Confession of a causal decision theorist

Adam Elga
April 29, 2021

God, grant me the serenity to accept the things I cannot change,
Courage to change the things I can,
And wisdom to know the difference.
– Reinhold Niebuhr

1 *Introduction*

Start with two questions:

1. Suppose that you care only about speaking the truth, and are confident that some particular deterministic theory is true. If someone asks you whether that theory is true (and you must answer either "yes" or "no") are you rationally required to answer "yes"?
2. Suppose that you face a problem in which (as in Newcomb's problem) one of your options—call it "taking two boxes"—*causally dominates* your only other option—call it "taking one box".[1] Are you rationally required to take two boxes?

Those of us attracted to causal decision theory are under pressure to answer "yes" to both questions. The first "yes" seems obvious, even prior to commitments to any particular decision theory. And the second "yes" reflects a core commitment of causal decision theory.

It has been shown that many existing decision theories are inconsistent with answering "yes" to both questions (Ahmed 2014a, §5, §7, Ahmed 2014b, §5.2.1). My aim is to give a simple proof that the same goes for an even wider class of theories: all "suppositional" decision theories (according to

[1] In other words: There is a proposition—call it "million-present"—whose truth value is settled and is outside of your causal influence. Furthermore you regard million-present scenarios in which you take two boxes as better than million-present scenarios in which you take just one box, and the same goes for no-million-present scenarios.

which the value of an option is its expected value on the supposition that it is selected). Such theories include ones described in Ahmed (2014b), Gibbard and Harper (1978), Jeffrey (1965), Joyce (1999), Lewis (1981), Skyrms (1980), Stalnaker (1981), and many others.

So causal decision theorists must either answer "no" to one of the above questions, or else abandon suppositional decision theories.

2 *Suppositional decision theories*

Many evidential and causal decision theories are unified by a guiding idea: The value of an option is a weighted average of the values of the possibilities compatible with that option. Weighted how? By the agent's probability function on the *supposition* that the option is realized.[2]

More precisely, these theories can all be expressed as the requirement that one select an option A that maximizes

$$U(A) = \mathbb{E}(v, P^A) \stackrel{\text{def}}{=} \text{expectation of } v \text{ with respect to } P^A,$$

where A ranges over one's *options* (assumed to be disjoint propositions[3]), v is one's *value function* (a function from possible worlds to real numbers), and P^A is the result of starting with one's probability function P and *supposing* A. The function that maps P to P^A is a *supposition function*, required only to be such that for all probability functions P and all propositions A in the domain of P, the following condition holds: P^A is a probability function with $P^A(A) = 1$. Different decision theories understand the supposition function in different ways. Call theories in this family *suppositional decision*

[2] I learned of this unification from Joyce (1999), which explicitly formulates a suppositional decision theory that has evidential and causal decision theories as special cases. See also Lewis (1981).

[3] I assume throughout that propositions are sets of possible worlds and that probability functions are defined over a suitable algebra of propositions.

theories.[4]

Prominent and popular suppositional decision theories abound:

- ▹ **Evidential Decision Theory** (Jeffrey 1965) is gotten by letting the supposition function be *conditionalization*: $P^A(\cdot) = P(\cdot|A)$.

- ▹ **Counterfactual** causal decision theories (Gibbard and Harper 1978, Stalnaker 1981) are gotten by letting the supposition function be *counterfactual supposition*: $P^A(\cdot) = P(A \mathbin{\Box\!\!\rightarrow} \cdot)$.

- ▹ ***K*-partition** causal decision theories are defined in terms of a privileged partition $\boldsymbol{K}$ of "dependency hypotheses", each of which specifies causal dependencies between options and outcomes.[5] For such theories, the supposition function is *imaging relative to* $\boldsymbol{K}$: $P^A(\cdot) = \Sigma_{K \in \boldsymbol{K}} P(K)P(\cdot|AK)$.[6]

In addition to the above, a wide range of decision theories count as suppositional. For example, suppose that one takes a counterfactual-based causal decision theory and replaces the normal counterfactual conditional with another conditional satisfying a few undemanding constraints.[7] The

[4]The above definition of a supposition function is slightly more permissive than the one in Joyce (1999, Ch 6), in order to maximize the generality of the impossibility proof below. There is some technical messiness associated with the case of $P(A) = 0$, orthogonal to present concerns, that I am here ignoring.

[5] Theories differ on the exact characterization of $\boldsymbol{K}$, whose members are also sometimes termed "states" or "act independent states" (see Joyce (1999, 165) for a survey). One influential proposal has it that each member of $\boldsymbol{K}$ fully specifies "how the things [one] cares about do and do not depend causally on [one's] present actions" (Lewis 1981, 11). Note that since each suppositional decision theory uses a single supposition function, a K-partition theory counts as suppositional only if it entails that K-partitions for different decisions always induce the same supposition function. I consider relaxing this assumption at the end of §6.

[6]Here "AK" denotes the conjunction (intersection) of propositions A and K. The stated definition of imaging assumes that the partition $\boldsymbol{K}$ is countable and that P is countably additive, assumptions I adopt going forward.

[7]To ensure that a conditional induces an imaging function, it is sufficient that it be a "centered conditional" (Joyce 1999, 64) which satisfies "Conditional Contradiction", "Harmony", and "Conditional Excluded Middle" (Joyce 1999, 168).

result would be another suppositional theory. Or suppose that one takes a partition-based theory and modifies the partition of act-independent states. Again the result would be a suppositional theory.

Though a wide range of decision theories count as suppositional, none of them are consistent with answering "yes" to both of the questions at the start of this paper. Or so I shall argue, by describing two cases.

3 *Betting on the laws*

Here is the first case:[8]

Betting on the laws Let D be the proposition that some particular deterministic regularities are exceptionless[9] laws of nature.[10] For example, D might be the proposition that the laws of nature include (a deterministic formulation of) Newtonian mechanics. Your total evidence favors D over its negation, and so $P(D) > 1/2$, where P is your probability function. You must choose between endorsing D by raising your hand (option A_1) and denying D by not doing so (option A_2). You are certain that your choice has no causal influence on whether D is true or on the state of the world in the distant past. You care only about endorsing truths and denying falsehoods on this occasion, as reflected by your value function v_1, pictured here:[11]

[8] This case due to Ahmed (2013), though I have changed some details. See also Ahmed (2014b, Ch. 5).

[9] Why "exceptionless"? To close off the potential escape route of claiming that a system of deterministic laws can obtain even though some violations of it occur (Braddon-Mitchell 2001).

[10] "Deterministic regularity" is here understood so that D, together which a full specification of the state of the world at any one time, entails a full specification of the state of world at all times. States of the world are assumed to be rich enough so that any two D-worlds whose states perfectly match at all times are alike with respect to which options you choose at all times.

[11] This table represents no more than the following: that $v_1(w) = 1$ for any world w in $A_1 D \cup A_2 \overline{D}$, and $v_1(w) = 0$ for any w in $A_1 \overline{D} \cup A_2 D$. (Notation: a horizontal line above the

v_1	D	$\overline{D}$
A_1	1	0
A_2	0	1

One verdict seems obvious about this case:

Bet "Betting on the laws" situations are possible, and in any such situation you should choose A_1 (endorse D).

Even prior to one's commitment to a particular decision theory, this verdict should have significant appeal. For **Bet** to be false, either the "Betting on the laws" situation would have to be impossible, or else it would have to be that even though you reasonably have $P(D) > 1/2$ (and you are sure that your bet has no causal influence on whether D is true), you are not rationally required to bet that D is true. Neither option has much independent appeal.[12]

4 *Newcomb on the past*

Here is the second case:[13]

Newcomb on the past Let H be the proposition that the truth about the state of the world at the first moment of 1900, when conjoined with D, entails that you raise your hand on the present occasion.[14] The

name of a proposition represents negation.) Subsequent tables for value functions are to be interpreted similarly. Note that unlike many decision tables, the "value tables" in this paper are not meant to convey or presuppose that the partition of propositions labeling the columns has any special status.

[12]This reasoning is similar in spirit to the defense of the "Causal Betting Principle" from Ahmed (2013, 292).

[13]This is a variant of the "Betting on the past" case from Ahmed (2014a). See also Ahmed (2014b, Ch. 5), Ahmed (2010, 123n5).

[14] I assume that H is consistent with D. H is a proposition entirely about the state of the world at the first moment of 1900. In particular, H neither entails D nor is entailed by D. Instead, H is the set of worlds w such that: at every $w' \in D$ such that w and w' are in the same state at the first moment of 1900, you raise your hand (in the present situation). (Compare Ahmed (2014a, 6).)

proposition D, your options, and your probability function P are all the same as they are in "Betting on the laws", but your value function v_2 (pictured below) differs as follows.[15] Here is all you care about:

- ▹ You greatly prefer that H be true (that is worth M utiles to you, where $M > 0$).
- ▹ You slightly prefer that you not raise your hand (that is worth T utiles to you, where $T > 0$).

v_2	H	$\overline{H}$
A_1	M	0
A_2	$M+T$	$0+T$

In this case the same sort of dominance reasoning supports A_2 as supports taking two boxes in a standard Newcomb problem (Nozick 1969). Indeed, the above table represents the payoffs in a standard Newcomb problem if one reads H as the proposition that the opaque box has $1 million, M as one million, T as one thousand, and A_1 and A_2 as the options of taking one or two boxes respectively.[16]

In a standard Newcomb problem there is a causal dominance argument for taking two boxes: "The $1 million is either there or it is not, and you have no causal influence on whether it is. Either way (and no matter what else is true), taking two boxes gets you a better outcome than taking just

[15] Since in the two situations you have the same probability function but different value functions, the setup entails that 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 the 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.

[16] Here I impose the harmless idealization that you care only about the combination of two factors: (1) whether the opaque box has $1 million, and (2) whether you take one or two boxes.

one. So you should take two boxes." Generalized and made more precise, the idea is that when choosing between A_1 and A_2 you should choose A_2 whenever:

(a) You are certain you have no present influence over whether H is true.

(b) You strictly prefer every A_2H world to every A_1H world.

(c) You strictly prefer every $A_2\overline{H}$ world to every $A_1\overline{H}$ world.

These conditions are satisfied in "Newcomb on the past" just as much as they are in a standard Newcomb problem. So those who are sympathetic to the spirit of causal decision theory are under some pressure to endorse:

Two-box "Newcomb on the past" situations are possible, and in any such situation you should choose A_2.

5 *Trouble for suppositional causal decision theories*

Here is where the trouble begins. There are sound arguments that several prominent counterfactual-based causal decision theories are committed to denying **Bet** (Ahmed 2014a, §5), that a number of other decision theories are similarly committed (Ahmed 2014a, §7), and that so are causal decision theories that operate by way of a particular conception of "causal reach" or objective chances (Ahmed 2014b, §5.2.1).

Still: the above arguments, however compelling, attempt to play a game of whack-a-mole, ruling out theories one by one. So a friend of causal decision theory might hope that some undreamt-of suppositional decision theory is immune to the arguments and consistent with both **Bet** and **Two-box**:

Suppositional Some suppositional decision theory—perhaps one yet to be formulated—delivers correct verdicts about all "Betting on the laws" and "Newcomb on the past" cases.

Contrary to the above hope, however, the following proof shows that **Bet**, **Two-box**, and **Suppositional** are jointly inconsistent.[17]

For the proof it will be convenient to have additional representations of the value functions in "Betting on the laws" (v_1) and "Newcomb on the past" (v_2):

v_1	HD	$H\overline{D}$	$\overline{H}D$	$\overline{H}\,\overline{D}$
A_1	1	0		0
A_2		1	0	1

v_2	HD	$H\overline{D}$	$\overline{H}D$	$\overline{H}\,\overline{D}$
A_1	M	M		0
A_2		$M+T$	T	T

Note that any D-world in which you raise your hand is a member of H (since any such world was in a state at the start of 1900 that together with D entails that you raise your hand), and so $A_1\overline{H}D = \emptyset$. Similarly $A_2HD = \emptyset$. That is why the corresponding cells in the above tables have been left blank.

Suppose for contradiction that **Bet**, **Two-box**, and **Suppositional** are true. Then by **Suppositional**, some suppositional decision theory delivers correct verdicts about "Betting on the laws" and "Newcomb on the past". Since the subjects in those cases have the same probability function P, we may write the values of the suppositional probabilities of this theory for both cases as follows:[18]

	HD	$H\overline{D}$	$\overline{H}D$	$\overline{H}\,\overline{D}$
P^{A_1}	a	b	0	c
P^{A_2}	0	d	e	f

where a through f are real numbers in the unit interval, $a+b+c = 1$ $= d+e+f$ (by the definition of a supposition function), and the 0 entries are forced because $A_2HD = A_1\overline{H}D = \emptyset$.

[17] Compare to the "Bonus Newcomb Problem" argument in Solomon (2019, §4).

[18] This table represents no more than that $P^{A_1}(HD) = a, P^{A_1}(H\overline{D}) = b, \ldots, P^{A_2}(\overline{H}\,\overline{D}) = f$. In particular it does not represent that any of HD, $H\overline{D}$, $\overline{H}D$, or $\overline{H}\,\overline{D}$ are dependency hypotheses.

On the one hand, by **Bet** and **Suppositional**:

$$\begin{aligned} \mathbb{E}(v_1, P^{A_2}) &< \mathbb{E}(v_1, P^{A_1}) \\ d + f &< a. \end{aligned} \tag{1}$$

On the other hand, by **Two-box** and **Suppositional**, for all $M, T > 0$:

$$\begin{aligned} \mathbb{E}(v_2, P^{A_1}) &< \mathbb{E}(v_2, P^{A_2}) \\ Ma + Mb &< (M+T)d + Te + Tf \\ M(a+b-d) &< T(d+e+f) \\ a+b-d &< T/M && \text{Since } d+e+f=1 \\ a+b-d &\leq 0 && \text{By continuity of the reals} \\ a &\leq d-b. \end{aligned} \tag{2}$$

Combining (1) and (2) we have that $d + f < d - b$, and hence that $f + b < 0$. But that contradicts the assumption that f and b are each nonnegative.

6 *Which assumption should be rejected?*

In the face of the above result, decision theorists must reject at least one of **Bet**, **Two-box**, and **Suppositional**. How might each rejection be motivated?

Start with **Bet**. One way to reject **Bet** is to claim that when determinism is true, one's present choice *does* have causal influence over what laws of nature obtain. Bales (2017, §4.3.1) expresses sympathy with this thought, assuming a deflationary Humean analysis of the laws of nature.

A second way is to give up causal decision theory (strictly so-called) in favor of "non-backtracking counterfactual dependence decision theory" (Hitchcock 2013, 139). One could then reject **Bet** on the grounds that one's present choice has not causal but "non-backtracking counterfactual" influence over what laws of nature obtain.[19]

Turn now to **Two-box**. Evidential decision theorists are of course happy to reject it. But in addition, inspired by Dorr (2016) and Loewer (Unpublished) one might argue as follows: No attractive theory of counterfactuals entails (roughly speaking) that both the past and the laws are counterfactually independent of one's present actions. Given this conflict, we should adopt a "causal-counterfactual" decision theory backed by a species of counterfactual according to which the laws are counterfactually independent of one's present actions, but the past is not. Such analyses of counterfactuals can be motivated by statistical mechanics (Albert 2000, 2015, Kutach 2002, Loewer 2007). The resulting decision theory would underwrite the denial of **Two-box**. Furthermore, any counterintuitiveness of rejecting **Two-box** should be tolerated because the subject in "Newcomb on the past" has such eccentric values.[20]

Another way to reject **Two-box** is to adopt a hybrid theory that rejects causal dominance reasoning in some special cases (such as "Newcomb on the past") while endorsing a version of it in many others. To arrive at such a theory one might exclude "unreachable" possibilities from expected utility calculations. For example, Sandgren and Williamson (forthcoming) spells out a way of excluding conjunctions of acts and dependency hypotheses thought to be inconsistent with the laws of nature.[21] And Kment (Unpublished) proposes (roughly) to calculate the utility of each option by first conditionalizing on this claim: that choosing that option is compatible with all truths beyond the decision-maker's causal influence.

[19]In particular one might endorse the a version of the partition-based decision theory of Lewis (1981) in which causal dependence is replaced by counterfactual dependence, and counterfactuals are given a "miracles" semantics (Lewis 2014).

[20]Here I mirror the discussion of the "Lavinia" case in Dorr (2016, 267).

[21]Note that the theory in Sandgren and Williamson (forthcoming) is not intended to on its own handle cases like "Betting on the laws". Rather it is intended to be deployed alongside strategies that reject **Suppositional** (Sandgren and Williamson forthcoming, n. 19, Williamson and Sandgren forthcoming, §6).

Another way to reject either **Bet** or **Two-box** is to argue that the cases figuring in them are not genuine decisions. Assume that the causal relations between acts and outcomes in any genuine decision can be represented by a suitable partition of dependency hypotheses. Then to show that a situation fails to be a genuine decision it is enough to show that no suitable partition exists. Joyce (2016, 225) and Williamson and Sandgren (forthcoming, §5.2.1) pursue this line, offering arguments that no partitions suitably represent "Newcomb on the past" and "Betting on the laws" respectively. In each instance candidate partitions are rejected because they violate an "Act-State Independence Principle" according which dependency hypotheses are counted as counterfactually independent of options.[22]

The above authors only argue against a restricted range of candidate partitions.[23] That leaves it open that a partition outside of that range might be adequate. But an argument in the spirit of Solomon (2019, §2.1) suggests that casting a wider net is unlikely to help.

To begin the argument, recall (from the discussion of K-partition theories in §2) that there is a canonical way for a partition of propositions to induce (determine) a supposition function. So we can impose constraints on partitions by imposing constraints to the supposition functions that they induce. Next note that causal decision theorists often enjoin us to

[22]In fact Joyce (2016) addresses "Betting on the past" (Ahmed 2014a), but similar considerations apply to "Newcomb on the past". "Act-state Independence" is Williamson and Sandgren's term; Joyce's principle $\mathbf{N_2}$ is similar (224). Note also that Williamson and Sandgren offer the above argument as just one of several alternative strategies, and that Joyce's view is more nuanced than the one described above, addressing not "genuine decisions" but rather "genuine Newcomb problems". What matters most in the present context is the denial that "Newcomb on the past" is the sort of decision to which a causal dominance argument applies (226). Such a denial would be a reason to reject **Two-box**.

[23]Joyce (2016) considers just the partition $\{HD, H\overline{D}, \overline{H}D, \overline{H}\,\overline{D}\}$ (notation adapted to the present context). The argument in Williamson and Sandgren (forthcoming) assumes (n. 3) that each conjunction of a dependency hypothesis and an option entails a unique outcome, and so does not cover indeterministic cases. It also assumes a "miracles" semantics for counterfactuals (Lewis 2014), and does not cover other sorts of counterfactuals, or other ways of characterizing dependency hypotheses.

compare options while holding fixed our opinions about matters beyond our influence. In a suppositional decision theory, doing so amounts to treating a proposition as *suppositionally independent* of your choice whenever you are certain that you have no causal influence on its truth.[24]

Now turn to "Betting on the laws" and "Newcomb on the past". In those cases you are certain that your choice has neither causal influence on the truth of H nor causal influence on the truth of D. So in modeling the cases it is natural to require a supposition function such that given your probability function P:

1. H is suppositionally independent of your choice, and
2. D is suppositionally independent of your choice.

Unfortunately, given that $P(D) > 1/2$ we can prove that no such supposition function meets those conditions.[25] It follows that no partition induces a supposition function that meets those conditions. That is reason to think that no partition of dependency hypotheses adequately represents both cases.

In the light of this result one might reject **Bet** or **Two-box** on grounds that "Betting on the past" or "Newcomb on the laws" are not genuine decisions. Alternatively, one might assume that the cases are genuine decisions and take the above argument as a reason to reject **Suppositional**.

[24]Definition: Proposition X is *suppositionally independent of your choice* (relative to P) if for each of your options A, your probability for X remains unchanged by supposing A: $P^A(X) = P(X)$. Solomon (2019, §2.1) motivates the requirement that propositions causally independent of your choice are suppositionally independent of it, and applies that requirement to propositions about the past and the laws of nature. The above definition of suppositional independence is slightly more demanding than the definition suggested by Joyce (1999, 162), but the two notions coincide for supposition functions determined by a K-partition.

[25]Proof: let $P^{\cdot}(\cdot)$ be the result of applying an arbitrary supposition function to your probability function P in "Betting on the laws" or "Newcomb on the past". Specify the values of $P^{\cdot}(\cdot)$ as in the last table of §5. We will show that if conditions (1) and (2) are satisfied, then $P(D) \leq 1/2$. By (1), $P^{A_1}(H) = P(H) = P^{A_2}(H)$ and so $a + b = P(H) = d + 0$. By (2), $P^{A_1}(D) = P(D) = P^{A_2}(D)$ and so $a + 0 = P(D) = 0 + e$. By the definition of a supposition function, $d + e + f = 1$. Using the substitutions $d = a + b$ and $e = a$ in this sum, we have $(a + b) + a + f = 1$, so $a = 1/2 - b/2 - f/2 \leq 1/2$. Hence $a = P(D) \leq 1/2$.

How might **Suppositional** be rejected? Bales (2017, 122–124) hopes for an attractive new decision theory consistent with analogs of **Bet** and **Two-box**, while cautioning that coming up with a suitable theory "might be a substantive challenge and might require substantial changes to existing accounts". Benchmark theory (Wedgwood 2013) is a notable non-suppositional decision theory consistent with **Bet** and **Two-box**, though it faces its own challenges (especially the proposed counterexample in Bassett (2015, §4.1)).

Other approaches for rejecting **Suppositional** include adopting a species of counterfactual whose semantics involves impossible worlds (Nolan 2017, Schwarz 2014, Williamson and Sandgren forthcoming, §5.1.1), and interpreting rigidified descriptions in novel ways (Williamson and Sandgren forthcoming, §5.2). Both approaches give up fairly fundamental framework assumptions, so their ultimate attractiveness will depend on whether alternative foundations can be constructed to support them.

A final approach worth exploring is a theory in which different supposition functions operate in different contexts.[26] Such a theory might recommend both that an agent facing "Betting on the laws" treat the laws (but not the distant past) as suppositionally independent of her present choice and that an agent facing "Newcomb on the past" treat the distant past (but not the laws) as suppositionally independent of her present choice. A challenge for this approach is to specify exactly how the appropriate supposition function depends on what decision an agent faces. It is not obvious how to do so in general, especially for hybrid choice situations that combine elements of "Betting on the laws" and "Newcomb on the past".

[26] This was independently suggested to me (but not endorsed) by Chris Meacham and Chris Register. A similar suggestion is considered but rejected by Solomon (2019, n. 21).

7 Conclusion

There are viable ways to reject each of **Bet**, **Two-box**, and **Suppositional**. Each way faces its own challenges and I won't try to choose between them here. But as a causal decision theorist I must confess: the independent appeal of each premise makes me wish I did not have to choose.[27]

References

A. Ahmed. Causal decision theory: a counterexample. *Philosophical Review*, 122(2):289–306, January 2013. doi: 10.1215/00318108-1963725.

A. Ahmed. Causal decision theory and the fixity of the past. *The British Journal for the Philosophy of Science*, 65(4):665–685, 2014a. doi: 10.1093/bjps/axt021.

Arif Ahmed. Causation and Decision. *Proceedings of the Aristotelian Society*, 110:111–131, 2010. ISSN 0066-7374.

Arif Ahmed. *Evidence, Decision and Causality*. Cambridge University Press, 2014b.

David Z Albert. *Time and Chance*. Harvard University Press, 2000.

David Z Albert. The difference between the past and the future. In *After Physics*, pages 31–70. Harvard University Press, 2015.

Adam Thomas Bales. *Decision and dependence: A defence of causal decision theory*. PhD thesis, University of Cambridge, 2017.

[27]For helpful discussion and feedback thanks to Arif Ahmed, Adam Bales, Richard Bradley, Cian Dorr, Jeremy Goodman, James Joyce, Boris Kment, Adam Kolber, Agustín Rayo, Alexander Sandgren, Toby Solomon, Timothy L. Williamson, two anonymous *Analysis* referees, and audiences at the University of Massachusetts at Amherst, the CUNY Graduate Center, Rutgers University, and the 2021 APA Pacific Division Meeting.

Robert Bassett. A critique of benchmark theory. *Synthese*, 192(1):241–267, 2015. doi: 10.1007/s11229-014-0566-3.

David Braddon-Mitchell. Lossy laws. *Noûs*, 35(2):260–277, June 2001. doi: 10.1111/0029-4624.00296.

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

Allan Gibbard and William Harper. Counterfactuals and two kinds of expected utility. In A. Hooker, J. J. Leach, and E. F. McClennen, editors, *Foundations and Applications of Decision Theory*, pages 125–162. D. Reidel, 1978.

Christopher Hitchcock. What is the 'cause' in causal decision theory? *Erkenntnis*, 78(S1):129–146, December 2013. doi: 10.1007/s10670-013-9440-9.

Richard Jeffrey. *The Logic of Decision*. University of Chicago Press, 1965.

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

James M. Joyce. Arif Ahmed: Evidence, Decision and Causality. *Journal of Philosophy*, 113(4):224–232, 2016. doi: 10.5840/jphil2016113413.

Boris Kment. Decision, causality, and pre-determination. Unpublished.

Douglas N. Kutach. The entropy theory of counterfactuals. *Philosophy of Science*, 69(1):82–104, 2002. doi: 10.1086/338942.

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

David Lewis. Counterfactual dependence and time's arrow. In *Philosophical Papers Volume II*. 2014.

Barry Loewer. Counterfactuals and the second law. In Huw Price and Richard Corry, editors, *Russell's Republic: The Place of Causation in the Constitution of Reality*. Oxford University Press, Oxford, 2007.

Barry Loewer. The consequence argument meets the mentaculus. Unpublished.

Daniel Nolan. Causal Counterfactuals and Impossible Worlds. In Helen Beebee, Christopher Hitchcock, and Huw Price, editors, *Making a Difference*, pages 14–32. Oxford, UK: Oxford University Press, 2017.

Robert Nozick. Newcomb's problem and two principles of choice. In Nicholas Rescher, editor, *Essays in Honor of Carl G. Hempel*, pages 114–146. Reidel, 1969.

Alexander Sandgren and Timothy Luke Williamson. Determinism, Counterfactuals, and Decision. *Australasian Journal of Philosophy*, forthcoming. doi: 10.1080/00048402.2020.1764073.

Wolfgang Schwarz. Decision-making under determinism. https://www.umsu.de/wo/2014/603, January 2014.

Brian Skyrms. *Causal necessity: a pragmatic investigation of the necessity of laws*. Yale University Press, New Haven, 1980.

Toby Charles Penhallurick Solomon. Causal decision theory's predetermination problem. *Synthese*, October 2019. doi: 10.1007/s11229-019-02425-0.

Robert C. Stalnaker. Letter to David Lewis. In William L. Harper, Robert Stalnaker, and Glenn Pearce, editors, *IFS: Conditionals, Belief, Decision, Chance and Time*, The University of Western Ontario Series in Philosophy of Science, pages 151–152. Springer Netherlands, Dordrecht, 1981. doi: 10.1007/978-94-009-9117-0_7.

Ralph Wedgwood. Gandalf's solution to the Newcomb problem. *Synthese*, 190(14):2643–2675, 2013. doi: 10.1007/s11229-011-9900-1.

Timothy Luke Williamson and Alexander Sandgren. Law-Abiding Causal Decision Theory. *British Journal for the Philosophy of Science*, forthcoming.