

ARTICLE

Unspecific Evidence and Normative Theories of Decision

Rhys Borchert 

University of Arizona, Tucson, AZ, USA

Email: rhysborchert@arizona.edu

(Received 12 October 2022; revised 31 January 2023; accepted 2 May 2023; First published online 05 June 2023)

Abstract

The nature of evidence is a problem for epistemology, but I argue that this problem intersects with normative decision theory in a way that I think is underappreciated. Among some decision theorists, there is a presumption that one can always ignore the nature of evidence while theorizing about principles of rational choice. In slogan form: decision theory only cares about the credences agents actually have, not the credences they should have. I argue against this presumption. In particular, I argue that if evidence can be unspecific, then an alleged counterexample to causal decision theory fails. This implies that what theory of decision we think is true may depend on our opinions regarding the nature of evidence. Even when we are theorizing about subjective theories of rationality, we cannot set aside questions about the objective nature of evidence.

Keywords: unspecific evidence; rational choice; decision theory; causal decision theory; Newcomb's problem

Philosophers tend to agree that a person's evidence constrains what doxastic attitudes they should take toward a proposition. What is disputed is the nature and extent of such constraints. When we are considering the evidential constraints on fine-grained doxastic attitudes, e.g., credences, some philosophers maintain that a body of evidence can be *unspecific*. A body of evidence that is unspecific does not justify forming a precise credence toward a proposition.¹ Instead of a precise credence, such a body of evidence supports some kind of imprecise doxastic attitude.^{2,3}

¹I have defined unspecific evidence such that it entails the normative thesis that forming a precise credence on the basis of such evidence would be irrational, so the pertinent question is whether a body of evidence can be unspecific. If one were to separate the definition of unspecific evidence from the normative thesis, then the pertinent question could instead be whether unspecific evidence requires forming an imprecise attitude toward a proposition. Carr (2020), for instance, frames this dispute this way.

²What exactly this imprecise attitude should be, and how it should be modeled, can vary. A common proposal is that an agent ought to have an imprecise credence toward a proposition where this doxastic attitude is represented by a set of probability distributions (see Joyce 2010; Levi 1974; van Fraassen 1990). Another proposal is that an agent ought to form a "thick confidence" toward a proposition, which is modeled with an interval-valued probability function (see Sturgeon 2020).

³Note that this dispute is tangential to the debate between uniqueness and permissivism. Both could accept that evidence can be unspecific. The defender of uniqueness would say that there is always only

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Accepting unspecific evidence introduces a conflict between straightforward conceptions of *evidentialism* and *probabilism*.

Evidentialism: A doxastic attitude *D* is rational for agent *S* at time *T* if, and only if, *D* fits the evidence that *S* possesses at *T*.

Probabilism: An agent is rational at time *T* only if their doxastic state at *T* can be modeled with a single probability distribution.⁴

The combination of evidentialism and unspecific evidence implies the falsity of probabilism. If one were to accept the combination of evidentialism and unspecific evidence, this would not imply that there are no cases where having probabilistically incoherent credences renders a person irrational. Instead, it would imply that when an agent is rationally required to have probabilistically coherent credences, it is because an agent's evidence is specific enough to justify precise credences.

Suspicion of probabilism on the basis of unspecific evidence is nothing new. However, I think that the relevance of unspecific evidence to decision theory is currently underappreciated.⁵ I argue that a particular challenge to *Causal Decision Theory* (CDT) fails on the assumption that evidence is unspecific.⁶ My aim is not simply to defend CDT, but also to show that in analyzing what theory of rational choice to accept, we cannot always remain neutral regarding the status of unspecific evidence. Even when we are concerned with *subjective* theories of rationality, questions about the nature of evidence are sometimes front and center.⁷ Also, my arguments may provide some motivation for causal decision theorists to consider how to incorporate imprecise decision principles into their overall decision-theoretic framework.

In section 1, I review some examples that support the idea of unspecific evidence. Of particular note is an example from Sturgeon (2020) which seems to show that evidence can be specific with regard to conditional credences but unspecific with regard to unconditional credences. In section 2, I apply considerations from section 1 to *Newcomb's Problem*. This leads to the (perhaps surprising) result that whether the

one rationally permissible imprecise attitude to form on the basis of unspecific evidence, while the permissivist would say that there can be multiple, equally rational imprecise attitudes to form on the basis of unspecific evidence.

⁴Terms like "probabilism" and "Bayesianism" sometimes differ in their usage. Often people are labeled "probabilists" or "Bayesians" when they endorse the thesis that a rational agent's doxastic state can be modeled by *either* a single probability function *or* a set of probability functions; they are dubbed "imprecise probabilists" or "imprecise Bayesians." However, I will restrict my use of "probabilism" to the thesis that an agent must be modeled by a single probability distribution.

⁵I do not mean to imply that this issue is totally underappreciated, since this would seem to overlook the literature on decision principles for agents with imprecise credences (e.g., Bradley 2019; Bradley and Steele 2014; 2016; Chandler 2014; Joyce 2010; Moss 2015; Seidenfeld 1988; 2004; Steele 2021; Troffaes 2007; Weatherson *ms*; Williams 2014). Instead, what I think is sometimes not appreciated is that the issue of unspecific evidence does not simply go away when we shift our focus to decision principles for agents with precise credences. Ideally, my arguments in this paper would spur more attention to decision principles with imprecise credence from those who typically only engage with "precise" theories of decision, like CDT and EDT.

⁶My discussion centers around a case from Spencer and Wells (2019), but I think the considerations in this paper also apply to cases from Ahmed (2014) and Oesterheld and Conitzer (2021).

⁷Arguably, Briggs' (2010) arguments also support this general claim since they connect the plausibility of thirdism or halfism in *Sleeping Beauty* to whether one accepts EDT or CDT.

principle of maximization of utility applies in a particular case may depend on which theory of maximization of utility is accepted. In section 3, I apply the considerations from the first two sections to show how unspecific evidence causes trouble for an alleged counterexample to CDT from Spencer and Wells (2019). I address possible objections in section 4 and conclude in section 5.

1. Unspecific Evidence

Taken literally, probabilism seems psychologically unrealistic. Aside from certain situations (e.g., where a body of evidence is strongly symmetric, or where the precise values of objective probabilities are known), it seems implausible to think that a person's fine-grained doxastic attitude toward a proposition picks out a unique real number. Yet rarely do those who tend to favor probabilism, e.g., Bayesians, insist that probabilism should be understood in such a literal manner. Instead, the claim from the Bayesian who accepts probabilism is often that a rational agent's doxastic state ought to be able to be *modeled* with a probability distribution. Models often make idealizing assumptions. We know that taken literally it is psychologically unrealistic to assume that an agent's doxastic state singles out a unique probability distribution in every situation. But we also know that taken literally it is physically unrealistic to assume that planets are perfectly spherical.

So while there are some who oppose probabilism on psychological grounds, I think a more forceful opposition comes from those who oppose probabilism on evidential grounds. The objection is not that conforming with the demands of probabilism is psychologically unrealistic, but rather that sometimes conforming with the demands of probabilism is irrational. Isaac Levi opposed probabilism on these grounds.

I am not concerned to speculate on our capacities for meeting strict Bayesian requirements for credal (and value) rationality. But even if men have, at least to a good degree of approximation, the abilities Bayesians attribute to them, there are many situations where, in my opinion, rational men *ought not* to have precise utility functions and precise probability judgments.⁸

As does James Joyce.

[T]he proper response to symmetrically ambiguous or incomplete evidence is not to assign probabilities symmetrically, but to refrain from assigning precise probabilities at all... It is not just that sharp degrees of belief are psychologically unrealistic (though they are). Imprecise credences have a clear epistemological motivation: they are the proper response to unspecific evidence.⁹

So multiple authors think that conforming with the demands of probabilism can be irrational, but what exactly are these situations? One illustrative example comes from Elga (2010).¹⁰

⁸Levi (1974: 394–95, emphasis in original).

⁹Joyce (2005: 171).

¹⁰Note that after providing this example, Elga proceeds to argue in favor of probabilism. See also Elga (2012), Chandler (2014), and Bradley and Steele (2014) for discussion.

Strange Items: A stranger approaches you on the street and starts pulling out objects from a bag. The first three objects he pulls out are a regular-sized tube of toothpaste, a live jellyfish, and a travel-sized tube of toothpaste. To what degree should you believe that the next object he pulls out will be another tube of toothpaste?¹¹

That an agent is required to form a non-arbitrary, precise credence in *Strange Items* seems ridiculous. I do not think that this seems ridiculous because it is *too difficult* to form a non-arbitrary, precise credence. Consider if, instead, the stranger said that there were nine dice in his bag: three six-sided, three eight-sided dice, and three ten-sided. What's your credence that if all the dice are rolled, the numbers will all be less than 3? I doubt that a number instantly popped into your head, but I expect that you would agree that this body of evidence does support a non-arbitrary, precise credence. If I gave you a pen and paper, you'd be able to figure it out. The lack of a pen and paper in *Strange Items* is not the issue. Rather the issue seems to be the lack of proper evidential support. Your evidence in *Strange Items* is not specific enough to warrant a precise credence, so forming a precise credence would not respect your evidence. And we should all agree that rational agents respect their evidence.

Note that it is not the case that a body of evidence is either completely specific or completely unspecific. It may be the case that a body of evidence is specific enough to warrant precise credences in some propositions, but not specific enough to warrant precise credences in other propositions. There are also cases where precise conditional credences are warranted but precise unconditional credences are not warranted. Consider the following example from Scott Sturgeon.

Red-Spotted Ball: There is an opaque box in front of you. You know that the box is filled with colored balls, some of which have red spots. You also know that there are exactly three blue balls in the box, one of which has a red spot.¹²

Suppose that you reach into the opaque box and grab a ball. First question: what should your conditional credence be that the ball you grabbed is a red-spotted ball given that the ball you grabbed is blue? First answer: $\frac{1}{3}$. Second question: what should your unconditional credence be that the ball you grabbed is a red-spotted ball? Second answer: I have no idea. And it seems that I have no idea because the evidence in this case is not specific enough to uniquely determine a precise unconditional credence that I ought to have.

I hope that the considerations of this section have shown that the idea that evidence can be unspecific is intuitively compelling. My aim in this paper was not to argue in favor of unspecific evidence, but rather to show that disputes over the correct theory of rational choice may not be independent of the dispute over unspecific evidence.

2. Unspecific Evidence and Newcomb's Problem

If one takes seriously the idea that evidence can be unspecific, then *Newcomb's Problem* deserves a second look.

¹¹Elga (2010: 1).

¹²Sturgeon (2020: 78).

Newcomb’s Problem: An agent has a choice to either take both a transparent box and an opaque box (“Two-Box”) or take only the opaque box (“One-Box”). The transparent box contains \$1,000 (*k*). The opaque box contains \$1,000,000 (*m*) if, and only if, a reliable predictor predicted that the agent would choose One-Box, otherwise the opaque box is empty.

Let P1 be the state where the predictor predicted that the agent would choose One-Box and P2 be the state where the predictor predicted that the agent would choose Two-Box. *Newcomb’s Problem* can be represented by the following decision matrix:

	P1	P2
One-Box	<i>m</i>	\$0
Two-Box	<i>m + k</i>	<i>k</i>

Here’s (a version of) the story many of us have heard before. *Newcomb’s Problem* reveals a conflict between *Evidential Decision Theory* (EDT) and *Dominance*. EDT says that when a rational agent faces a decision, he selects the option that maximizes expected utility according to the following formula:

$$EEU(O) = \sum_i U(O \ \& \ S_i)Cr(S_i|O)$$

where {*S*_{*i*}} is a partition of causally option-independent possible states of the world, *U* (*O* & *S*_{*i*}) is the agent’s subjective utility in the option-state pair *O* & *S*_{*i*}, and *Cr*(*S*_{*i*}|*O*) is the agent’s conditional credence in *S*_{*i*} conditional on *O*. “*EEU*(*O*)” should be read as “the evidential expected utility of *O*.”

As long as the predictor is slightly better than chance, EDT will say that the rational choice in *Newcomb’s Problem* is One-Box. For example, suppose that the predictor is 51% accurate and the agent knows this. This leads to the following evidential expected utilities:

$$EEU(\text{One-Box}) = 0.51(m) + 0.49(\$0) = \$510,000$$

$$EEU(\text{Two-Box}) = 0.49(m + k) + 0.51(k) = \$491,000$$

Given that *EEU*(One-Box) > *EEU*(Two-Box), EDT says that one ought to choose One-Box. This recommendation from EDT conflicts with *Dominance*:

An option *O* *dominates* an option *Q* if, and only if, for all causally option-independent possible states of the world {*S*_{*i*}}, *U*(*O* & *S*_{*i*}) > *U*(*Q* & *S*_{*i*}). *Dominance* says that when *O* dominates *Q*, then it is irrational to choose *Q* over *O*.¹³

¹³The “causally option-independent” qualification is important. If it is absent, then *Dominance* would give silly recommendations (see Jeffrey (1983: 8–9) or Joyce (1999: 114–19) for discussions on this point). If we replace it with “probabilistically option-independent” then *Dominance* would not conflict with EDT in *Newcomb’s Problem* since the states are probabilistically correlated with the options.

In *Newcomb's Problem*, Two-Box dominates One-Box. So Dominance says that One-Box is irrational. Thus, EDT and Dominance are in conflict.

Enter the hero of the story: CDT. CDT says that when a rational agent faces a decision, they select the option that maximizes expected utility according to the following formula¹⁴:

$$CEU(O) = \sum_i U(O \& S_i)Cr(O \Rightarrow S_i)$$

where $Cr(O \Rightarrow S_i)$ is the agent's unconditional credence in the subjunctive conditional "if I were to choose O, then S_i would be the case." " $CEU(O)$ " should be read as "the causal expected utility of O."

CDT does not conflict with Dominance. When the states of a decision are causally option-independent, an informed, rational agent's unconditional credence in $O \Rightarrow S_i$ ought to be equal to their unconditional credence in S_i (O has no causal influence on the state so the probability of "S if I were to O" is simply the probability of S). So in *Newcomb's Problem*, we have the following credences:

$$Cr(\text{One-Box} \Rightarrow P1) = Cr(\text{Two-Box} \Rightarrow P1) = Cr(P1)$$

$$Cr(\text{One-Box} \Rightarrow P2) = Cr(\text{Two-Box} \Rightarrow P2) = Cr(P2)$$

Given that the agent's choice cannot causally affect the contents of the opaque box, the agent's credence that taking only the opaque box will cause there to be money in the opaque box ought to be equal to their unconditional credence that the money is in the opaque box, and so on. From this, it is easy to show that CDT recommends Two-Box for any probabilistically coherent pair of credences $Cr(P1)$ and $Cr(P2)$:

$$CEU(\text{One-Box}) = Cr(P1)(m) + Cr(P2)(\$0)$$

$$CEU(\text{Two-Box}) = Cr(P1)(m + k) + Cr(P2)(k)$$

$$\therefore CEU(\text{Two-Box}) = CEU(\text{One-Box}) + k(Cr(P1) + Cr(P2))$$

So CDT cannot conflict with Dominance in *Newcomb's Problem*.

This is where the story typically ends. We are told that no matter what unconditional credences an agent has, so long as they are probabilistically coherent, CDT will tell them they should choose Two-Box. But we are never told what unconditional credences an agent *should* have. So, I ask you, what unconditional credences should the agent have? The only non-arbitrary unconditional credence would be 0.5, but that does not seem warranted. It is not the case that the agent has equal evidence for and against the money being in the opaque box. The only way I could see getting a non-arbitrary unconditional credence would be to use option-credences. However, as I argue in section 4, option credences are not the proper *base* for a decision – at least in this kind of context. So it seems plausible to me to say that there are no unconditional credences

¹⁴Different authors have formulated CDT in slightly different ways. I follow the formulation from Gibbard and Harper (1978). Lewis (1981a) formulates his version in terms of "dependency hypotheses" while Joyce (1999) favors a formulation of CDT that uses imaging.

that the agent should have. In other words, *Newcomb's Problem* is a case with unspecific evidence. This implies that forming a precise unconditional credence in the contents of the opaque box is not warranted. This is so in spite of the fact that precise conditional credences do seem warranted. Suppose the agent saw a detailed track record of the predictor's success in 10,000 cases. It seems very plausible to think that one is warranted in forming precise conditional credences on the basis of this track record.

If this is right, then the traditional story about *Newcomb's Problem* is only half-correct. It is true that CDT does not conflict with Dominance, but it is not true that CDT explicitly recommends Two-Box. This is because if an agent does not have precise unconditional credences, then CDT does not recommend anything. One cannot calculate CEU unless one has precise credences, and if one cannot calculate CEU, then one cannot extract a recommendation from CDT. This implies that in one sense of "maximization of expected utility" – the EDT sense – a rational agent is in a position to choose an option that maximizes expected utility; however, in a different sense of "maximization of expected utility" – the CDT sense – a rational agent is not in a position to choose an option that maximizes expected utility.¹⁵ I think that this conclusion follows naturally from the thesis that evidence can be unspecific.

Given that the historical development of CDT seems to be an attempt to formulate a decision theory that vindicates two-boxing, this may seem like a bad, or at least awkward, result for the causal decision theorist since it seems that I am implying that CDT is silent on *Newcomb's Problem*.¹⁶ But I think that the causal decision theorist could respond by saying that the development of CDT has really been an attempt to formulate a way of calculating expected utility that does not conflict with Dominance. One could say that CDT, strictly speaking, is a conjunction of Dominance and CEU, or one could endorse a generalization of Dominance, what I call Credal Dominance in section 3.2 – these combinations would give an explicit verdict in *Newcomb's Problem*. I also discuss some possible imprecise decision principles the causal decision theorist could incorporate into their overall decision-theoretic framework in section 3.4.

What is more troubling for the causal decision theorist is a case like *The Frustrater*, which many consider to be a counterexample to (orthodox) CDT. However, in the next section I argue that the causal decision theorist can appeal to unspecific evidence in order to evade this alleged counterexample.

3. Unspecific Evidence and The Frustrater

3.1 The Frustrater

Before I discuss Spencer and Wells' case, consider the following simple case:

Envelope and Boxes: You have a choice between an envelope, a red box, and a blue box. The envelope contains \$40. The two boxes together contain \$100, however, you do not know how the \$100 is distributed between the two boxes.

¹⁵Levi (1975) makes essentially this point. He argues against Nozick's (1969) claim that *Newcomb's Problem* demonstrates a conflict between maximization of utility and Dominance. Levi says that while Nozick presumes that maximization of expected utility "required the computation of expected utility for an option where the probabilities used to compute expected utilities are probabilities of states of nature conditional on the choice of the option whose expected utility is being evaluated" one could instead calculate expected utility "relative to the unconditional probabilities of the states of nature" (168–69, emphasis in original). He then claims that *Newcomb's Problem* is underspecified so an agent is not in a position to act according to a principle of maximization of expected utility in this latter sense.

¹⁶Thanks to an anonymous reviewer for bringing up this worry.

What option do you choose? Perhaps you take the safe choice and select the guaranteed \$40, or maybe you take the risk and choose one of the boxes. Both seem reasonable to me. The point of this simple case is not to make an argument about which option is the rational one to choose. Instead, my point is that if one is on board with the possibility of unspecific evidence, then this seems like a case with unspecific evidence. The evidence regarding the contents of the boxes is incomplete and symmetrically ambiguous. So it seems reasonable to me to suggest that it would be irrational for an agent to form precise credences when they face *Envelope and Boxes*. If this is correct, then *Envelope and Boxes* is a case where maximization of expected utility does not apply.¹⁷ And I think it is due to this fact that all three choices strike me as reasonable choices.

Spencer and Wells' case is a variation of *Envelope and Boxes*¹⁸:

The Frustrater: The setup is the same as in *Envelope and Boxes*, except now you learn that the distribution of the \$100 between the red box and blue box is determined as follows. A reliable predictor, called the Frustrater, predicts whether you choose the envelope ("Envelope"), the red box ("Red"), or the blue box ("Blue"). If the Frustrater predicts Blue, they will put \$100 in the red box. If they predict Red, they will put \$100 in the blue box. If they predict Envelope, then they will put \$50 in the red box and \$50 in the blue box. They make their prediction, and they distribute the money accordingly, one week before your choice.

Let PR be the state where the Frustrater predicts Red, PB be the state where the Frustrater predicts Blue, and PE be the state where the Frustrater predicts Envelope. We can represent *The Frustrater* with the following decision matrix:

	PR	PB	PE
Red	\$0	\$100	\$50
Blue	\$100	\$0	\$50
Envelope	\$40	\$40	\$40

Spencer and Wells' argument against CDT based on this case is straightforward:

- P1. In *The Frustrater*, Envelope is the uniquely rational choice.
- P2. In *The Frustrater*, CDT says that Envelope is an irrational choice.
- C. Therefore, CDT is false. [P1, P2]

¹⁷It's tempting to use the *Principle of Indifference* in this case. Two concerns. First, a liberal use of the Principle of Indifference would simply rule out the possibility of unspecific evidence. Given that my aim of this paper was to highlight some implications of accepting that evidence can be unspecific, it seems I must also assume that the Principle of Indifference cannot be applied in every case. Second, even if the Principle of Indifference can be used here, such an application would still not work for *The Frustrater*. Any attempt at establishing the relevant symmetries required for the Principle of Indifference to apply in *The Frustrater* will be affected by one's option-credences. In section 4, I show how using option-credences in this manner leads to difficulties.

¹⁸This case also makes an appearance in Ahmed (2020) and Spencer (2021).

In defense of P1, Spencer and Wells appeal to a “why ain’cha rich?” (WAR) argument. The idea is that if we imagine two agents, the envelope-taker and the box-taker, who face *The Frustrater* many times, we expect that the envelope-taker will win \$40 every time and that the box-taker will win nothing nearly every time. The envelope-taking strategy will likely make an agent wealthier than the box-taking strategy. Given that the assumption for a case like this is that the hypothetical agent only cares about maximizing their monetary wealth, then it seems reasonable to conclude that the envelope-taking strategy is the rational strategy. Hence, Envelope is the rational choice in *The Frustrater*.

The original WAR argument is an argument for one-boxing in *Newcomb’s Problem*. If we imagine that two agents, the one-boxer and the two-boxer, face *Newcomb’s Problem* many times, we expect that the one-boxer will end up richer than the two-boxer. On the assumption that the predictor’s predictions are very accurate, we expect that the one-boxer will win \$1,000,000 nearly every time, while we expect that the two-boxer will win \$1,000 nearly every time. If the goal is to win the most amount of money possible, how could it be that two-boxing is rational? As Lewis (1981b) puts it, the one-boxer asks the two-boxer, “why ain’cha rich?”

But the WAR argument for one-boxing in *Newcomb’s Problem* faces serious challenges. First, imagine a case that is similar to *Newcomb’s Problem* but now both boxes are transparent. Box 1 has 1,000 dollars in it and a reliable predictor puts 1 million dollars in box 2 if, and only if, they predict that the agent will take only box 2. Call this *Transparent Newcomb*. While there is entrenched disagreement regarding the rational choice in *Newcomb’s Problem*, both sides typically agree that taking both boxes is the rational choice in *Transparent Newcomb*.¹⁹ Both CDT and EDT agree with this as well. But imagine that there are one-box fetishists who take only box 2 no matter what. Notice that these one-box fetishists are in a position to run a WAR argument in favor of one-boxing in *Transparent Newcomb*. Assuming that the predictor’s predictions are very accurate, we expect that the one-box fetishists will win \$1,000,000 nearly every time, while we expect the ordinary agents who take both boxes will win only \$1,000 every time. The reasoning given in favor of one-boxing in *Transparent Newcomb* seems to be identical to the reasoning given in favor of one-boxing in *Newcomb’s Problem*. So if the WAR argument in favor of one-boxing in *Newcomb’s Problem* is sound, then the WAR argument in favor of one-boxing in *Transparent Newcomb* is sound. Given that we all reject the rationality of one-boxing in *Transparent Newcomb*, this ought to make us suspicious of the soundness of the WAR argument in the original *Newcomb’s Problem*.²⁰

Second, consider the opportunities that one-boxers and two-boxers had when they made their respective decisions. When the one-boxer makes their decision, they likely have a choice between an option worth \$1,000,000 and an option worth \$1,001,000. So the worst they could possibly do is win \$1,000,000. And they knowingly choose the worst option.²¹ Compare this to the two-boxer. When they make their decision they likely have a choice between an option worth \$0 and an option worth \$1,000. The

¹⁹I say “typically” because there are some who defend one-boxing in *Transparent Newcomb* (e.g., Greene 2018).

²⁰This point was made by Gibbard and Harper (1978) in their discussion of *Newcomb’s Problem*.

²¹The worst option in terms of actual utility, that is. The reflective one-boxer knows that the actual utility of one-boxing is less than the actual utility of two-boxing. Their justification for one-boxing is not based on actual utility, but rather on the basis that One-Box is more *choiceworthy* than Two-Box.

best they could possibly do is win \$1,000. And they knowingly choose the best option. It is hard to insist that a person who won \$1,000,000 by knowingly doing the worst that they possibly could is more rational than the person who wins \$1,000 by knowingly doing the best that they possibly could. Another way to put the point, is that in order for a WAR argument to be sound, the explanation for a disparity in wealth must refer to an agent's irrationality. It seems that in *Newcomb's Problem*, a perfectly good explanation for the disparity in wealth is the unequal opportunities that the two agents face. The one-boxer has an opportunity that the two-boxer never has, namely the opportunity to win \$1,000,000.²²

The WAR argument in support of Envelope in *The Frustrater* seems to avoid the problems with the WAR argument in support of one-boxing in *Newcomb's Problem*. First, there is not a large disparity in opportunity. The envelope-taker typically has a choice between an option worth \$40, an option worth \$50, and an option worth \$50, while the box-taker typically has a choice between an option worth \$40, an option worth \$0, and an option worth \$100.²³ Second, I do not think it is even possible to imagine a transparent version of *The Frustrater*, so long as we are assuming that the agents facing *The Frustrater* are not absolutely insane. If all three boxes were transparent, then any reasonable person will take the box that contains \$100 or one of the boxes that contain \$50, depending on how the money is distributed. There simply cannot be a reliable predictor in a transparent variation of *The Frustrater* since the predictor is forced to indicate their prediction by how they distribute the money between the boxes, but, by doing that, any reasonable agent facing the transparent version of *The Frustrater* will falsify the prediction.²⁴ This is in contrast to *Transparent Newcomb*, where it is quite straightforward to imagine a reliable predictor. If a reasonable person is facing *Transparent Newcomb*, then they will choose both boxes no matter what. Presumably the predictor knows this, so if they are predicting the actions of a reasonable person, they will leave one box empty. Taking both boxes is the rational choice for the agent and it *confirms* the prediction, unlike the transparent version of *The Frustrater* where the rational choice for the agent will always *disconfirm* the prediction.

So, unlike the WAR argument in support of one-boxing, the WAR argument in support of Envelope seems to be in good standing. And if the WAR argument in support of Envelope is in good standing, then it is reasonable to think that P1 of Spencer and Wells' argument is true.

The reasoning in support of P2 is quite straightforward. Given that possible states of *The Frustrater* are causally option-independent, the agent's credences in the relevant subjunctive conditionals should be equal to their unconditional credences in the possible states, just like in *Newcomb's Problem*:

$$\text{Cr}(\text{Envelope} \Rightarrow \text{PR}) = \text{Cr}(\text{Red} \Rightarrow \text{PR}) = \text{Cr}(\text{Blue} \Rightarrow \text{PR}) = \text{Cr}(\text{PR}) = a$$

²²This point was made by Lewis (1981b), but see Wells (2019) for an extended defense of this response to the WAR argument in *Newcomb's Problem*.

²³Although both the envelope-taker and the box-taker have three options worth \$140 in total, one may still be worried about the differences in how the money is distributed. While there is not a large disparity in opportunity, we cannot say that they have the *same* opportunities. However, one can slightly modify *The Frustrater* such that both the box-taker and the envelope-taker have the same opportunities. See Ahmed (2020: 883).

²⁴This is reminiscent of the discussion from Scriven (1965). See also Ismael (2016: ch. 7).

$$\text{Cr}(\text{Envelope} \Rightarrow \text{PB}) = \text{Cr}(\text{Red} \Rightarrow \text{PB}) = \text{Cr}(\text{Blue} \Rightarrow \text{PB}) = \text{Cr}(\text{PB}) = b$$

$$\text{Cr}(\text{Envelope} \Rightarrow \text{PE}) = \text{Cr}(\text{Red} \Rightarrow \text{PE}) = \text{Cr}(\text{Blue} \Rightarrow \text{PE}) = \text{Cr}(\text{PE}) = c$$

Taking the envelope is a guaranteed \$40, so $\text{CEU}(\text{Envelope}) = \40 .

Here are the causal expected utilities of Red and Blue:

$$\text{CEU}(\text{Red}) = a(\$0) + b(\$100) + c(\$50) = b(\$100) + c(\$50)$$

$$\text{CEU}(\text{Blue}) = a(\$100) + b(\$0) + c(\$50) = a(\$100) + c(\$50)$$

Given that $a + b + c = 1$, this means that $\text{CEU}(\text{Red}) + \text{CEU}(\text{Blue}) = \100 . So no matter the values of a , b , and c , either $\text{CEU}(\text{Red}) > \text{CEU}(\text{Envelope})$ or $\text{CEU}(\text{Blue}) > \text{CEU}(\text{Envelope})$. As long as an agent has probabilistically coherent unconditional credences, CDT will never recommend Envelope.

This seems to be good so far, but let us return to the topic of unspecific evidence. If *Newcomb's Problem* and *Envelope and Boxes* are cases with unspecific evidence, then it seems reasonable to think that *The Frustrater* is also a case with unspecific evidence. If *The Frustrater* is a case with unspecific evidence, then if we are to imagine a *rational* agent facing *The Frustrater*, we must imagine that they do not form precise unconditional credences in the possible states. From this, it follows that $\text{CEU}(\text{Red})$ and $\text{CEU}(\text{Blue})$ will be undefined, thus CDT is inapplicable in *The Frustrater*. This implies that P2 of Spencer and Wells' argument is false. CDT does not say that Envelope is irrational because CDT does not say anything. Thus, on the assumption that the evidence in *The Frustrater* is unspecific, Spencer and Wells' argument against CDT is unsound.

3.2 Bad Dominance

It is tempting to try to avoid the conclusion that CDT is silent in *The Frustrater* by appealing to a kind of dominance reasoning. The argument goes like this. Even if it is granted that an agent cannot rationally assign unconditional credences to the possible states in *The Frustrater*, it is still the case that for any probabilistically coherent credences that one *could* assign, CDT has the consequence that it *would* be irrational for someone with those credences to choose Envelope. Therefore, CDT has the consequence that it is irrational to choose Envelope even for an agent who does not have any unconditional credences.

The above reasoning is not sound. To see why, first consider a version of Dominance where instead of using actual utilities we use expected utilities:

An option O *credal dominates* an option Q if, and only if, for any probabilistically coherent credence function $\text{Cr}(\cdot)$, $\text{CEU}(\text{O}) > \text{CEU}(\text{Q})$. *Credal Dominance* says that when O credal dominates Q, then it is irrational to choose Q over O.

Credal Dominance represents a natural extension of Dominance to expected utility. From this, one could combine Credal Dominance with CDT to get an explicit verdict from CDT in *Newcomb's Problem* since Two-Box credal dominates One-Box. Note that this is possible *even if* the agent does not, in fact, have any precise credences. Credal Dominance quantifies over the probabilistically coherent credence functions the agent *could have*, not the credence function (or functions) they *do have*.

Unlike in *Newcomb's Problem*, there is no dominant option in *The Frustrater*, so we cannot use Dominance to justify a choice. There is also no credal-dominant option in *The Frustrater*, so we cannot use Credal Dominance to justify a choice of option. It is not the case that for any probabilistically coherent credence function $\text{CEU}(\text{Red}) > \text{CEU}(\text{Envelope})$ and it is not the case that for any probabilistically coherent credence function $\text{CEU}(\text{Blue}) > \text{CEU}(\text{Envelope})$. Rather, in *The Frustrater*, for any probabilistically coherent credence function *either* $\text{CEU}(\text{Red}) > \text{CEU}(\text{Envelope})$ *or* $\text{CEU}(\text{Blue}) > \text{CEU}(\text{Envelope})$.

One may say that there is still a *sense* in which Envelope is dominated since CDT never says it is rational, but trying to use this sense of dominance in order to finagle an explicit verdict from CDT would be based on a false principle of rational choice. It would be the extension of the following principle from actual utilities to expected utilities:

Bad Dominance: If there is an option O such that for every state S_i there is another option Q such that $U(Q \wedge S_i) > U(O \wedge S_i)$, then it is irrational to choose O .

Note the scope of the quantifier. It is not that there is a Q such that for every S_i , $U(Q \wedge S_i) > U(O \wedge S_i)$ – that's Dominance – but rather it is that for every state S_i there is a Q such that $U(Q \wedge S_i) > U(O \wedge S_i)$. This change in the scope of the quantifier matters. Dominance is a plausible principle of rational choice, whereas Bad Dominance is an implausible principle of rational choice. Consider the following decision matrix:

	S1	S2
A	\$10	\$10
B	\$11	\$0
C	\$0	\$11

According to Bad Dominance, it would be irrational to choose A, but this is wrong. Unless one was very confident in S1 or S2, then A is the rational choice. Thus, Bad Dominance is false. But notice that the structure of the actual utilities in this case parallels the causal expected utilities in *The Frustrater*. Trying to force the conclusion that CDT explicitly rejects Envelope would not be based on dominance reasoning but on bad dominance reasoning. So while using an extension of Dominance to expected utilities in order to get an explicit verdict from CDT in *Newcomb's Problem* seems acceptable, trying to do the same in *The Frustrater* is unacceptable, since it would rely on a false principle of rational choice.

3.3 Trying to Remove the Unspecificity

Another tempting response is to say that the unspecificity of a case like *The Frustrater* is a mere quirk of presentation and not a deep consideration that impacts how we think about theories of decision. The idea is that the evidence in the case is unspecific because the case is under-described. So all we need to do is provide a richer description of the case such that precise unconditional credences *are* demanded. Once we do this, then a case like *The Frustrater* would constitute a clear counterexample to CDT.

The difficulty with this response is that it is unclear how details can be added to a case like *The Frustrater* such that all three of the following conditions are met:

- (1) The relevant precise unconditional credences are epistemically justified.

- (2) The relevant precise unconditional credences are stable.
- (3) Envelope is the uniquely rational choice.

As I mentioned in section 3.1, I think that *The Frustrater* as described fails to meet condition (1). The evidence relevant to the agent's unconditional credences is imprecise, so precise unconditional credences are not epistemically justified.

One way to meet condition (1) might be to incorporate option-credences – one's credence that one will choose an option – into the determination of one's unconditional credence. So, for instance, if I'm certain that I will choose red, then I should be confident that the state of the world is PR. If we were to supplement this with precise information about the Frustrater's prediction record, e.g., that they predict correctly 90% of the time, then we would get a precise unconditional credence in PR. Namely, $Cr(PR) = 0.9$. The problem with this proposal, as I argue in more detail in section 4, is that this way of determining precise unconditional credences in this manner would fail to satisfy condition (2). One's unconditional credences would be unstable.

The most straightforward way of satisfying (1) and (2) would be to simply provide the agent with strong evidence about what state has obtained. However, it is hard to do this without undermining the significance of the frustrater. For instance, suppose a trusted friend says that she got a peek into the blue box and saw that it was empty. This would strengthen the case for a precise unconditional credence in PB; however, it would undermine the claim that Envelope is the rational choice. One could get a tip that the frustrater was lazy this time around and flipped a coin to determine whether they would put \$100 in the blue box or the red box. This would justify precise credences $Cr(PR) = Cr(PB) = 0.5$, but, again, this would make it so that choosing Envelope would be irrational.

My conjecture is that the instability when using option-credences and/or the unspecific nature of evidence goes hand-in-hand with the intuitiveness of Envelope as the rational choice. I cannot *prove* that this will hold for *every possible* case involving frustrating predictors, however, I can leave it as a challenge for those who think a case like *The Frustrater* is a counterexample to CDT. If you can provide a case where conditions (1)–(3) are clearly met, then you will have made a stronger case against CDT.

3.4 Incompleteness and Imprecise Decision Principles

Even if all I have said so far is correct, there does seem to be a weaker – but perhaps still strong – argument to be made against CDT on the basis of *The Frustrater*. Call a theory of rational choice *T* *sound* if, and only if, whenever it is irrational to Φ , *T* does not endorse Φ . Call a theory of rational choice *T* *complete* if whenever it is rational to Φ , *T* endorses Φ . So far I have been addressing an argument that targets the soundness of CDT; however, it seems that one could instead use *The Frustrater* to target the completeness of CDT.

- P1. In *The Frustrater*, Envelope is the uniquely rational choice.
- P2*. In *The Frustrater*, CDT does not endorse Envelope.
- C*. Therefore, CDT is incomplete. [P1, P2*]

I have argued that P2 is false, but my arguments imply that P2* is true. CDT does not endorse Envelope because CDT does not endorse any action. So it seems that Spencer and Wells could still maintain that their case shows that CDT is incomplete.

This point can be granted, but the charge of incompleteness in this context does not strike me as very worrying for the causal decision theorist. It is plausible to think that

any particular theory of rational choice will have a limited domain of applicability.²⁵ So it may be that a theory *does not* apply in a case because it *should not* apply in that case. For the charge of incompleteness against CDT to have any bite, it must be argued that CDT should apply in this case. However, it seems to me that whether CDT should apply in a case depends on whether an agent *should* have precise unconditional credences in a case. If *The Frustrater* is a case with unspecific evidence, then the agent *should not* have unconditional credences.

Perhaps one insists that we ignore the issue of what credences the agent *should* have and just stipulate that an agent does, in fact, have unconditional credences. After making this stipulation we can argue as follows. Given this stipulation, CDT will explicitly endorse the box-taking strategy over the envelope-taking strategy. Thus, the follower of CDT will end up financially impoverished relative to the envelope-taker. From this, we conclude that CDT is unsound.

If we are operating under the assumption that the evidence in *The Frustrater* is unspecific, then this reasoning is unconvincing. If the evidence is unspecific, then it is irrational to have unconditional credences. So even if we accept that Envelope is the rational choice in *The Frustrater*, the stipulation that an agent has precise credences only supports the following conditional: if an agent has irrational credences in *The Frustrater*, then following CDT will lead them to be financially impoverished. But this should not count against CDT. A theory of rational choice cannot be blamed for bad outcomes if an agent has irrational credences. It is a case of garbage in, garbage out.

This may seem like a punt. Or, worse, a turnover on downs. But it is neither. It is a handoff. If we accept that unspecific evidence is possible, then we accept that precise credences are sometimes irrational. If we accept that precise credences are sometimes irrational, then we need theories and/or principles of decision that do not rely on precise credences. Luckily, we *do* have theories and principles of decision that do not rely on precise credences. The seminal work on this goes back to Isaac Levi (1974; 1980), but the work continues.²⁶ So the causal decision theorist who accepts my arguments has a few different options available to them regarding what to say in cases of unspecific evidence.

A conservative approach would be to simply accept Dominance for all decisions, Credal Dominance for agents with imprecise credences, and CEU for agents with precise credences. This combination would recommend Two-Box in *Newcomb's problem* but would remain silent on *The Frustrater*.

A more interesting approach would be to accept an imprecise decision rule that would not conflict with Dominance but would also vindicate Envelope in *The Frustrater*. Interestingly, one rule that could *not* do this is Levi's E-Admissibility.²⁷ Roughly, E-Admissibility says that an option is rationally permissible if it maximizes expected utility relative to at least one probability function in an agent's representor.²⁸

²⁵Note that I am *not* suggesting that there are domains where no theory of rational choice applies; rather I am suggesting that it is possible that, for any particular theory of rational choice, its domain of application is smaller than the total domain of application for theories of rational choice.

²⁶See, e.g., Seidenfeld (1988; 2004), Troffaes (2007), Joyce (2010), Chandler (2014), Williams (2014), Moss (2015), Bradley and Steele (2014, 2016), Bradley (2019), Steele (2021), and Weatherston (ms).

²⁷See, e.g., Levi (1974) and Seidenfeld (2004).

²⁸A *representor* is a (typically convex) set of probability functions that model an agent with an imprecise attitude toward a proposition. See Levi (1974), van Fraassen (1990), Seidenfeld (2004), and Joyce (2010).

But we've already seen that Envelope will never maximize CEU, so accepting E-Admissibility would say that Envelope is not permissible.²⁹

One rule that could endorse Envelope is Γ -Maximin, which is essentially a version of the traditional maximin principle where *actual* utilities are replaced with *expected* utilities. Adopting this approach, Γ -Maximin would say that an agent ought to choose the option that maximizes the minimum value of CEU relative to the probability functions in their representor. Depending on the agent's representor, Γ -Maximin could recommend Envelope since the minimum value of CEU for Envelope is \$40, while the minimum value of CEU for Red or Blue could be as low as \$0, depending on the nature of the agent's representor.^{30,31}

Admittedly both of the approaches have challenges. Accepting only Credal Dominance would invite the aforementioned complaints of incompleteness and Γ -Maximin seems to endorse a (problematic kind of) refusal of new information.³² My point for the present purposes, however, is not to argue for a particular imprecise decision theory. Instead, my point is that even if CDT is silent on *The Frustrater*, it does not mean that the causal decision theorist has nothing to say.

4. Objections

In this section, I respond to two objections to my reasoning in section 3. The first objection says that talk of unspecific evidence is simply irrelevant to decision theory. The second objection says that in *The Frustrater* an agent can form precise unconditional credences by incorporating option-credences – an agent's unconditional credence that they will choose one of the available options – into their deliberation.

The first objection goes something like this:

You are misunderstanding decision theory. You keep talking about the evidence that an agent has, but EDT and CDT are *subjective* theories of rational choice.

²⁹A slight modification to the scope of E-Admissibility could say that Envelope is rational. E-Admissibility says that an option A is permissible if, and only if, there exists a credence function in an agent's representor such that the expected utility of A is greater than (or equal to) every other option B. If instead we have a principle that says that A is permissible if, and only if, for every other option B, there exists a credence function in the agent's representor such that the expected utility of A is greater than (or equal to) B. Thanks to an anonymous reviewer for proposing this principle. Note that it would say that *all* of the options in *The Frustrater* are permissible. So it would still seem to be vulnerable to the argument from Spencer and Wells, since they argue that Envelope is *uniquely* rational.

³⁰Interestingly, Γ -Maximin seems to endorse the "correct" verdict in other alleged counterexamples to CDT that are plausibly cases that call for imprecise credences. For example, for the same reasons that it seems plausible that *The Frustrater* calls for imprecise credences, Egan's (2007) *The Psychopath Button* plausibly calls for imprecise credences. While it will depend on the agent's utilities and their representor, Γ -Maximin could endorse not pressing the button in *The Psychopath Button*.

³¹Another interesting option would be a version of Savage's (1951) Minimax-Regret where actual utilities are replaced by expected utilities. This principle would say that the agent ought to choose the option that minimizes the maximum expected "regret" of an option, where the expected regret of an option relative to a particular credence function is the difference between the value of CEU for that option and the maximum value of CEU among all other options. This principle could also recommend Envelope in *The Frustrater* since the maximum expected regret value for Envelope is \$60, while the maximum expected regret for either Red or Blue is \$100. And, depending on the agent's utilities and representor, this principle could endorse not pressing the button in Egan's (2007) *Psychopath Button*, since the maximum expected regret of pressing the button could be higher than the maximum expected regret of not pressing the button.

³²See Seidenfeld (2004), Bradley and Steele (2016), and Steele (2021) for discussions on this point.

So the only thing we, decision theorists, need to care about is what credences agents *actually* have, not what credences agents *should* have. We are not concerned with whether their credences are evidentially justified – that is something for epistemologists to worry about. A *decision* is simply a quadruple $\langle O, S, Cr, U \rangle$ where O is the set of options available to the agent, S is the possible states, Cr is the agent's credence function, and U is the agent's utility function. Unspecific evidence is irrelevant to decision theory, because evidence *tout court* is irrelevant to decision theory.

This is a solipsistic conception of decision theory – decisions are just in the head. And I have no qualms with understanding *decisions* in this manner. What I cannot make sense of is how we are supposed to judge whether things go well or poorly for an individual if the only information is the agent's opinions about the world and *nothing else* – we cannot judge whether their opinions are true/false, good/bad, rational/irrational. This conflicts with how we actually analyze different decision theories and principles.

Consider *Newcomb's Problem*. Suppose an agent can be represented with the proper quadruple. When we imagine that they choose One-Box, we imagine them looking inside and seeing \$1,000,000. This is what we imagine because we imagine that that is the most likely outcome. But facts about the agent's *opinions* don't demand that we imagine this, it is facts about the *case*. We assume that the predictor is, in fact, very reliable. If we are not allowed to assume this and, furthermore, we are not allowed to assume that a high value for the conditional credence $Cr(P1|One-Box)$ is rational, then there is no reason to assume that there is \$1,000,000 in the opaque box when the agent chooses One-Box. Mere subjective facts on their own may tell me what the agent will *expect* to happen, but they do not tell me what will *actually* happen. Facts *about the case* tell me what will happen. It is the fact that the predictor *is* reliable that tells me to imagine that the one-boxer will be richer than the two-boxer in *Newcomb's Problem*, not the fact that a person *thinks* that the predictor is reliable.

I emphasize that I am not rejecting the view that says the decision an agent faces is determined by their actual credences, rather I am saying that when we are analyzing what theory of rational choice to accept by imagining cases, we ought to imagine that any hypothetical agent in an imagined case has rational credences. So my claim is not that an agent is rationally required to act against their credences if their credences are not supported by the evidence. Rather, my claim is that if an agent acts on credences that are not supported by the evidence, then we cannot use the fact that things go poorly for them as reason for rejecting the theory of rational choice they follow. It only makes sense to use the fact that things go poorly for an agent against a theory of rational choice – as one tries to do in a WAR argument – under the assumption that the agent's credences are supported by the evidence.

The upshot is this. It may be the case that EDT and CDT are subjective theories of rational choice, but this does not mean that evidence is irrelevant to decision theory. When we are trying to assess the status of a particular theory of rational choice by imagining cases, we have to imagine that the doxastic attitudes of the hypothetical agent fit the evidence of the case. But this brings the considerations of this paper front and center. In a case like *The Frustrater*, we should only imagine an agent whose credences fit the evidence. If precise unconditional credences do not fit the evidence, then what an agent with precise credences would or would not do should not constrain our theorizing about the correct theory of rational choice because those precise credences would be irrational.

One might respond to this by receding even further and suggesting that decision theory is merely about articulating principles of means-end structural rationality. Just like one could consider probabilism to be a structural constraint on doxastic attitudes, theories of decision are mere structural constraints on practical rationality. This is indeed one conception of decision theory, but it cannot work as a *response* to my arguments. This is because, on this construal of decision theory, a case like *The Frustrater* cannot possibly be a counterexample to CDT. If we suppose an agent has precise unconditional credences and perpetually takes boxes, there is nothing structurally incoherent with this behavior. The argument from Spencer and Wells is not that there is structural incoherence in the agent that perpetually takes boxes, but rather that taking one of the boxes in *The Frustrater* is irrational.

The next objection goes like this. Even if we accept that there can be cases of unspecific evidence (e.g., in *Strange Items* and *Envelope & Boxes*), the hypothetical agent in *Newcomb's Problem* and *The Frustrater* is in a position to justifiably form precise credences. The agent can use the combination of their confidence in the agent's prediction and their confidence in what option they will choose in order to form precise unconditional credences in the relevant states. For example, if the agent is very confident that they will choose Two-Box in *Newcomb's Problem*, then they ought to be very confident that the opaque box is empty, and *vice versa* if they are very confident that they will choose One-Box. If they have precise values for the relevant conditional credences and option-credences, then the values for the relevant unconditional credences simply follow from the probability calculus.

I think that there are serious difficulties with the use of option-credence in the suggested manner, at least for cases like *The Frustrater*. Using option-credences in the manner suggested for *The Frustrater* will lead to an instability. And though it is possible to stipulate away this instability, doing so would demand using option-credences in a way that is incoherent and/or practically irrational.

Regarding the point about instability, consider a well-known case discussed by Gibbard and Harper (1978).

Death in Damascus: Death works from an appointment book that states time and place; a person dies if, and only if, the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions of your actions. An appointment for tomorrow has been inscribed for you; you know that it is either for Aleppo or for Damascus. You must decide now whether to stay in Damascus overnight, or ride to Aleppo to arrive tomorrow morning.

The suggestion I am considering is to use my confidence in death's predictive powers and my confidence in where I will go in order to determine my confidence in where death is. So if I am very confident that I will stay in Damascus, then this means I should be very confident that death is in Damascus. And if I am very confident death is in Damascus, then I should ride to Aleppo. An instability looms.

The same kind of instability arises in *The Frustrater*. Perhaps one thinks that this kind of instability is problematic in both *Death in Damascus* and *The Frustrater*, so sees this as a problem for CDT.³³ But if this were all *The Frustrater* showed, then it would be old news – it would not constitute a new challenge to CDT. But it *does*

³³For instance, Ritcher (1984) uses instability cases to argue against CDT.

constitute a new challenge. Namely, that it is intuitively compelling that Envelope is a rationally permissible choice in *The Frustrater*, yet CDT can never endorse Envelope. Furthermore, this instability only arises when one uses option-credences in order to determine unconditional credences in the calculation of CEU. So even if it is a problematic kind of instability, there are ways of rectifying it without rejecting CDT.³⁴

But suppose the suggestion is to stipulate away the instability. So, in *The Frustrater*, it is stipulated that the agent has a high credence that they will take the red box and, on that basis, has a high credence that \$100 is in the blue box. It is also stipulated that they do not change their credences. From this, it would be argued that CDT would say that the agent ought to take the blue box, since taking the blue box has the highest causal expected utility. But then, it would be argued, we should expect the box to be empty given the reliability of the Frustrater's predictions. Thus, CDT is to blame for the agent's financial woes.

I cannot make sense of this. Borrowing a term from Spencer (2021), say that an agent *embodies* a theory of rational choice T if, and only if, they always act according to T and they know that they always act according to T. Another way of putting the argument from Spencer and Wells, then, is that an agent who embodies CDT will be financially impoverished relative to the envelope-taker in *The Frustrater*. With this in mind, I think it is easier to see the incoherence in what is being suggested.

We are told that the agent chooses Blue because Blue has the highest causal expected utility. We are also told that Blue has the highest causal expected utility because the agent has a high credence that \$100 is in the blue box. Their high credence that the \$100 is in the blue box is based on their high credence that they will choose Red. But, if they embody CDT, then they *know* that they will choose Blue given that $CEU(\text{Blue}) > CEU(\text{Red})$. So we are told to imagine an agent that knows that they will choose Blue because they have a high credence that they will choose Red. This seems incoherent to me.

Whether option-credences should be incorporated into one's deliberative process in general is itself a controversial topic. Isaac Levi famously coined the phrase "deliberation crowds out prediction." It may look as though I am appealing to Levi's thesis – call it DCOP – in order to defend this last point. Perhaps I am, but different authors seem to understand DCOP in different ways. If I am appealing to DCOP here, then I think I am only appealing to a very weak version of DCOP.

In Hájek's (2016) critique of DCOP, he defined it as follows: *while deliberating about what you'll do, you cannot rationally have credences for what you'll do*. This definition is unfortunate because it is ambiguous. What exactly is being claimed in this definition depends on how one understands "cannot" and "rationally." Many of Hájek's criticisms of DCOP seem to be predicated on understanding "you cannot rationally have credences for what you'll do" as "option-credences are epistemically irrational." In other words, you cannot have option-credences on pain of epistemic irrationality. This is a possible formulation of DCOP, but it is not the only possible formulation of DCOP. I expect that for many, DCOP is not supposed to be an epistemic constraint on what credences would be epistemically rational or irrational, but rather it is supposed to capture a conceptual point about what it is to be a deliberating agent. The idea is that when an agent takes the deliberative stance, they bracket option-credences. This does not mean that one

³⁴E.g., one could accept a deliberative causal decision theory in order to address these cases (see Joyce 2012; Skyrms 1982).

cannot have option-credences, or that it would be irrational to have them, but rather that they should be bracketed when one is deliberating about what one should do.³⁵

While there is a dispute regarding the nature and status of DCOP, I do not think that my objection to the use of option-credences in *The Frustrater* is based on Hájek's definition of DCOP nor is it predicated on the conceptual understanding of DCOP.³⁶ Instead, my objection is that the particular use of option-credences required to establish precise unconditional credences in *The Frustrater* is practically irrational. Joyce (2002) points out that there are three worries one could have about incorporating option-credences into practical deliberation.³⁷

Worry-1: Allowing option-credences might make it permissible for agents to use the fact that they are likely (or unlikely) to perform an act as a reason for performing it.

Worry-2: Allowing option-credences might destroy the distinction between options and states that is central to most decision theories.

Worry-3: Allowing option-credences multiplies entities needlessly by introducing quantities that play no role in decision making.

I think that both Worry-2 and Worry-3 are related to the conceptual understanding of DCOP that I mentioned, however, my worry is Worry-1. What I cannot make sense of is an agent who uses the fact that they have a high credence that they will choose Red as *justification* for choosing Blue. Here is what Joyce has to say about Worry-1.

As to Worry-1, I entirely agree that it is absurd for an agent's views about the advisability of performing any act to depend on how likely she takes that act to be. Reasoning of the form "I am likely (unlikely) to A, so I should A" is always fallacious.³⁸

I think Joyce is right on this point. Such reasoning is fallacious. And notice that accepting that such reasoning is fallacious does not mean that one must maintain that option-credences are epistemically irrational, nor does it mean that one must maintain that

³⁵Hájek calls DCOP the DARC thesis: deliberation annihilates reflexive credences. But to bracket something is not to annihilate it, so it seems the thesis Hájek attacks is not the thesis I am suggesting. Indeed, Hájek may even be somewhat sympathetic to this understanding of DCOP since he seems to be somewhat sympathetic to a suggestion he attributes to Huw Price that option-credences should be "muted."

³⁶Indeed, I think the conceptual point, even if accepted, ought to be only accepted in a restricted form. For example, I think it is both coherent and permissible to use option-credences in some decisions. One notable example is from White (2021). It's Mother's Day and I know the right thing to do is to call my mom and be nice. However, I know how our conversations tend to go, and I know if I call her we will end up getting into an argument and I will likely say something mean. So I decide not to call her. In this case, I'm using credences about what I will do in order to justify my choice of action. While this case brings up many interesting philosophical issues (see the discussion in White 2021), I am not contesting the use of option-credences in this manner. I am only contesting a particular use of option-credences, namely, using one's credence that one will do A in order to justify doing A (or not doing A). White's case does not have this structure. I am using my credence *that I will say something mean* to justify *not calling my mom*. I am not using my credence *that I will call my mom* to justify *not calling my mom*.

³⁷Joyce (2002: 79).

³⁸Joyce (2002: 79).

using option-credences is *always* practically irrational. Rather it is to say that using option-credences in a particular way is practically irrational. But to get precise unconditional credences from option-credences in *The Frustrater* would require using option-credences in precisely this way.

There are cases where reasoning from “I am likely to A” to “I should A” seems innocuous. Consider a friendly (and very reliable) predictor who put \$100 in a black box and \$0 in a silver box if, and only if, they predicted you would choose the black box, and \$0 in a black box and \$100 in a silver box if, and only if, they predicted you would choose the silver box.³⁹ Suppose you have a slight compulsion toward shiny things, so have a high credence that you will take the silver box. Consequently, you have a high credence that the friendly predictor puts \$100 in the silver box. Here it seems fine, or at least benign, to reason from “I am likely to choose the silver box” to “I should choose the silver box.”

But I maintain that it is not. Suppose instead that the friendly predictor will only put \$10 in the silver box if they predicted you would choose the silver box. Furthermore, while your compulsion makes you think it likely that you will choose the silver box, whenever you do choose shiny things it makes you slightly sad. So, while you have a high credence that you will choose the silver box, you know you could prefer to choose the silver box (setting aside the possible contents of the boxes). Admitting that “I am likely to choose the silver box” can be a proper justification for one’s choice results in saying that choosing the silver box could be the rational choice in this situation. The high credence that you will choose the silver box leads to a high credence that there is \$10 in the silver box, so, for a high enough credence, $CEU(\text{silver box}) > CEU(\text{black box})$. But you know that the black box’s potential value is greater, you know that you would prefer to choose the black box, and you know that your choice does not causally influence the contents of either box. This seems to me to illustrate the problem with allowing option-credences to play a *justificatory* role in determining rational choice, regardless of whether the hypothetical predictor is friendly or frustrating.

One possible response to this is to say that, in a case like *The Frustrater*, it is not the fact that I have a high credence *that I will choose Red* that is justifying my choice of Blue, but rather it is the fact that I have a high credence *that the money is in Blue*. So while Worry-1 is a legitimate worry, it is not something one should be worried about in *The Frustrater*.⁴⁰ This is essentially Joyce’s view, so while I have invoked some of what he has said in defending my view, he ultimately would disagree with me on this point. Joyce acknowledges that one can present CDT without the use of option-credences, as I have, however, he prefers an account of CDT that does make use of option-credences (see Joyce 1999; 2002; 2012). Joyce suggests that his formulation of CDT cannot fall prey to Worry-1, because the values of option-credences are not inputted *directly* into the calculation of CEU (Joyce 2002: 80). So Joyce would say that one can and should use option-credences in *The Frustrater*. Due to the unratifiability of Red or Blue in *The Frustrater*, the agent that embodies CDT ought to be 50/50 regarding what option they will choose. So $Cr(\text{Red}) = Cr(\text{Blue}) = 0.5$. Thus, the agent that embodies CDT will be indifferent between Red and Blue. The idea is that, at deliberative equilibrium, it is the agent’s credences in where the money is, not the credences in what they will do, that justifies their indifference.

³⁹Thanks to an anonymous reviewer for this case.

⁴⁰Thanks to an anonymous reviewer for raising this response.

I still find this suggestion to be problematic. Consider that if the agent were to lose their option-credences, then they would also lose their unconditional credences in the states. This seems to me to indicate that the values of the unconditional credences depend on the values of the option-credences. If this is correct, then the values of CEU(Red) and CEU(Blue) ultimately depend on the values of the option-credences. And given that the justification for choosing Red or Blue is based on the values of CEU(Red) and CEU(Blue) *which are themselves based on the agent's option-credences*, then it is still a case of Worry-1. If the values that are directly inputted into the calculation of CEU are themselves directly dependent on the values of one's option credences, how is this not still reasoning from the form "I am likely (unlikely) to A, so I should A" but with one extra step added?

I admit that I do not take the above considerations to be decisive, but I hope I have made the case for my way of conceptualizing how a rational, deliberative agent approaches decisions. What I think divides philosophers on this issue is, in part, what stage of the deliberative process one focuses on. I think the kind of perspective that I am advocating for looks bizarre to my opponent, in part, because they are focusing on the *endpoint* of deliberation. Even in my aforementioned case of an asymmetric friendly predictor, it seems as though my position is that picking the silver box is irrational even if you are overwhelmingly confident that the silver box has \$10 and the black box has \$0. This is, of course, counterintuitive. However, I think that the opposing view is problematic, in part, because I am focusing on the *starting point* of deliberation. It seems strange to let one's confidence that one will do A affect the reasonableness of doing A, instead of assessing the reasonableness of A independently.⁴¹ So, in the asymmetric friendly predictor, even if you *start* with the thought you will probably take the silver box, if you assess the options independently of their option-credences, you will conclude that taking the black box is more reasonable. When you make this conclusion, this will presumably affect your option-credences, shifting from a high credence that you will take the silver box to a high credence that you will take the black box. At this point both positions would be in agreement that taking the black box is the rational choice.

Summing up. In section 3, I suggested that if an agent's credences are irrational, then a theory of rational choice cannot be blamed for choices based on these irrational credences. It is a case of garbage in, garbage out. The suggestion I have been considering is that unconditional credences in states in *The Frustrater* could be formed in a rational manner by using option-credences. However, using one's option-credences in the manner suggested leads to instability in *The Frustrater*, and stipulating the instability away would force an appeal to the fallacious reasoning "I am likely (unlikely) to A, so I should A." In the same way that a theory of rational choice cannot be blamed for choices based on irrational credences, a theory of rational choice cannot be blamed for choices based on fallacious reasoning. So it is another case of garbage in, garbage out.

5. Conclusion

Trying to understand the nature of evidence and its relation to rational doxastic attitudes raises a host of difficult epistemological issues. It would be nice if decision theory could ignore these issues. Analyzing theories of rational choice would be a lot easier if

⁴¹Note that my point here is restricted to cases of the form of Worry-1, not cases involving the use of option-credences in general. See footnote 36.

we did not have to take a stand on the fundamental nature of evidence. Unfortunately, I do not think that this is the case. How to properly understand decisions may depend on how to properly understand evidence. So whether an *alleged* counterexample is a *bona-fide* counterexample may depend on how evidence constrains doxastic attitudes. In this paper, I have tried to show that the dispute over the status of unspecific evidence intersects with the dispute over the status of CDT. At the very least, I hope that both opponents and proponents of CDT pay heed to this intersection moving forward.⁴²

References

- Ahmed A. (2014). 'Dicing with Death.' *Analysis* 74(4), 587–92.
- Ahmed A. (2020). 'Equal Opportunities in Newcomb's Problem and Elsewhere.' *Mind* 129(515), 867–86.
- Bradley S. (2019). 'A Counterexample to Three Imprecise Decision Theories.' *Theoria* 85, 18–30.
- Bradley S. and Steele K. (2014). 'Should Subjective Probabilities be Sharp?' *Episteme* 11(3), 277–89.
- Bradley S. and Steele K. (2016). 'Can Free Evidence be Bad? Value of Information for the Imprecise Probabilist.' *Philosophy of Science* 83(1), 1–28.
- Briggs R. (2010). 'Putting a Value on Beauty.' *Oxford Studies in Epistemology* 3, 3–34.
- Carr J.R. (2020). 'Imprecise Evidence without Imprecise Credences.' *Philosophical Studies* 177, 2735–58.
- Chandler J. (2014). 'Subjective Probabilities Need Not be Sharp.' *Erkenntnis* 79, 1273–86.
- Egan A. (2007). 'Some Counterexamples to Causal Decision Theory.' *The Philosophical Review*, 116, 93–114.
- Elga A. (2010). 'Subjective Probabilities Should be Sharp.' *Philosophers' Imprint* 10(5), 1–11.
- Elga A. (2012). Errata for subjective probabilities should be sharp. Unpublished.
- Gibbard A. and Harper W.L. (1978). 'Counterfactuals and Two Kinds of Expected Utility'. In W.L. Harper, R. Stalnaker and G. Pearce (eds), *Ifs*, pp. 153–90. Dordrecht: Springer.
- Greene P. (2018). 'Success-First Decision Theories.' In A. Ahmed (ed), *Newcomb's Problem*, pp. 115–37. Cambridge: Cambridge University Press.
- Hájek A. (2016). 'Deliberation Welcomes Prediction.' *Episteme* 13(4), 507–28.
- Ismail J. (2016). *How Physics Makes Us Free*. New York: Oxford University Press.
- Jeffrey R. (1983). *The Logic of Decision*. 2nd edition. Chicago: The University of Chicago Press.
- Joyce J.M. (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
- Joyce J.M. (2002). 'Levi on Causal Decision Theory and the Possibility of Predicting One's Own Actions.' *Philosophical Studies* 110(1), 69–102.
- Joyce J.M. (2005). 'How Probabilities Reflect Evidence.' *Philosophical Perspectives* 19, 153–78.
- Joyce J.M. (2010). 'A Defence of Imprecise Credences in Inference and Decision Making.' *Philosophical Perspectives* 24, 281–323.
- Joyce J.M. (2012). 'Regret and Instability in Causal Decision Theory.' *Synthese* 187(1), 123–45.
- Levi I. (1974). 'On Indeterminate Probabilities.' *Journal of Philosophy* 71(13), 391–418.
- Levi I. (1975). 'Newcomb's Many Problems.' *Theory and Decision* 6(2), 161–75.
- Levi I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability and Chance*. Cambridge, MA: MIT Press.
- Lewis D.K. (1981a). 'Causal Decision Theory.' *Australasian Journal of Philosophy* 59(1), 5–30.
- Lewis D.K. (1981b). 'Why Ain'cha Rich?' *Noûs* 15(3), 377–80.
- Moss S. (2015). 'Credal Dilemmas.' *Noûs*, 49(4), 665–83.
- Nozick R. (1969). 'Newcomb's Problem and Two Principles of Choice.' In N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*, pp. 114–46. Dordrecht: Springer.
- Oesterheld C. and Conitzer V. (2021). 'Extracting Money from Causal Decision Theorists.' *The Philosophical Quarterly* 71(4), 701–16.
- Richter R. (1984). 'Rationality Revisited.' *Australasian Journal of Philosophy* 62(4), 392–403.

⁴²My thanks to Zak Woodman, Caspar Oesterheld, Andy Egan, Terry Horgan, Jason Turner, Juan Comesaña, Tim Kearn, Luke Goleman, Jack Spencer, and Sara Aronowitz. Thanks also to audiences at the Minnesota Philosophical Society Conference (2019) and the Western Michigan University Graduate Philosophy Conference (2019).

- Savage L.J.** (1951). 'The Theory of Statistical Decision.' *Journal of the American Statistical Association* **46** (253), 55–67.
- Scriven M.** (1965). 'An Essential Unpredictability in Human Behavior.' In B.B. Wolman and E. Nagel (eds), *Scientific Psychology: Principles and Approaches*, pp. 411–425. New York: Basic Books.
- Seidenfeld T.** (1988). 'Decision Theory without "Independence" or without "Ordering": What is the Difference?' *Economy and Philosophy* **4**, 267–315.
- Seidenfeld T.** (2004). 'A Contrast between Two Decision Rules for Use with (Convex) Sets of Probabilities: γ -Maximin versus E-Admissibility.' *Synthese* **140**, 69–88.
- Skyrms B.** (1982). 'Causal Decision Theory.' *Journal of Philosophy* **79**(11), 695–711.
- Spencer J.** (2021). 'An Argument against Causal Decision Theory.' *Analysis* **81**(1), 52–61.
- Spencer J. and Wells I.** (2019). 'Why Take Both Boxes?' *Philosophy and Phenomenological Research* **99**(1), 27–48.
- Steele K.** (2021). 'How to be Imprecise and Yet Immune to Sure Loss.' *Synthese* **199**, 427–44.
- Sturgeon S.** (2020). *The Rational Mind*. Oxford: Oxford University Press.
- Troffaes M.C.M.** (2007). 'Decision Making under Uncertainty Using Imprecise Probabilities.' *International Journal of Approximate Reasoning* **45**, 17–29.
- van Fraassen B.** (1990). 'Figures in a Probability Landscape.' In M. Dunn and A. Gupta (eds), *Truth or Consequences*, pp. 345–56. Dordrecht: Springer.
- Weatherston B.** (ms). Decision Making with Imprecise Probabilities. Unpublished manuscript.
- Wells I.** (2019). 'Equal Opportunity and Newcomb's Problem.' *Mind* **128**(510), 429–57.
- White S.J.** (2021). 'Self-Prediction in Practical Reasoning: Its Role and Limits.' *Noûs* **55**(4), 825–41.
- Williams J.R.G.** (2014). 'Decision Making under Indeterminacy.' *Philosophers' Imprint* **14**(4), 1–34.

Rhys Borchert is a PhD candidate in philosophy at the University of Arizona. His research interests include decision theory, epistemology, and normativity.