Decision and Foreknowledge

J. Dmitri Gallow*

Dianoia Institute of Philosophy · April 1st, 2021

1. A lot of philosophical work on decision theory focuses on the question of how you should deliberate about what to do when your decision requires you to think of that very deliberation as embedded in the world's causal order. (If the deliberation is so embedded, then the *outcome* of your deliberation is in-principle predictable.)

2. Today I want to focus on another philosophical aspect of decision theory: how should you deliberate about what to do when your decision requires you to think of the future as (metaphysically) *settled*.

   ▷ I'm going to focus on these two decisions:

     STICKER
     Under the Christmas tree are two gifts: one for you, and one for your sister. You know that one contains a toy, and the other contains a lump of coal, but you don't know which is which. You are absent-mindedly placing decorative stickers on the gifts. Before you place the reindeer sticker, an oracle informs you: the gift on which you'll put the reindeer sticker contains the toy.

     INADMISSIBLE FOREKNOWLEDGE
     Before a fair coin is flipped, you're offered a bet which pays out $150 if the coin lands heads, and costs $50. Before you choose whether to take the bet, the oracle tells you: the coin will land on tails.

3. Many have suggested that these kinds of decisions pose a distinctive and novel threat to causal decision theory (CDT).

   ▷ For instance, LEWIS (1981) says that they are "much more problematic for decision theory than the Newcomb problems"

   ▷ PRICE (2012) thinks that these kinds of decisions show that we must be subjectivists about causation itself.

   ▷ HITCHCOCK (2016) and STERN (forthcoming) think that CDT must be modified to deal with these kinds of decisions.

   ▷ SPENCER (2020) argues against the possibility of a problematic kind of foreknowledge, explicitly in order to avoid decisions in which he thinks CDT gives the wrong advice.

4. My thesis: foreknowledge poses no new problems for CDT. The supposed problem cases are either...

   ▷ ...not problems,

   ▷ ...not problems for CDT, but rather problems for our theory of subjunctive supposition, or

   ▷ ... not *new* problems for CDT. [I won't be talking about this final class of problems here today; if you're interested, see the paper.]

5. Nonetheless, these decisions teach and illustrate important lessons for causalists. In particular, ...

   ▷ ...they give a vivid illustration of the lesson that your intuitive judgements about instrumental value are not to be trusted when you suffer from an illusion of control; and

▷   ...they teach us to not confuse the probability that an outcome out result, were you to choose A, with the chance of that outcome, conditional on your choosing A.

[I think that these kinds of decisions have other lessons to teach us, too, but I won't be talking about them today. See the paper if you're interested.]

1   CAUSAL DECISION THEORY

6.   When you're making a decision, there are:

▷   some available acts, $\mathcal{A} = \{A_1, A_2, \ldots, A_M\}$, between which you must choose, and

▷   some ways the world might be, 'worlds', $\mathcal{W} = \{w_1, w_2, \ldots, w_N\}$

And you have:

▷   a *credence* distribution, C, over subsets of $\mathcal{W}$, and

▷   a *desire* function, $\mathcal{D}$. For each $w \in \mathcal{W}$, this function tells you the degree to which you desire that w is actual, $\mathcal{D}(w)$.

In addition, the causal decision theorist says that you must consider what each act $A \in \mathcal{A}$ would bring about, were you to select it.

▷   We model this with a function, $would_A$, where $would_A(w)$ is a probability distribution over subsets of $\mathcal{W}$.

▷   The interpretation is that $would_A(w)(w^*)$ tells us how likely you think it is that w^* would result, were you to choose A at w.

▷   Since A would certainly result, were you to choose A, we should stipulate that $would_A(w)(A) = 1$.

▷   Functions like $would_A$ are known as *imaging* functions.

7.   Fix some enumeration of the worlds in $\mathcal{W}$, and let C be a $1 \times N$ vector where the ith column is your credence in w_i. Let $\mathcal{D}$ be an $N \times 1$ vector where the ith row is $\mathcal{D}(w_i)$. And let $would_A$ be an $N \times N$ matrix, where the ith row and jth column is $would_A(w_i)(w_j)$. Then, CDT says to choose an option which maximises 'utility', $\mathcal{U}$, where

$$\mathcal{U}(A) \overset{\text{def}}{=} C \cdot would_A \cdot \mathcal{D}$$

▷   That is: the utility of A consists of three ingredients: your credences, your desires, and your opinions about what would happen, were you to choose A. Multiply the ingredients together, and you get the utility of A.

▷   In contrast, evidential decision theory (EDT) says to choose an option which maximises 'news-value', $\mathcal{V}$, where

$$\mathcal{V}(A) \overset{\text{def}}{=} C \mid A \cdot \mathcal{D}$$

('$C \mid A$' is your credences *conditioned* on the proposition that you've performed A. That is: $C \mid A$ is a $1 \times N$ vector whose ith column is $C(w_i \mid A)$.)

▷   Utility is sometimes explained by saying that $\mathcal{U}(A) = C_A \cdot \mathcal{D}$, where $C_A \overset{\text{def}}{=} C \cdot would_A$ is your credence function *imaged* on A. However, I think that CDT's commitments about the instrumental value of acts are clearer if we instead group the imaging function with your *desires*.

▷ $\mathcal{D}_A \overset{\text{def}}{=} would_A \cdot \mathcal{D}$—your desires imaged on A—tells you how desirable the world would be, were you to choose A. Then, utility is just your expectation of this quantity.

▷   So: CDT says that an act is instrumentally valuable to the extent that the world would be desirable, were you to choose it

8.   A nice illustration of this commitment:

NO DIFFERENCE
Before you are two boxes. You may either take the box on the left, 'Lefty', or the box on the right, 'Righty'. There is no difference between them. If it was predicted that you'd take Lefty, then there's $100 in both boxes. If it was predicted that you'd take Righty, then there's nothing in either box. You take these predictions to be very reliable.

	Predicted Lefty	Predicted Righty
Take Lefty	$100	$0
Take Righty	$100	$0

▷   In this decision, if you find yourself inclining towards taking Lefty, then you will give yourself evidence that it was predicted that you'd take Lefty.

And this is evidence that you're going to get $100. On the other hand, if you find yourself inclining towards taking Righty, then you will give yourself evidence that it was predicted that you'd take Righty. And this is evidence that you're not getting any money.

▷ In this decision, your *rational credence* that there's money in the boxes is under your control. By taking Lefty, you can raise your rational credence that there's money there. And by taking Righty, you can lower your rational credence that there's money. This can make it *feel* as though you have control over whether the money is there. But *ex hypothesi*, you have no control over whether the money's in the boxes. This is an *agential illusion* of control.

▷ There is a strong intuition that you have instrumental reason to take Lefty. But causalists should diagnose this intuition as a consequence of the agential illusion of control.

9. To correct for the illusion of control, we may consider the decision from a better-informed, third-personal perspective.

▷ Suppose that your friend is choosing between Lefty and Righty, and you can look inside the boxes.

▷ From this perspective, it no longer appears that there is any more instrumental value in taking Lefty than there is in taking Righty.

▷ In fact, from this perspective, a preference for taking Lefty can appear irrational. To appreciate this, consider a new decision in which, in order to take Lefty, your friend must pay $90.

	Predicted Lefty	Predicted Righty
Take Lefty	$10	-$90
Take Righty	$100	$0

Then, evidential decision theory will advise your friend to pay the $90. From your perspective, it appears clear that paying the $90 only makes things worse, and has negative instrumental value.

Lesson #1 When you have control over your rational credence that ϕ, but you know for sure that you do not have control over whether ϕ, your intuitive judgements about instrumental value can lead you astray by conflating control over your epistemic state with control over the world.

In these cases, you should consider what instrumental value a choice has when viewed from a better informed, third-personal perspective.

▷ As Lewis puts it: it is irrational to "counsel [a] policy of managing the news so as to get good news about matters which you have no control over" (Lewis, 1981, p. 5)

2 Managing the News from the Future

10. Return to the decision

STICKER
Under the Christmas tree are two gifts: one for you, and one for your sister. You know that one contains a toy, and the other contains a lump of coal, but you don't know which is which. You are absent-mindedly placing decorative stickers on the gifts. Before you place the reindeer sticker, an oracle informs you: the gift on which you'll put the reindeer sticker contains the toy.

	You gifted toy	Sister gifted toy
Sticker on yours	☺	☹
Sticker on sister's	☺	☹

▷ You take the world's causal structure to be like this:
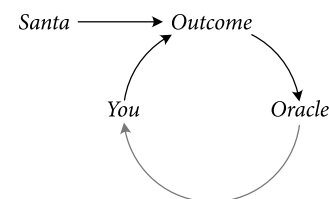
▷ *Santa* says whether Santa gifted the toy to you or your sister

▷ *You* says whether you put the sticker on you or your sister's gift. It may or may not be causally influenced by *Oracle*.

▷ *Outcome* says who got the toy and which gift has the sticker on it. It is causally determined by *Santa* and *You*.

3

- ▷ *Oracle* says whether the oracle tells you that the gift with the sticker has the toy or the coal. It is causally determined by *Outcome*.

11. In this decision, putting the sticker on yours or your sister's gift wouldn't make any difference to who gets the toy. So CDT says that you have no more instrumental reason to place the sticker on your gift than you have to place the sticker on your sister's gift.

12. Hitchcock (2016) thinks that this is the wrong result. He proposes a modification of CDT which will tell you to place the sticker on your gift in this decision.

   - ▷ To appreciate the revision, let E be your ordinary evidence, let F be your foreknowledge, and let C_0 be your *prior* credences. Then, we can suppose that your current credences are just your prior credences conditioned on your total evidence, $C = C_0 \mid EF$. Then, CDT says that you should choose an act which maximises $\mathcal{U}$:

   $$\mathcal{U}(A) = (C \mid EF) \cdot would_A \cdot \mathcal{D}$$

   - ▷ In contrasts, Hitchcock says that you should choose an act which maximises $\mathcal{H}$:

   $$\mathcal{H}(A) = (C_0 \mid E \cdot would_A) \mid F \cdot \mathcal{D}$$

   That is: you should take your prior credences, condition them on your ordinary evidence E, image that function on the performance of A, and only *then* condition on your foreknowledge F.

   - ▷ In STICKER, we will have:

   $$\mathcal{H}(\text{sticker on yours}) = ☺$$
   $$\mathcal{H}(\text{sticker on sister's}) = ☹$$

13. Evidentialists should be happy with this verdict, but if we've learnt **Lesson #1**, we should be more cautious.

   - ▷ This is a decision in which you have control over your rational credence that Santa gifted you the toy, but you know for sure that you do not have control over whether Santa gifted you the toy. So, in STICKER, you are subject to an agential illusion of control.

- ▷ To correct for this illusion, we should consider the decision from a better-informed, third-personal perspective. Suppose that it is your sister making the decision, she wants you to get the gift, and you can see inside of the packages.

- ▷ From this perspective, it no longer appears that there is any instrumental value in placing the sticker on your gift.

- ▷ In fact, from this perpsective, a preference for placing the sticker on your gift can appear irrational. To appreciate this, consider a new decision in which, in order to place the sticker on your gift, your sister has to pay.

	You get toy	Sister gets toy
Sticker on yours	☺	🤑
Sticker on sister's	☺	☹

Then, $\mathcal{H}(\text{sticker on yours}) = ☺$ and $\mathcal{H}(\text{sticker on sister's}) = ☹$, so Hitchcock will tell your sister to pay to put the sticker on your gift. But from your perspective, it appears clear that paying only makes things worse, and has negative instrumental value.

14. So: CDT says that you have no instrumental reason to place the sticker on one gift over the other. But causalists should be happy with this verdict. There is a strong inclination to place the sticker on your own gift, but causalists should diagnose this inclination as a result of an agential illusion of control (just as they do in NO DIFFERENCE).

3   Foreknowledge and Chance

15. Call foreknowledge *inadmissible* if, when you have that information, you know for sure that the chance of ϕ is x, but your credence in ϕ should not be x.

16. Hall (1994), Meacham (2010), and Spencer (2020) all say that inadmissible foreknowledge is *impossible*.

17. Spencer says this in part because he thinks that inadmissible foreknowledge would lead to counterexamples to CDT. For instance,

INADMISSIBLE FOREKNOWLEDGE

Before a fair coin is flipped, you're offered a bet which pays out $150 if the coin lands heads, and costs $50. Before you choose whether to buy the bet, the oracle says: the coin will land on tails.

	Heads	Tails
Buy the bet	$100	-$50
Don't buy the bet	$0	$0

▷ It's natural to describe this case by saying: you know for sure that the chance of heads is 50%, but you should be less than 50% confident in heads. However, Spencer disagrees. He wants to say that, precisely because it's rational for you to be less than 50% confident in heads, the *chance* of heads is less than 50%.

▷ One of his reasons for wanting to say this is that Spencer thinks it is irrational to buy the bet (I agree), but he thinks that, if the chance of the coin landing heads is 50%, then CDT will advise you to buy the bet (I disagree).

18. First, let's understand why Spencer thinks CDT will advise you to buy the bet.

    (a) In this decision, there are four possibilities:

    w_{HB}, in which the coin lands *h*eads and you *b*uy

    w_{HN}, in which the coin lands *h*eads and you do *n*ot buy

    w_{TB}, in which the coin lands *t*ails and you *b*uy

    w_{TN}, in which the coin lands *t*ails and you do *n*ot buy

    (b) Given your foreknowledge, your credences in these possibilities are as shown:

    $C(w_{HB}) = 0$
    $C(w_{HN}) = 0$
    $C(w_{TB}) = C(B)$ (your credence that you bet)
    $C(w_{TN}) = C(T)$ (your credence that you do not bet)

    (c) We can assume that your desires are linear in dollars, so that

    $\mathcal{D}(w_{HB}) = 100$

$\mathcal{D}(w_{HN}) = 0$
$\mathcal{D}(w_{TB}) = -50$
$\mathcal{D}(w_{TN}) = 0$

(d) Spencer *additionally* assumes that, were you to buy the bet, it's 50% likely that the coin would land heads, no matter which world you're at.

$$would_B = \begin{array}{c} \\ w_{HB} \\ w_{HN} \\ w_{TB} \\ w_{TN} \end{array} \begin{array}{cccc} w_{HB} & w_{HN} & w_{TB} & w_{TN} \\ \left[\begin{array}{cccc} 50\% & 0 & 50\% & 0 \\ 50\% & 0 & 50\% & 0 \\ 50\% & 0 & 50\% & 0 \\ 50\% & 0 & 50\% & 0 \end{array} \right] \end{array}$$

(e) Then, it follows that

$$\mathcal{U}(B) = 25 \qquad \text{and} \qquad \mathcal{U}(N) = 0$$

So CDT says you should buy the bet.

19. The crucial step comes at (18d), when Spencer assumes—following LEWIS (1980, 1981)—that, for any w, $would_B(w)(-) = Ch_w(- \mid B)$. As RABINOWICZ (1982, 2009) notes, this assumption conflicts with **Strong Centering**.

Strong Centering If w is a world at which you choose A, then were you to choose A at w, w is the world which would result.

$$\text{if } A \text{ is true at } w, \text{ then } would_A(w)(w) = 100\%$$

▷ Both LEWIS (1979) and STALNAKER (1968) validate the analogue of **Strong Centering** for counterfactuals: $A \wedge C \vdash A \square\rightarrow C$.

▷ To get from this principle to the one I've called **Strong Centering** above, assume that $w \models A \square\rightarrow C \Leftrightarrow would_A(w)(C) = 100\%$.

▷ But LEWIS (1986) rejects **Strong Centering** in his decision theory.

20. I accept **Strong Centering**, but it won't help with INADMISSIBLE FORE-KNOWLEDGE.

▷ If we impose **Strong Centering** on the imaging function $would_B$, but otherwise leave everything unchanged, we'll get

$$
would_B = \begin{array}{c} \\ w_{HB} \\ w_{HN} \\ w_{TB} \\ w_{TN} \end{array} \begin{array}{cccc} w_{HB} & w_{HN} & w_{TB} & w_{TN} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 50\% & 0 & 50\% & 0 \\ 0 & 0 & 1 & 0 \\ 50\% & 0 & 50\% & 0 \end{array} \right] \end{array}
$$

and the utility of taking the bet will become $\mathcal{U}(B) = 25 - 75 \cdot C(B)$. So, as you give yourself more evidence that you'll take the bet, taking the bet will have more expected instrumental value than not taking it. And, as you give yourself more evidence that you'll *not* take the net, taking the bet will have less expected instrumental value than not taking it. But this seems wrong. It seems like, *no matter how confident you are that you'll take the bet*, you shouldn't take it.

21. In my view, CDT's difficulties with INADMISSIBLE FOREKNOWLEDGE are just an instance of a well-known issue with our theories of subjunctive supposition. Suppose I offer you a bet on whether a flipped coin lands heads, you refuse, and the coin lands on heads. Sidney Morgenbesser observed that, in this context, it sounds true to say:[1]

(MC) If you had taken the bet, you would have won

and that's so, *even though* at the moment of choice, it was a matter of chance whether the coin lands heads or not.

22. There is a general lesson to be learnt here: when we make a subjunctive supposition that we choose differently, we should hold fixed things which are *causally independent* of our choice—even if those things are matters of chance at the moment of choice.

In particular, the causal decision theorist should accept:

Causal Independence If whether ϕ is causally independent of your choice, then ϕ would not change its truth-value, were you to choose an $A \in \mathcal{A}$.

$$
would_A(w)(\phi) = \begin{cases} 1 & \text{if } \phi \text{ is true at } w \\ 0 & \text{if } \phi \text{ is false at } w \end{cases}
$$

[1] See (SLOTE, 1978, fn. 33).

23. **Causal Independence** implies that the imaging function $would_B$ is:

$$
would_B = \begin{array}{c} \\ w_{HB} \\ w_{HN} \\ w_{TB} \\ w_{TN} \end{array} \begin{array}{cccc} w_{HB} & w_{HN} & w_{TB} & w_{TN} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right] \end{array}
$$

And, now—as you can verify for yourself—taking the bet has a negative utility,

$$\mathcal{U}(B) = -50 \qquad \text{and} \qquad \mathcal{U}(N) = 0$$

so CDT will tell you to not take the bet.

24. Once we've imposed **Causal Independence**, CDT will reason about the decision like this: if you were to buy the bet, this wouldn't make any difference to how the coin lands. And you know that the coin lands tails—so you know that the bet is a loser. So, you shouldn't buy it.

Lesson #2 The probability that ϕ would result, were you to choose A, is not always just the chance that ϕ would has, conditional on your choosing A.

If you choose A and ϕ is true, then ϕ would be true, were you to choose A.

And if ϕ is causally independent of your choice, then ϕ wouldn't change its truth-value, were you to choose A.

4   IN SUMMATION

25. Foreknowledge poses no new problems for CDT

▷ Decisions like STICKER are not problems for CDT, because causalists should think that CDT gives the correct advice in those cases.

▷ Decisions like INADMISSIBLE FOREKNOWLEDGE *are* problems, but they are problems for our theories of subjunctive supposition, not for CDT

26. These kinds of decisions teach us—or vividly illustrate for us—two important lessons about instrumental value.

- ▷ Lesson #1: when you have control over *what to believe* about whether ϕ, but no control over *whether* ϕ, intuitions about instrumental value can be distorted by an agential illusion of control.

- ▷ Lesson #2: the probability an outcome would result, were you to choose A, is not always the chance of that outcome, conditional on *A*.

References

Hall, Ned. 1994. "Correcting the Guide to Objective Chance." *Mind*, vol. 103 (412): 505–517.

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

Lewis, David K. 1979. "Attitudes De Dicto and De Se." *The Philosophical Review*, vol. 88 (4): 513–543.

—. 1980. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*, Richard C. Jeffrey, editor, vol. II, 263–293. University of California Press, Berkeley.

—. 1981. "Causal Decision Theory." *Australasian Journal of Philosophy*, vol. 59 (1): 5–30.

—. 1986. "Postscripts to 'Causation.'" In *Philosophical Papers*, vol. II. Oxford University Press, New York.

Meacham, Christopher J. G. 2010. "Binding and Its Consequences." *Philosophical Studies*, vol. 149 (1): 49–71.

Price, Huw. 2012. "Causation, Chance, and the Rational Significance of Supernatural Evidence." *Philosophical Review*, vol. 121 (4): 483–538.

Rabinowicz, Wlodek. 2009. "Letters from Long Ago: On Causal Decision Theory and Centered Chances." In *Logic, Ethics, and All That Jazz—Essays in Honour of Jordan Howard Sobel*, L-G. Johansson, editor, vol. 56, 247–273. Uppsala Philosophical Studies.

Rabinowicz, Włodzimierz. 1982. "Two Causation Decision Theories: Lewis vs Sobel." In *Philosophical Essays Dedicated to Lennart Åqvist on His Fiftieth Birthday*, Tom Pauli, editor, vol. 34, 299–321. Uppsala Philosophical Studies, Uppsala.

Slote, Michael A. 1978. "Time in Counterfactuals." *The Philosophical Review*, vol. 87 (1): 3–27.

Spencer, Jack. 2020. "No Crystal Balls." *Noûs*, vol. 54 (1): 105–125.

Stalnaker, Robert C. 1968. "A Theory of Conditionals." In *Studies in Logical Theory*, N. Rescher, editor, chap. 4, 98–112. Oxford University Press, Oxford.

Stern, Reuben. forthcoming. "An Interventionist's Guide to Exotic Choice." *Mind*.