoboken	Pred Secaucus
A_1 (viz., Hob)	\$30	\$0
A_2 (viz., Sea)	\$25	\$20

FIGURE 7 Decision Dependence in the "Stacked" problem. [Color figure can be viewed at wileyonlinelibrary.com]



The intersection point in Figure 7, where $CEU(\text{Sea}) = CEU(\text{Hob})$, is at $P(\text{Sea}) = .2$. Since, by hypothesis, your prior P assigns 25% to Sea, Skyrmsian dynamics will favor riding to Secaucus in

this problem. As the payoff matrix makes plain, **Stacked Predictors** is *not* a Newcomb Problem—the second row of the payoff matrix does not dominate the first. So the CDTer faces no particular conflict of loyalties here. That is to say: it is no betrayal of CDT's characteristic commitments vis-à-vis Newcomb Problems for the theory to recommend either A_1 *or* A_2 in **Stacked Problems in New Jersey**.

4.2 | Dynamics Unbound

Above, I signaled affection for a form of CDT with deliberational dynamics that, unlike the versions explored by Skyrms *et. al.*, is not constrained to mechanisms of expected utility maximization that are local with respect to the agent's prior. Such a view would permit an increase in the probability of acts which *suppress* the causal expected utility of the status quo in the *short* term, but *increase* it the *long* term.

In the remainder of this section, I take up the point of view of such a form of CDT. In asymmetric nice cases like **Nice Choices in New Jersey**, this opens up new deliberational pathways that seem practically advantageous and epistemically innocent. At a first pass, this theory simply takes as input the entire space of Jeffrey-conditionalizations available from the prior—the equivalent of the whole curve in Figure 5. It is *permissive*, allowing the agent to e.g. ride all the way to Secaucus in **Nice Choices in New Jersey** in light of the fact that, once she gets close, she converges upon confidence that her choice maximizes causal expected utility after all.

(For those EDTers keeping score: Skyrmsian dynamics relative to the Hob-skewed prior *reverses* the recommendations of EDT for *both* **Nice Choices in New Jersey** and **Newcomb in New Jersey**. The latter is obvious; the former holds because EDT simply "pre-conditionalizes" on each available act, assigning Hob and expected utility of 10 and Sea an expected utility of 15. The alternative to Skyrmsian dynamics on offer here does the latter but not the former; it permits, without requiring, Sea in **Nice Choices in New Jersey**.)

With this much of the positive view on the table, we return to betting on the laws. Recall that in our gloss on **Problem 1**, we were aware that some predictor was present and using the true laws to predict our actions; while we weren't certain the laws were D, we were very confident of it. The payouts for A_1 and A_2 were:

TABLE 3 (repeated): Bet 1.

v_2	$(H \wedge D)$	$(H \wedge \overline{D})$	$(\overline{H} \wedge D)$	$(\overline{H} \wedge \overline{D})$
A_1	m	m	0	0
A_2	$m+t$	$m+t$	$0+t$	$0+t$

For the sake of argument, I granted in §2 that you should do A_1 if you face this choice by itself and $k > 0$. But since H is, given D, equivalent to the prediction that you would choose A_1, your credence across $\{HD, \overline{H}D\}$ is sensitive here to your act-probabilities, which are in turn (via the feedback loop) sensitive to your utilities. It might be that *new* information about prizes substantively increases your credence that the predictor predicted A_2 instead. This will have knock-on effects on CEU and, through Jeffrey Conditioning, would ramify into your credences across your own actions.

Problem 2 (repeated here) is a standard Newcomb problem, where the good news hypothesis is H. In this problem dynamically permissive CDT, like Skyrmsian CDT, says that A_2 is always better

TABLE 2 (repeated): Bet 2.

v_2	$(H \wedge D)$	$(H \wedge \overline{D})$	$(\overline{H} \wedge D)$	$(\overline{H} \wedge \overline{D})$
A_1	m	m	0	0
A_2	$m+t$	$m+t$	$0+t$	$0+t$

than A_1 if $m, t > 0$, no matter what one's act-probabilities are or how they evolve (as illustrated in Figure 3). So we get the two verdicts Elga wants regarding the choices the theory makes in each problem.

Now for Elga's impossibility proof. My permissive causalist—seconded by Skyrms et al—argues that although the priors with which you face Problems 1 and 2 start out the same, they need not *stay* the same; they can permissibly evolve toward A_1 (and thereby, towards H) in the first problem and toward A_2 (and thereby, towards $\neg H$) in the second problem, as a result of the very prizes you have been offered. That means that, *by the time you act*, you aren't working with the same credence function across the two problems. Once again, the analogue of P1 holds: by Elga's proof, no rational agent can choose A_1 in **Problem 1** and A_2 in **Problem 2** while calculating causal expected utilities with respect to the same probability function P. But the analogue of C1 does not hold: it is *false*, on my theory, that an agent with prior P therefore cannot choose A_1 in **Problem 1** and A_2 in **Problem 2**.

As we saw above, a proponent of Elga's argument can reply by stipulating that the two problems are faced at the same world and time—hence, with summed payoffs. This is the analogue of **Stacked Problems in New Jersey** (Table 6).

Stacked Problems 1 and 2. You face **Problem 1** and **Problem 2** simultaneously.

TABLE 6 Bets 1 & 2, payouts summed.

$v_1 + v_2$	$(H \wedge D)$	$(H \wedge \overline{D})$	$(\overline{H} \wedge D)$	$(\overline{H} \wedge \overline{D})$
A_1	$m+k$	m	k	0
A_2	$m+t$	$m+t+k$	$0+t$	$0+t+k$

Stacked Problems 1 and 2 is a decision problem in good stead. Once again, though, the thing to note about Table 6, is that, unless $t > k$, it is *not* a Newcomb Problem: the second row does not (strictly) dominate the first. So unless $t > k$, the *sine qua non* of CDT—that thou shalt two-box in a Newcomb Problem—does not constrain the **Stacked Problem**.

Instead, permissive CDT will analyze this decision problem as a series of tradeoffs, as we can illustrate by considering (silly, low-stakes) proxies for m, t, and k. To narrate: once again, you believe yourself to be observed by an infallible predictor. You are very (but not *absolutely*) sure that the observer is Dolly (D). Focusing for the moment on this possibility (the odd columns of Table 6), we see that: should you do A_1, you will win a *Pabst Blue Ribbon* (worth k) iff the predictor *is* Dolly. Depending on how much you value cheap beer, then, this may give you a reason to do A_1. (It could be a very weak reason.) And if you do so—since A_1 entails one-boxing—you will also sacrifice t, the contents of a tiny Newcomb box. On the other hand, if you do A_2, you secure the tiny-box Newcomb box prize t, but sacrifice the beer. So in all likelihood, the choice you face is a tradeoff between t and beer. Finally, in the background: taking the beer rather than the small-box

prize is a good *sign* that a medium box, the contents of which you will also receive in either case, contains a medium prize m. But it cannot increase the chances of that.

t	tiny Newcomb prize
m	medium Newcomb prize
k	a Pabst Blue Ribbon beer

Finally, the even columns of Table 6 address the unlikely event that the predictor *isn't* Dolly. In that case, A_2 is strictly better than A_1, since, in addition to the contents of the medium Newcomb box, you also get t and a beer to boot. And, of course—to rehearse standard two-boxer reasoning—the medium box might contain m and it might not. But nothing you can do now will make any difference to *that*.

…It seems clear—at least, to me—that what you should do next will be influenced by how much you value Pabst Blue Ribbon (viz., k). Suppose you value it a lot. Then you have a very strong reason to do A_1. If you are rational and follow through, then as you execute A_1—the equivalent of riding the bus to Secaucus, dutifully updating your act-probabilities along the way—your credence in H rises. So your expectation of getting m rises. But you still see yourself as sacrificing t for the (nearly sure-thing) beer. Suppose on the other hand that you value Pabst Blue Ribbon very little. Then your additional reason to do A_1 is correspondingly weak. If you are rational and follow through, then as you execute A_2, hanging around Hoboken until the sun goes down, your credence in $\neg H$ rises. Your expectation that the medium box contains m thus falls. You could perfectly well choose t and think (since the predictor is almost surely Dolly) "it was worth it, though, by doing A_2 instead of A_1, I am poorer by one cheap beer."

It's worth noting that the way I just told the story said nothing about my *initial* confidence that I'd go for the beer (viz., choose A_1, the act that gives me a beer only in the very likely event that the predictor is Dolly). That is the difference between my view and Skyrms's—mine does not privilege the prior over acts.

…That was a bit complicated, so I'd like to tell the same story (the story of Table 6) again, beginning with Newcomb aspect of the problem this time. Here goes. You are (a two-boxer and you're) in a Newcomb problem. There is thus a transparent box before you with a free \$1000 in it. You can take it just in case you *don't* raise your arm (A_2). So, offhand, you think the laws of nature—which are very likely to be D—predict that you won't raise your arm. But now: an eccentric offers you \$$k$ iff:

(i) D is true and you *do* raise your arm, or
(ii) D is false and you do *not* raise your arm.

Whether the combined choice you now face is decision-dependent depends on the value of k; it *is* decision-dependent when $k > 1000$. Suppose, for example, that k is 2000. You *were*, recall, *initially* quite confident that D predicted you would keep your arm down. If causal expected utilities of A_1 and A_2 are calculated relative to this "locked" prior, then you believe that doing A_1 is very likely to net you a marginal gain of \$0, whereas A_2 is a sure-thing marginal \$1000 (Figure 8a). The *CEU* of the status quo thus drops locally if you increase your confidence in A_1. If you nevertheless persist

FIGURE 8a $k >$ \$1000. Shading indicates high terminal confidence.

Before additional offer	$D \wedge D$ pred A_1	$D \wedge D$ pred A_2
A_1 (raise)	[big box]	[big box]
A_2 (don't raise)	[big box] + \$1000	[big box] + \$1000

After	$D \wedge D$ pred A_1	$D \wedge D$ pred A_2
A_1 (raise)	[big box] + \$2000	[big box]+ \$2000
A_2 (don't raise)	[big box] + \$1000	[big box] + \$1000

FIGURE 8b $k <$ \$1000. Shading indicates high terminal confidence.

Before additional offer	$D \wedge D$ pred A_1	$D \wedge D$ pred A_2
A_1 (raise)	[big box]	[big box]
A_2 (don't raise)	[big box] + \$1000	[big box] + \$1000

After	$D \wedge D$ pred A_1	$D \wedge D$ pred A_2
A_1 (raise)	[big box] + \$100	[big box]+ \$100
A_2 (don't raise)	[big box] + \$1000	[big box] + \$1000

in increasing your confidence in A_1, you thereby increase your confidence in ($D \wedge D$ predicted A_1)—viz., in ($H \wedge D$). Hence you become more confident as you execute your act that A_1 will gain you \$2000.

If $k <$ 1000, you have no reason to alter your current credences in the predictions of D, because you are still in a Newcomb Problem. Say \$$k$ is only \$100 (Figure 8b). Then keeping your arm down dominates raising it, no matter what you believe about what D predicts. You decline, so you get the \$1000 and your arm stays put. This is a standard Newcomb Problem, except that the margin by which the expected utility of two-boxing exceeds the expected utility of one-boxing has narrowed slightly (viz., it has shrunk by \$100).

Summing up, the eccentric has hardly spoiled your day: you'll either get \$1000, or you will get \$$k$, whichever is bigger. In each case, by following CDT, you will wind up certain you did better than you otherwise would have.

4.3 | Summing Up

In this section, I aim to lay out a preliminary statement of my differences with Skyrms before concluding.

Skyrms (1990) directs an agent to choose, if possible, an act A:

(i) which maximizes *CEU* after the agent's prior P evolves according to a local dynamical law.
(ii) whose *CEU* does not fall below the *CEU* of any other option as its probability approaches 1.

I have hedged the statement of Skyrms's view with "if possible", because in nasty cases like **Death in Damascus**, whether making such a choice *is* possible will turn on the availability of mixed acts.[15] Should mixed acts be admitted, Skyrms's (i) entails that there is an equilibrium probability

[15] This depends, in turn, on the agent's *ability* to perform mixed acts—at least, on the usual construal of what equilibrium distributions $\{P(A) : A \in \{A\}\}$ represent. On the alternative view of Arntzenius (2008), by contrast, decision theory

distribution $P_{t'}$, such that $P_{t'+1}(A) = P_{t'}(A)$ for all $A \in \{A\}$.[16] If mixed acts are admitted, therefore, Skyrms's (i) entails (ii).

The bare-bones positive view I inclined towards above, by contrast, was simply (ii) without (i): choose, if possible, an act A:

(ii) whose CEU does not fall below the CEU of any other option as its probability approaches 1.

(ii) is known in the literature as (causal) *ratifiability*, and my bare-bones view thus coincides with the views endorsed by other ratifiability-inclined causal decision theorists, such as Harper (1986, 1992), Weirich (1985), Joyce (2007)[17], and Armendt (2019).

5 | CONCLUSION(S)

In this paper, after presenting Elga's impossibility proof for CDT under determinism, I explored a way of resisting it. My argument leveraged decision dependence—which arises natively out of Jeffrey Conditioning and CDT's characteristic equation—to work around the key assumption of Elga's proof: to wit, that in **Problems 1** and **2**, the CDTer must employ subjunctive-suppositional (rather than evidential) transformations of a shared prior. Achieving absolution did involve some evolution: I stepped slightly away from Skyrms's received view, sketching a form of CDT with deliberational dynamics that helps itself to the idea that act-probabilities can pursue the distal at the expense of the local good.

I am unsure whether there is a substantive disagreement between Skyrms and myself on the issue of locality. Many of the dynamical laws Skyrms considers make their home in the setting of evolutionary games and population dynamics, rather than in (subjective, one-shot) decision theory. In evolutionary game theory contexts—such as the ones studied by Maynard-Smith (1982)—the prior over acts represents proportions of strategies played by a whole interacting species population, and the dynamics are stipulated to be driven by Darwinian mechanisms that are clearly local in character. The considerations raised against locality in this paper do not carry over to these settings.[18]

In concluding, I would like to touch briefly on three matters relevant to the recent dialectic surrounding Elga's paper. The first is a dilemma for the semantics of counterfactuals raised by Cian Dorr (2016). The second is James Joyce (2016)'s response to Arif Ahmed's deterministic counterexamples to CDT, which bear important similarities (flagged by Elga) to the dilemma discussed here. The third is an intriguing footnote of Elga's that highlights the possibility of an agent's being ignorant of her own value function.

does not *aim* to issue in acts (mixed or otherwise), but *only* in (possibly mixed) credal states. Hence for Arntzenius the recommendations of CDT do not, strictly speaking, depend on the *ability* of an agent implement mixing.

[16] In the symmetric **Death in Damascus** case, for example, this point is $P(\mathsf{Alep}) = P(\mathsf{Dam}) = .5$.

[17] Though compare Joyce (2012), who argues against ratifiability in Nasty Problems.

[18] Intriguingly, it is worth noting that even within the context of evolutionary game theory, one can find dissatisfaction with locality (sometimes called *history-dependence*). In particular, *stochastic evolutionary game dynamics* studies the effect of perturbations—environmental shocks, spontaneous mutations in strategies, and other probabilistic phenomena—that make equilibrium selection less history-dependent (Wallace & Young, 2014, pg. 239). This treatment ends up with solution concepts, like risk-dominant equilibria, that more closely approximate ratifiability; see in particular the treatment of the coordination game in Wallace & Young *op. cit*., pg. 335.

5.1 | A Question from Dorr (2016)

The commitments I have made in this paper about the imaging operation, which transforms $P(\cdot)$ into $P^A(\cdot)$, are minimal in the extreme. They include the **Constraint on Imaging**, which is best understood as a claim about what imaging does *not* do (rather than what it does do).[19] Beyond that, I have accepted, for the sake of engaging with Elga's argument, only a very specific further claim: to wit, that where

- P is one's prior and $\{A_1, A_2\}$ are "mundane" actions such as raising one's arm,
- D is a deterministic theory of the laws of nature, and
- H is the most inclusive specification of initial conditions which, under D, entail A_1,

…that we may conclude that $P^{A_1}(D \wedge \overline{H}) = P^{A_2}(D \wedge H) = 0$. (As noted in §2, the best form of an argument for this conclusion likely goes not *directly* through imaging, but first through an appeal to chance: for example, an argument that $Ch(D \wedge \overline{H} \wedge A_1) = 0$ is entailed by the description of the case, which entails in turn that $P^{A_1}(D \wedge \overline{H}) = 0$ for any P that treats chance as an expert.)

One might argue that, as a defender of CDT, I owe a more positive, and metaphysically general, account of how imaging *does* shift values assigned by the prior. Questions of this kind have been raised about the very general modal relation—sometimes called the "selection function" (Stalnaker, 1968)—that is held in the literature to underwrite both imaging and the semantics of counterfactuals.[20] Take a question posed by Dorr (2016): supposing Determinism is true, and that (to use Dorr's example) I did not blink a moment ago, which of the following counterfactuals:

(1) If I had blinked, the past would have been different.

or

(2) If I had blinked, the laws would have been different.

shall we say is (nonvacuously) true? Each option seems very strange.

In Elga's dilemma, it looks like we must answer a similar question. Let A^* rigidly denote the act in $\{A_1, A_2\}$ that the agent in fact performs, and let H^* be the actually true member of $\{H, \overline{H}\}$. Since $A_2 = \overline{A_1}$ and $A^* \in \{A_1, A_2\}$, we know that $P^{\overline{A*}} \in \{P^{A_1}, P^{A_2}\}$. Since we've granted that $P^{A_1}(D \wedge \overline{H}) = P^{A_2}(D \wedge H) = 0$, we can conclude that if D is true, and *either* (3) or (4) below is false, the other must, surprisingly, be true:

(3) If the agent had done otherwise, the past would have been different.
$P^{\overline{A*}}(H^*) = 0$

(4) If the agent had done otherwise, the laws would have been different.
$P^{\overline{A*}}(D) = 0$

[19] That is to say: when $\{A\}$ is counterfactually independent of $\{S\}$, the Constraint says imaging on $\{A\}$ does *not* change the probability of $S \in \{S\}$.

[20] See Lewis (1981, §10) and Lewis (1976, §6). Stalnaker (1968)'s selection function semantics for counterfactuals $A \mathbin{\square\!\!\rightarrow} B$ is the source of the bridge principle $P(A \mathbin{\square\!\!\rightarrow} B) = P^A(B)$ Lewis uses in the 1976 article.

And from this observation about counterfactuals, two claims seem to follow.[21]

> **Claim 1 (tradeoff)**. In Elga's problem, if $\{D, \neg D\}$ is (believed by an agent to be) counterfactually independent of $\{A_1, A_2\}$, then $\{H, \neg H\}$ is (believed by that agent to be) counterfactually *de*-pendent on $\{A_1, A_2\}$, and vice-versa.

> **Claim 2 (powers)**. In Elga's problem, if $\{D, \neg D\}$ is (believed by an agent to be) outside her power to change, then $\{H, \neg H\}$ is (believed by that agent to be) *within* her power to change, and vice-versa.

I do not think the decision-dependent approach to CDT defended in this paper is committed to the truth of either Claim 1 or Claim 2, because it is not committed to the idea that even *one* of (3)-(4) is nonvacuous in the relevant sense. I will try here to briefly say why.

Suppose, instead of Elga's mere high confidence, that we are *completely certain* that some deterministic theory—which we might as well continue to call 'D'—holds, and consider this simple variation on **Nice Choices in New Jersey**.

> **Explicitly Deterministic Nice Choices in New Jersey (EDNCNJ)**. Yesterday, just as before, a predictor left $10 in Hoboken iff she believed you would go there, and $15 in Secaucus iff she believed you would go there. You are certain that she used the true deterministic theory D to make her prediction. So you are certain that: either the objective chance you will go to Hoboken is 0, or the objective chance that you will go to Secaucus is 0.

As far as I can see, this twist on **Nice Choices in New Jersey** is entirely compatible with the permissive deliberational dynamics approach I advocated above. But nothing analogous to Claims 1-2 hold in **EDNCNJ**: the bare fact that I consider {Predicted Hoboken, Predicted Secaucus} to be counterfactually independent of my actions—*ergo*, not within my power to change—does not entail that I believe of some *other* state-partition $\{Y, \overline{Y}\}$ that's relevant to determining my utilities that it *is* within my power to change.

Rather, **EDNCNJ** is compatible with the analysis sketched above because all of the following can still be stipulated to hold in the story:

(1) P(Sea) and P(Hob) are intermediate at $t = 0$;
(2) P(Predicted Hoboken | Hob) and P(Predicted Secaucus | Sea) are both approximately 1;
(3) {Predicted Hoboken, Predicted Secaucus} is nonetheless counterfactually independent of {Hob, Sea}, and thus subject to the **Constraint on Imaging**;
(4) I can alter my act-probabilities across {Hob, Sea} at will;
(5) I update as I do so, *if* I do so, by Jeffrey Conditionalization;
(6) CEU_t(Hob) and CEU_t(Sea) are (therefore) decision-dependent.

CDT with deliberational dynamics thus enjoys a tactical advantage in the context of debates about Determinism: it provides prescriptions for rational decisions even under such (apparently paralyzing) constraints as Determinism's fully believed truth.

[21] I am indebted to an anonymous referee for encouraging me to consider these claims.

This is not, of course, to say that **Claims 1-2** are false. It is only to say that the answer to the practical question of what to do in **EDNCNJ**, like the question of what to do in the original **Nice Choices in New Jersey**, does not seem to depend on their truth.

5.2 | The view from Joyce (2016)

This neutrality within deliberational dynamics also bears on the comparison between my view and that of Joyce (2016). Joyce aims, as I do here, to defend CDT from the charge that it cannot bet rationally on deterministic laws. In a nutshell, his response to gambles where (theories like) D play a role in individuating prizes is this. Given what you believe, if D is true, you are not really in a decision problem at all. This is because one of your apparently available options is (nomically) impossible for you, though you know not which. It follows, Joyce argues, that decision theorists can de facto ignore the D-columns of any such decision problem, even when $P(D)$ is high. For example, in **Stacked Problems 1 and 2**—the decision problem illustrated in Table 6— we should pay attention to the recommendations of CDT only where D is false. The problem is quite easily decided in that case, since, from the CDTer's point of view, one of the options strictly dominates the other.[22]

On a "hard" construal of determinism, Joyce's analysis seems right to me. I suspect that the permissive form of causalist deliberational dynamics I explored in this paper can be seen as another way of making the same point. If we are to take naturalistic decision theory seriously, then we should take seriously the idea that the agent can increase her confidence in any option $A \in \{A\}$ at will. In the case of D, this means that, even if one of my options is objectively impossible, I can choose which option I (justifiedly!) believe it is. To put things the other way around: suppose an agent *cannot* choose which option she (justifiedly) believes she is bringing about in the way the view I sketched presupposes. Then I think there is reason to be skeptical, with Joyce, that she is really in a decision problem at all.

5.3 | A footnote from Elga

Finally, there is the matter of Elga's footnote. Here is what he has to say (emphasis added):

> Since in [**Problem 1**] and [**Problem 2**] you have the same probability function but different value functions, *in at least one of the two situations you are ignorant or incorrect about your value function*. In response to *the worry that such ignorance compromises verdicts about what it is rational to do in [these] situations*, there are at least two options. (1) one might hold that rationality requires one to maximize expected utility even when one is less than omniscient about one's values. (2) One might model the whole setup with probabilities defined over a space of 'coarsened' elementary possibilities each of which is silent about the subject's values. Doing so would remove the need to say that in either situation the subject is mistaken about her values. (Elga *op cit*., ftn. 8, quoted in its entirety; emphasis mine)

[22] See in particular Joyce *op. cit*., pg. 226, where he makes a this point about dominance in the context of a very similar decision problem.

What I take Elga to be highlighting here is the possibility of taking a strategy quite like the one I took in this paper. He is flagging that his impossibility result can be blocked if the prior evolves in response to the prizes featured in the relevant problem(s) before (causal) expected utility is calculated.

Because I do not know how to state a theory of CDT with deliberational dynamics under the more relaxed assumptions about value Elga sketches in the quoted passage—and he does not offer one himself—this line of response lies outside the scope my discussion. I confess do not straightaway see how Elga's sketch would go. Suppose I know that I'll get a Coffee Roll iff I raise my arm and a Boston Kreme otherwise, but I am given no information about how v(Coffee Roll) compares to v(Boston Kreme). Can expected utility require anything of me here?

If the issue is expectation of v, rather than v itself, then we are in somewhat more familiar territory. In that case I will say only this: I find it odd to conceptualize the potential weakness of Elga's proof in terms of *ignorance* of one's (expected) value function. What I, following Skyrms, emphasized was that the expected value function *adapts* in response to prizes the world throws up at us. The possibility of such adaptation seems fundamental—not to some weird ideal of self-transparency, but to the concept of decisionmaking as such.

APPENDIX

Skyrms CDT sees asymmetric nice cases as ones where your act-credences are constrained to seek the *local* good; you must, in that sense, remain in the grip of your prior. Below are some rules that appear in Skyrms (1990); we consider them with respect to a prior P according to which $P(\text{Hob}) = .75$ and $P(\text{Sea}) = .25$ (the vertical line in Figure 4).

(1) Nash map: where $cov(A_i)$, the *covetability* of an act A_i, is $\max[CEU(A_i) - CEU(SQ), 0]$:

$$P^+(A_i) = \frac{P(A_i) + cov(A_i)}{1 + \sum_i cov(A_i)}$$

In **Nice Choices in New Jersey**, $cov(\text{Hob})$ is positive with respect to the prior P, while $cov(\text{Sea})$ is 0; hence $P^+(\text{Sea}) < P(\text{Sea})$.

(2) The general Nash map family is parameterized to a weight k:

$$P^+(A_i) = \frac{kP(A_i) + cov(A_i)}{k + \sum_i cov(A_i)}$$

Once again, $cov(\text{Hob})$ is positive with respect to the prior P, while $cov(\text{Sea})$ is 0. Thus $\sum_i cov(A_i) = cov(\text{Hob})$, and so we have $P^+(\text{Sea}) = \frac{k \cdot P(\text{Sea})}{k + cov(\text{Hob})}$. Hence whenever $k > 1$, $P^+(\text{Sea}) < P(\text{Sea})$.

(3) Darwin (or Maynard Smith[23]) map:

$$P^+(A_i) = P(A_i)\frac{CEU(A_i)}{CEU(SQ)}$$

[23] Maynard-Smith (1982).

…this map requires that utilities are nonnegative, a condition which is already met by the prizes in **Nice Choices in New Jersey**. With respect to the prior P, $CEU(\text{Sea}) < CEU(SQ)$. Hence $\frac{CEU(\text{Sea})}{CEU(SQ)} < 1$ and so again $P^+(\text{Sea}) < P(\text{Sea})$.

ACKNOWLEDGMENTS

I am grateful to Simon Huttegger, Matt Mandelkern, Calum McNamara, audiences at Carnegie Mellon and New York University, and two anonymous *Noûs* referees for generous commentary and discussion.

ORCID

Melissa Fusco https://orcid.org/0000-0001-6512-2357

REFERENCES

Ahmed, A. (2013). Causal decision theory: A counterexample. *Philosophical Review*, *122*(2).

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

Albert, D. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.

Armendt, B. (2019). Causal decision theory and decision instability. *Journal of Philosophy*, *116*(5), 263–277. doi: https://doi.org/10.5840/jphil2019116517

Armstrong, D. (1983). *What Is a Law of Nature?* Cambridge: Cambridge University Press.

Arntzenius, F. (2008). No regrets, or: Edith Piaf revamps decision theory. *Erkenntnis*, *68*(2), 277–297.

Dorr, C. (2016). Against counterfactual miracles. *Philosophical Review*, *125*(2), 241–286. doi: https://doi.org/10.1215/00318108-3453187

Egan, A. (2007). Some counterexamples to causal decision theory. *The Philosophical Review*, *116*(1), 93–114. doi: https://doi.org/10.1215/00318108-2006-023

Elga, A. (2022). Confessions of a causal decision theorist. *Analysis*, *82*(2), 203–213. doi: https://doi.org/10.1093/analys/anab040

Gärdenfors, P. (1982). Imaging and conditionalization. *Journal of Philosophy*, *79*(12), 747–760.

Gibbard, A., & Harper, W. (1978). Counterfactuals and two kinds of expected utility. In C. A. Hooker, J. J. Leach, & E. F. McClennen (Eds.), *Foundations and Applications of Decision Theory, Vol 1*. Dordrecht: D. Reidel.

Goodman, N. (1965). *Fact, Fiction, and Forecast*. Indianapolis: Bobbs-Merrill.

Hare, C., & Hedden, B. (2016). Self-reinforcing and self-frustrating decisions. *Noûs*, *50*(3), 604–628. doi: https://doi.org/10.1111/nous/12094

Harman, E. (2009). 'I'll be glad I did it' reasoning and the significance of future desires. *Philosophical Perspectives*, *23*.

Harper, W. (1986). Ratifiability and causal decision theory: Comments on Eells and Seidenfeld. *Philosophy of Science*, *2*, 213–228.

Harper, W. (1992). Dynamic deliberation. *Philosophy of Science Association*, *2*, 353–364.

Harper, W. (2022). Decision dynamics and rational choice. In B. Dunaway & D. Plunkett (Eds.), *Meaning, Decision, & Norms: Themes from the Work of Allan Gibbard*. Michigan Publishing.

Jeffrey, R. (1983). *The Logic of Decision*. University of Chicago Press.

Joyce, J. (2007). Are Newcomb problems really decisions? *Synthese*, 537–562. doi: https://doi.org/10.1007/sl1229-006-9137-6

Joyce, J. (2010). Causal reasoning and backtracking. *Philosophical Studies*, *147*, 139–154. doi: https://doi.org/10.1007/s11098-009-9454-y

Joyce, J. (2012). Regret and instability in causal decision theory. *Synthese*, *187*, 123–145. doi: https://doi.org/10.1007/s11229-011-0022-6

Joyce, J. (2016). Review of *Evidence, Decision and Causality*, by Arif Ahmed. *Journal of Philosophy*, *113*(4), 224–232. doi: https://doi.org/10.5840/jphil2016113413

Latham, N. (1987). Singular causal statements and strict deterministic laws. *Pacific Philosophical Quarterly*, *68*, 29–43.

Lauro, G., & Huttegger, S. (2022). Structural stability in causal decision theory. *Erkenntnis*, *87*, 603–621. doi: https://doi.org/10.1007/s10670-019-00210-6
Levinstein, B., & Soares, N. (2020). Cheating death in Damascus. *Journal of Philosophy*, *CXVII*, 237–266.
Lewis, D. (1971). A subjectivist's guide to objective chance. *Studies in Inductive Logic and Probability 2*.
Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*, 297–315.
Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, *59*(1), 5–30.
Maudlin, T. (2004). Causation, counterfactuals, and the third factor. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals*. MIT Press.
Maynard-Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
Meacham, C. (2010). Binding and its consequences. *Philosophical Studies*, *149*, 49–71. doi: https://doi.org/10.1007/s11098-010-9539-7
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
Richter, R. (1984). Rationality revisited. *Australasian Journal of Philosophy*, *62*(4), 392–403.
Skyrms, B. (1981). The prior propensity account of subjunctive conditionals. In W. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*. D. Reidel, Dordrecht.
Skyrms, B. (1982). Causal decision theory. *The Journal of Philosophy*, *79*(11), 695–711.
Skyrms, B. (1990). *The Dynamics of Rational Deliberation*. Harvard University Press.
Skyrms, B. (2022). Appendix 2a: Death in Damascus: Continuous time dynamics. In B. Dunaway & D. Plunkett (Eds.), *Meaning, Decision, & Norms: Themes from the Work of Allan Gibbard*. Michigan Publishing.
Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*. Oxford: Blackwell.
Stalnaker, R. (2018). Game theory and decision theory (causal and evidential). In A. Ahmed (Ed.), *Newcomb's Problem* (pp. 180–200). Cambridge University Press.
Stanley, J. (2005). *Knowledge and Practical Interests*. Oxford University Press.
Wallace, C., & Young, H. P. (2014). Stochastic evolutionary game dynamics. In H. P. Young & S. Zamir (Eds.), *Handbook of Game Theory with Economic Applications*, volume 4 (pp. 327–380). Elsevier.
Weirich, P. (1985). Decision instability. *Australasian Journal of Philosophy*, *63*(4), 465–473.

How to cite this article: Fusco, M. (2023). Absolution of a Causal Decision Theorist. *Noûs*, 1–28. https://doi.org/10.1111/nous.12459