Actual Value Decision Theory

1 Two Puzzles: Newcomb & Vacation Boxes

Let's start by looking at two puzzles.

> **Newcomb Problem.**[1] You have two boxes before you: an opaque box, which either contains a million dollars or nothing, and a transparent box, which contains a thousand dollars. You have the option to, either, take only the opaque box (*One-Box*) or to take both the opaque and the transparent box (*Two-Box*). Whether the opaque box contains a million dollars or no dollars has been determined by a super-reliable predictor. If the predictor predicted that you'd One-Box, she put a million dollars in the opaque box; if the she predicted that you'd Two-Box, she put nothing in the opaque box.

	Predicts: "One-Box"	Predicts: "Two-Box"
One-Box	$\$M$	\$0
Two-Box	$\$M + K$	$\$K$

Assume that you take the predictor to be so reliable that your credence that she predicted correctly is close to one. And, for simplicity, assume that you value money linearly. Which should you prefer to do: to *One-Box* or to *Two-Box*?

The second puzzle is an example of what Hare [2010] calls "opaque sweetening."

[1] The puzzle was first discussed in print by Nozick [1969], who credits its construction to the physicist William Newcomb.

> **Vacation Boxes.** There are two opaque boxes: a Larger box (L) and a Regular box (R). A fair coin has been tossed. If the coin landed heads, then a voucher for an all-expenses-paid Alpine ski vacation (A) was placed in the Larger box and a voucher for an all-expenses-paid beach vacation (B) was placed in the Regular box; if the coin landed tails, then B was placed in the Larger box and A was placed in the Regular box. In either case, you don't know which prize is in which box.
>
> $$\textbf{Larger box} = \begin{cases} A & \text{if Heads} \\ B & \text{if Tails.} \end{cases} \qquad \textbf{Regular box} = \begin{cases} B & \text{if Heads} \\ A & \text{if Tails.} \end{cases}$$
>
> Now imagine that \$1 is added to the Larger box. If you choose the Larger box, you will win whichever prize it contains plus a \$1. Nothing is added to the Regular box. You are asked to choose one of the two boxes, taking home whichever prize is in the box you choose.

Suppose your attitude toward the two vacations, A and B, is like this: you don't strictly prefer one to the other, nor are you indifferent between the two. Following Chang [2002], let's say that the two are *on a par.*[2] The difference between parity, on the one hand, and indifference, on the other, is that the former is insensitive to mild sweetening while the latter is not. Let us suppose, for example, that you don't prefer the alpine ski vacation plus a dollar (A^+) to the beach vacation, nor do you prefer the beach vacation plus a dollar (B^+) to the alpine ski vacation. (If you were *indifferent* between the alpine ski vacation and the beach vacation, however, then you would prefer the alpine ski vacation plus a dollar to the beach vacation, and you would prefer the beach vacation plus a dollar to the alpine ski vacation).[3]

[2] For Chang, *parity* is a fourth *sui generis* value relation that hold between two comparable goods. The other philosophers who argue that the Completeness Axiom should be relaxed, not because there is a fourth value relation, but rather because, e.g., preferences can be vague or indeterminate ([Broome, 1997], [Gert, 2004]). I don't intend to take sides on this issue. When I say that two things are "on a par," one should feel free to substitute in whichever analysis of the phenomenon one likes.

[3] The "sweetener" needn't be a dollar. The same issue would arise if we sweetened one of the options with 50¢, or an ice-cream cone, or 1¢, or a lottery ticket with a one-in-a-millionth chance at netting 1¢, etc. So long as a good contributes *some* positive value to outcome A and B, it's a potential sweetener.

	Heads	Tails
Take Larger box	A^+	B^+
Take Regular box	B	A

Does rationality require you to take the Larger box, or is it rationally permissible to take either?

I think that you should *Two-Box* in the Newcomb Problem and that it is rationally permissible to take either box (*Either*) in Vacation Boxes. My plan, however, is not to argue in favor of these two positions; instead, I want to suggest that there's an affinity between them. I think that the best argument for *Two-Boxing* in the Newcomb Problem is, also, the best argument for *Either* in Vacation Boxes.

I will use this observation to articulate a conception of rational decision-making — which I call "the Actual Value Conception" — and argue that it underlies causal decision theory. I will then sketch a way of generalizing the view to cases, like Vacation Boxes, in which you regard the potential outcomes of your options as on a par. The decision theory developed at the end of the paper, unlike its competitors, is motivated by the same considerations that underpin causal decision theory.[4]

2 Reflection, Deference, and Dominance

Here's a reason for thinking that you should *Two-Box* in the Newcomb Problem. You know that after making your decision you'll learn whether there's a million dollars in the opaque box or not. And you know that if you will learn that there *is* a million dollars in the opaque box, you will want yourself to have taken both boxes. You also know that if you will learn that there *is not* a million dollars in the opaque box, you will want yourself to have taken both boxes. Either way, then, you know

[4] Its competitors are those views that evaluate options primarily in terms of their corresponding *prospects*: the probability-distribution over its potential outcomes. Here are some examples of views that fall into this class: I.J. Good's Quantizationism [Good, 1952]; Caspar Hare's Prospectism [Hare, 2010]; Isaac Levi's V-admissibility [Levi, 1986, 2008]; Amartya Sen's Intersection Maximization [Sen, 2004]. There are also a number of decision theories designed to handle similar cases that arise not because of incomplete preferences but because of imprecise (or unsharp) credences: for example, Susanna Rinard's Moderate [Rinard, 2015]; Weatherson's Caprice [Weatherson, 2008]; and [Joyce, 2010].

that you will want yourself to have *Two-Boxed* rather than *One-Boxed*. And so, if you're rational, you should now prefer *Two-Boxing* to *One-Boxing*.

Call this **The Argument from Reflection.** The last step in the argument implicitly appealed to a principle of rationality linking your current preferences to what you know, if anything, about the preferences you will have in the future.

> [Preference Reflection]
>
> If you are now in a position to know that you will prefer having φed to having ψed, then you should now prefer φ to ψ.

A similar argument, relying on a similar principle, can be made in support of *Either* in Vacation Boxes. You know that after you make your decision you'll learn which box contained which vacation prize. You know that if what you will learn is that the Larger box contains A^+ and the Regular box contains B, you won't prefer having taken the Larger box over the Regular one. And you know that if what you will learn is that the Larger box contains B^+ and the Regular box contains A, you won't prefer having taken the Larger one. So, either way, you know that you will *not* prefer having taken the Larger box over having taken the Regular box. And so you are not rationally required to prefer taking the Larger box over taking the Regular box.

> [No Preference Reflection]
>
> If you are now in a position to know that you will not prefer having φed to having ψed, then you are not rationally required to now prefer φ to ψ.

The idea behind both of these principles is this: if you're in a position to know what your future preference-like attitudes will be, you should adopt those attitudes now. If you find this idea compelling, you should support *Two-Boxing* in the Newcomb Problem and *Either* in Vacation Boxes; in both of these cases, you are in a position to know something about what your future preference-like attitudes will be, and so you should adopt those attitudes now.

Should you find this idea compelling, though? First, it's not obvious why rationality would require you to now adopt the attitudes you know you will adopt in the future. If the two Reflection Principles are true, what *explains* why they are true?

But, second, even if there's something right about these two principles, neither *are* true in their current formulations.

Counterexamples abound.[5] Suppose you know that you will prefer φ to ψ, in part, because you know that you will soon forget some crucial bit of relevant information. You know that you are now in a better epistemic situation with respect to the relevant features of φing vs ψing than you will be in the near-future. Are you really rationally required to defer to your less-informed future-self? Or suppose that you know you will come to prefer φ to ψ via some a- or ir- rational process — you will take a pill that causes you to adopt such a preference, or you will be conked on the head, or you will adopt the preference on a whim, or what have you. Does rationality require you to adopt preferences you currently consider irrational simply because you know they will be yours? Or suppose you know that your preference for φ over ψ will be the result of a fundamental change to your core values. Why adopt future-you's opinions about the means when you know don't share the same ends?

These problems can be avoided, of course, by amending our Reflection principles with some caveats. We could, for example, strengthen their antecedents by adding: "If you know that you won't be less-informed than you are now, and you know that you won't suffer any failures of rationality, and you know that there will be no fundamental changes to your core values, and . . . ". This might rescue the principles from counterexamples, but the maneuver seems *ad hoc* unless there is some more general idea from which these caveats follow. And there is. It goes like this: if you think someone is, by your own lights, in a better position to make a wise decision than you are, you should heed their advice (assuming you know what it would be). And if each of the caveats hold — that is, if you know that the only change (if any) you'll undergo between now and then is that you'll become better-informed about some relevant matter — then, in general, you should now think that you will be in a better position to make a wise decision than you are now.

Let's call someone who knows at least as much as you do, who evaluates options according to rational standards that you endorse, and who shares your ends (i.e.,

[5] These worries about Reflection are not new. The first two have been raised by Talbott [1991] and Christensen [1991], respectively, as a problem for Belief Reflection [van Fraassen, 1984]. All three worries are discussed in [Joyce, 2007].

has the exact same array of preferences over outcomes as you) an *expert advisor.*[6] In many cases, we have good reason to regard our future-self as an expert advisor. And if you know that your expert advisor prefers φ to ψ, you should too; and you're not required to prefer φ to ψ if you know that your expert advisor doesn't prefer φ to ψ.

> [(No) Preference Deference]
>
> If you know that your expert advisor prefers φ to ψ, then you should prefer φ to ψ. (If you know that your expert advisor doesn't prefer φ to ψ, you aren't rationally required to prefer φ to ψ).

Deference can explain the appeal of Reflection: you should defer to your future-self insofar as you regard your future-self to be an expert advisor. Furthermore, *the spirit* of the arguments given above for *Two-Boxing* in the Newcomb Problem and for *Either* in Vacation Boxes did not essentially rely on it being *you* whose better-informed preference-like attitudes should be adopted. A better-informed, well-wishing bystander would work just as well.

Suppose you have a well-wishing friend who's equipped with X-ray specs. She can see the contents of the opaque boxes in the Newcomb Problem and Vacation Boxes. In the Newcomb Problem, she would prefer for you to *Two-Box*; in Vacation Boxes, she wouldn't prefer you taking the Larger box over the Regular box (or *vice versa*). You are in a position to know that your friend would have these attitudes. And because you know that she is better-informed than you, that she is rational, and that she wants what is best for you, you should regard her as an expert advisor. (Whether or not your X-Ray bespectacled friend *actually* exists is neither here nor there. What's required is only that you be in a position to know what array of preferences such a friend *would* have were she to exist.) According to Deference, then, you should prefer *Two-Boxing* to *One-Boxing* in the Newcomb Problem, and you aren't rationally required to prefer taking the Larger box over taking the Regular box in Vacation Boxes. Call this **The Argument from Deference.**[7]

---

[6] Joyce [2007] calls such a person "a decision-making expert." See [Gaifman, 1988] for a related discussion about "epistemic experts."

[7] Nozick [1969] offers something like this argument in order to motivate the intuition behind *Two-*

Let's look at one more argument that both supports *Two-Boxing* in the Newcomb Problem and *Either* in Vacation Boxes: **The Dominance Argument.** But, first, some terminology.

Decision problems — like the ones you face in the Newcomb Problem and Vacation Boxes — can be represented with three different entities: there are your *options* (or "alternatives," or "acts"), which are the objects of your instrumental preferences; there are the *outcomes* that might result from performing your options, which are the objects of your *non*-instrumental preferences; and there are *states*, which are those features of the world not under your control that influence the outcomes that might result from performing one of your options.[8] Following Savage [1954], we can think of an option as a *function* from states to outcomes. Or, following Jeffrey [1983], we can think of all three of these entities as *propositions* (which I'll take to be sets of possible worlds), where an option φ is a proposition of the form ⌜I do such-and-such⌝, a state S is a proposition concerning how (for all you know) the world might be, and the outcomes are propositions of the form $(\varphi \wedge S)$.

In the Newcomb Problem, the predictor has either placed a million dollars in the opaque box or she hasn't. If she has and you *One-Box*, you'll receive a million dollars; if you *Two-Box*, however, you'll receive the million dollars plus an additional thousand. On the other hand, if she hasn't placed a million dollars in the opaque box and you *One-Box*, you'll get nothing; whereas, if you *Two-Box*, you'll receive a thousands dollars. No matter which prediction the predictor has predicted, *Two-Boxing* does better than *One-Boxing*. Therefore, you ought to prefer *Two-Boxing* to *One-Boxing*.[9]

Boxing in the Newcomb Problem. Schlesinger [1974, 1976] also makes a similar argument in favor of *Two-Boxing*. Hare [2010] discusses this argument in the context of cases, like Vacation Boxes, of "opaque sweetening."

[8] See [Briggs, 2014] or [Resnik, 1987] as examples of explanations of decision theory that set things out in this manner. Also, not that "not under your control" is intentionally ambiguous between a *causal* and an *evidential* reading so as to remain neutral between causal and evidential decision theory. Later, this intentional ambiguity will be disambiguated: we should understand the states (relevant to decision theory) to be *dependency hypotheses,* which are "maximally specific proposition[s] about how the things [you] care about do and do not depend causally on [your] present actions." [Lewis, 1981].

[9] Nozick [1969] originally presented the Newcomb Problem as a case where the recommendations of expected utility theory conflicts with the Dominance Argument. Nearly all of the subsequent work on the Newcomb Problem has discussed, or at least made passing reference to, the Dominance

The argument appeals to a dominance principle:

> [Dominance]
>
> Let $\mathbf{S} = \{S_1, S_2, \ldots, S_n\}$ be a partition of the ways the world might be into states. If, for all $S_i \in \mathbf{S}$, you prefer $(\varphi \wedge S_i)$ to $(\psi \wedge S_i)$, then you ought to prefer φ to ψ.

A very similar dominance argument can be, and has been, given in support of *Either* in cases like Vacation Boxes.[10] The coin has either landed heads or it has landed tails. If the coin has landed heads, then the outcome that would result from taking the Larger box isn't preferred to the outcome that would result from taking the Regular box. If the coin has landed tails, then, similarly, the outcome that would result from taking the Larger box isn't preferred to the outcome that would result from taking the Regular box. Therefore, if the following dominance principle is correct, you are not rationally required to prefer taking the Larger box over taking the Regular box:

> [No Preference Dominance]
>
> Let $\mathbf{S} = \{S_1, S_2, \ldots, S_n\}$ be a partition of the ways the world might be into states. If, for all $S_i \in \mathbf{S}$, you don't prefer $(\varphi \wedge S_i)$ to $(\psi \wedge S_i)$, then you aren't rationally required to prefer φ to ψ.

Reflection, Deference, and Dominance each support both *Two-boxing* in the Newcomb Problem and *Either* in Vacation Boxes. This is no coincidence. The three

---

Argument for *Two-Boxing*. See, for example, [Jeffrey, 1983], [Joyce, 1999, 2007], [Sobel, 1985].

[10] Both Bales et al. [2014] and Rabinowicz [2016] discuss the Dominance Argument (approvingly, in the former case, but disapprovingly in the latter). Rabinowicz [2016] discusses a principle he calls *Complementary Dominance (V)*, which says: "One action is not better than another if it under every state yields an outcome that is not better than the outcome of the other action." Bales et al. [2014] argue, on intuitive grounds, for a principle they call *Competitiveness*. It says that it's rationally permissible to perform a competitive action, where an action is *competitive* if "for every way the world could be, its consequences are no worse than the consequences of all alternative actions." (pg. 460). These two principles are formulated differently, but the differences stop there.

arguments are closely related.[11] In fact, I think that these arguments owe their appeal to the very same underlying idea — roughly, that the way you evaluate your options should be sensitive to what you know about how the *actual values* of your options compare — and that this idea provides the best argument for *Two-Boxing* in the Newcomb Problem and *Either* in Vacation Boxes. In their current form, however, the Reflection, Deference, and Dominance arguments don't work; they all fall prey to the same class of counterexamples.

There are cases in which the Reflection, Deference, and Dominance principles offer bad advice. Here's one.[12]

> **The Big Test.** You have an important test tomorrow. You'd very much like to pass the test rather than fail it. Tonight, you have two options: you can *Study* or you can *Goof*. All else equal, you prefer goofing around to studying. What should you do?

	Pass	Fail
Study	20	0
Goof	25	5

Suppose that the test results will be mailed to your grandma. Grandma will open the envelope, read the results, and come to know whether you passed or failed the test. (Grandma won't know whether you opted to *Study* or to *Goof*, however. You should call her more.) Once Grandma sees the results, she will be better-informed than you currently are. Grandma is impeccably rational. And it goes without saying that Grandma wishes you well. You should, then, regard Grandma as an expert advisor. If Grandma learns that you passed, she will hope that you opted to *Goof* rather than *Study*. If Grandma learns that you failed, she will also hope that you

[11] As we've seen, the Deference principles entail suitably qualified versions of the Reflection principles. Furthermore, Deference also entails Dominance. If one option dominates another, then you are in a position to know that any *fully-informed* well-wisher would prefer you to choose the former option over the latter. Being fully-informed is one way of being better-informed, so if the antecedent of Dominance holds, then, by Deference, so does the consequent.

[12] This problem is well-known, and widely discussed. For some examples, see ([Arntzenius, 2008], pg. 287-9), (Jeffrey [1982], pg. 719-22), ([Joyce, 1999], pg. 115-8), ([Joyce, 2007], pg. 551). Nozick [1969], when first discussing the Newcomb Problem, raises this worry for Dominance.

opted to *Goof* rather than *Study*. So, either way, Grandma will hope that you opted to *Goof* over *Study*. You are in a position to know this. You know that an expert advisor — namely, Grandma — prefers you to *Goof* over *Study*. By Deference, then, you should prefer to *Goof* rather than to *Study*.

The same goes for Reflection and Dominance. For Reflection, tell the story above but replace Grandma with your future-self (who, for whatever whimsical reason you'd like, has selective-amnesia concerning whether you opted to *Study* or *Goof*). For Dominance, notice that you prefer the outcome $(\textit{Goof} \wedge \textsc{Pass})$ to the outcome $(\textit{Study} \wedge \textsc{Pass})$, and that you prefer the outcome $(\textit{Goof} \wedge \textsc{Fail})$ to the outcome $(\textit{Study} \wedge \textsc{Fail})$. $\{\textsc{Pass}, \textsc{Fail}\}$ is a partition of the ways the world might be. Relative to that partition, *Goofing* dominates *Studying*. By Dominance, then, you ought to prefer *Goofing* to *Studying*.

But it's obviously false that you ought to prefer *Goofing* to *Studying*. It's not (always) irrational to study! What's gone wrong here?

The standard diagnosis goes like this.[13] The problem is with the partition of states $\{\textsc{Pass}, \textsc{Fail}\}$. Studying makes it more likely that you'll pass the test, while goofing makes it more likely that you'll fail. What you decide to do influences what Grandma (or future-you) will learn when she (or future-you) opens the envelope. In general, we can't expect the Reflection, Deference, and Dominance principles to offer sensible advice when your options fail to be independent of the states (where "independence" means that your "choice of one act or another is thought to have no tendency to facilitate or impede the coming about of any of the possible states of nature" [Jeffrey, 1982]).

The standard diagnosis is standard for a good reason: it's undoubtedly correct. However, it is incomplete in two respects.

First, the way I've presented it here leaves "independent" ambiguous. An option φ is *evidentially independent* of a state S just in case you don't think that φing will provide you with any evidence that S is the case. (Formally, when your cre-

[13] For example, see ([Arntzenius, 2008], pg. 288-9), ([Jeffrey, 1983], pg. 21-22), ([Jeffrey, 1982], pg. 720), ([Joyce, 1999], pg. 116-9), ([Joyce, 2007], pg. 551). Arntzenius [2008] is concerned with the Reflection principle, Joyce [2007] with Reflection and Deference (as well as the Sure-Thing Principle [Savage, 1954], which he argues is equivalent to a version of Deference), and Jeffrey [1983, 1982] and Joyce [1999] with Dominance.

dences are such that $Cr(S \mid \varphi) = Cr(S \mid \neg\varphi)$, which is equivalent to $Cr(S \mid \varphi) = Cr(S)$, φ and S are evidentially independent). An option φ is *causally independent* of a state S just in case you think φing won't causally influence whether S is the case. (Formally, when $Cr(\varphi \mathrel{\square\!\!\rightarrow} S) = Cr(\neg\varphi \mathrel{\square\!\!\rightarrow} S)$, where '$\mathrel{\square\!\!\rightarrow}$' denotes a non-backtracking, causally-understood, subjunctive conditional, φ and S are causally independent).[14] If your options are evidentially independent of the states, then they will also be causally independent. After all, one way that φing might provide you with evidence for S is by causing it to be true. Because studying causally influences how you will fare on the test, in The Big Test, your options are neither causally nor evidentially independent of the states. When your options and the states fail to be *evidentially* independent, the Reflection, Deference, and Dominance principles aren't guaranteed to issue sensible advice. But what about cases — like the Newcomb Problem — in which your options and the states are causally, but not evidentially, independent?

Second, whether we interpret "independence" causally or evidentially, what *explains why* the principles go awry when it fails to hold? And what explains why the principles appear to get the right results when they're applied to cases in which your options *are* independent of the states?

In the next section, I will work toward answering these questions.

3 The Actual Value Conception

The Reflection, Deference, and Dominance principles are false. They issue incorrect verdicts in decision-problems where your options fail to be independent of the states. We can rescue the principles by restricting them to cases in which the appropriate kind of independence holds — although, as we've seen, what counts as "the appropriate kind of independence" is a matter of dispute — but what explains why they work (if they do) in these cases? And what explains why they don't work in the cases where independence doesn't hold?

[14] This way of cashing out causal independence — with non-backtracking, causally-understood, subjunctive conditionals — is by no means uncontroversial. The world might be indeterministic. Some of these subjunctive conditionals might be indeterminate. But this works as, at least, a rough approximation of the idea.

Here's what I will do. First, I will present what I take to be the best argument for *Two-Boxing* in the Newcomb Problem. I'll then show that this argument also supports *Either* in Vacation Boxes. Then I'll argue that it captures the underlying idea behind the Reflection, Deference, and Dominance arguments in a way that explains why they work only when there's independence between your options and the states of the world.

3.1 The Actual Value Conception: *Two-Boxing* & *Either*

Roughly, the argument goes like this. Ideally (if you were perfectly rational and omniscient), you would prefer one option to another when, and only when, it *actually* does a better job promoting your ends. So, if you're certain that φing would better serve your ends than ψing would, you ought to prefer φ to ψ. In the Newcomb Problem, you can be certain that taking both boxes would better serve your ends than taking only the one. So you should prefer *Two-Boxing* to *One-Boxing*.

In the service of presenting the argument more carefully, let's introduce some terminology. Let Cr be a probabilistically coherent credence-function representing your *subjective degrees of belief*. Let V be a utility-function, defined over the possible outcomes of your decisions, representing your *ends*.[15] Let's say that the *actual value* of an option is equal to the value you assign to the outcome that would, as a matter of fact, result from performing it.[16]

---

[15] When your preferences are incomplete (as they are in Vacation Boxes), it's unclear what your "ends" are. In the standard case, when your preferences *are* complete, by representing you with a utility-function, we treat you *as if* you have a single unified end — there is a single measurable quantity of value that we represent you as seeking to maximize. Of course, utility is not some precious fluid that decision theorists presuppose we all want to amass; rather, utility is a theoretical posit whose extreme flexibility allows it to represent the intrinsic valuing of anything, whatever it happens to be: money, happiness, pleasure, other people's happiness, other people's pain, jumping jacks, etc., or even things that cannot be easily expressed by finitely long strings of English. That being said, by assuming that your various ends — in all their variety and complexity — can be represented with a single utility-function (or, more precisely, a set of utility-functions all of which are positive linear transformations of each other), we thereby assume that all the potential tensions between your ends are (or, would be were you to consider them) resolved. And, in so doing, we treat you as if you have a single overarching end.

[16] I define actual value in terms of dependency hypotheses, but we could just as well define it in terms of non-backtracking, causally-understood, subjunctive conditionals. Let $X \mathrel{\square\!\!\rightarrow} S$ be such a conditional. (It says: *if it were true that X, then it would be true that S*.) Then, we can say: if $\varphi \mathrel{\square\!\!\rightarrow} o$, then

Actual Value. Let $K_@$ pick out the state of affairs that actually obtains.[17] It specifies how things are with respect to all of the features of the world that you care about which are outside your present influence.

$$V_@(\varphi) = V(\varphi \wedge K_@)$$

The actual value of φing is equal to the value you assign to the outcome picked out by $(\varphi \wedge K_@)$, which is the outcome that would actually result were you to φ.[18]

The available option with the most actual value is, in an objective sense, the best means to your ends. The regulative ideal governing instrumental rationality is to take the best means to your ends. Ideally, your preferences over your options would align with the facts concerning those option's actual values. To the extent that you can, you should adopt those preference that satisfy this regulative ideal. This is easy to do if you know the actual values of your options. Of course, in all but the most mundane of examples, you won't know the actual values of your options.

Preference, by nature, is comparative. So the facts about your options' actual values that are relevant to satisfying the Regulative Ideal should, likewise, be comparative. Whether your preference for φ over ψ conforms to the Regulative Ideal depends on how the actual value of φ *compares* to the actual value of ψ — the actual values of φ and ψ don't themselves matter *per se*. What matters is whether φ has more, or less, or the same amount of actual value as ψ (as well as the *extent* to which φ has more, or less, or the same amount of actual value as ψ.) The absolute amount of actual value had by φ (and ψ) matter only derivatively: the absolute amounts of

---

$V_@(\varphi) = V(o)$. In words: if, were you to φ, doing so would result in outcome o, then the actual value of φing is equal to the value you assign to outcome o. In some recent work, Spencer and Wells [2016] cash out actual value these terms.

[17] Understand $K_@$ to be a *dependency hypothesis:* a maximally specific proposition about how the things you care about depend causally on your options [Lewis, 1981]. The dependency hypotheses form a partition, and each dependency hypothesis is causally independent of your options.

[18] It's important that $K_@$ be a dependency hypothesis as opposed to just any state that actually obtains. What you do can affect which state is actual — studying will make it more likely that you'll pass the test, for example — and, so, the outcome that would result were you to perform one of your options isn't guaranteed to be the outcome that option has in the state that is actual unless, like dependency hypotheses, the states are causally independent of your options.

actual value *determine* the comparative facts, but it's the comparative facts — not the absolute ones — that matter.

In order to make the comparative nature of the Regulative Ideal clearer, let's introduce another bit of terminology. Let $\mathcal{CV}_@(\varphi, \psi)$ denote the extent to which option φ's actual value exceeds the actual value of option ψ. And let's say, in the special case when $V_@(\varphi)$ and $V_@(\psi)$ are both well-defined, that:

$$\mathcal{CV}_@(\varphi, \psi) = V_@(\varphi) - V_@(\psi)$$

More generally, let the function $\mathcal{CV}_K(\varphi, \psi)$ measure the degree to which option φ does better than option ψ in state K. And, again, when $V_K(\varphi)$ and $V_K(\psi)$ are both well-defined, $\mathcal{CV}_K(\varphi, \psi) = V(\varphi \wedge K) - V(\psi \wedge K)$. In cases like Vacation Boxes, where your ends cannot be represented with a utility-function, this equivalence won't hold. But let's accept it for the time being.[19]

> **The Regulative Ideal:** "Aim to be such that you strictly prefer one option to another if and only if the actual value of the former exceeds the actual value of the latter; aim to be indifferent between two options if and only if their actual values are equal."[20]

$$\varphi \succ \psi \text{ when, and only when } \mathcal{CV}_@(\varphi, \psi) > \mathcal{CV}_@(\psi, \varphi)$$
$$\varphi \approx \psi \text{ when, and only when } \mathcal{CV}_@(\varphi, \psi) = \mathcal{CV}_@(\psi, \varphi)$$

If you know how to satisfy the Regulative Ideal, you should do it. If you're certain that the actual value of φ exceeds the actual value of ψ (whether or not you know what φ's and ψ's actual values are), then you should prefer φ to ψ. Instrumental

[19] Toward the end of the paper, I'll offer a more general way of cashing this out to handle cases, like in Vacation Box, where $V_@(\varphi)$ and $V_@(\psi)$ aren't well-defined because you regard at least some of the outcomes of your decision to be on a par. There are other phenomena, in addition to having incomplete preferences, that might make taking the *comparisons* of your options' actual values to be the primitive notion helpful. For example, if you have intransitive preferences, $V_@$ isn't well-defined, but $\mathcal{CV}_@(\varphi, \psi)$ might be. Or if you regard some possible outcomes (e.g., spending an eternity in heaven) as infinitely valuable, $V_@(\varphi) - V_@(\psi)$ might fail to be well-defined but not $\mathcal{CV}_@(\varphi, \psi)$.

[20] Assuming, as we are in this section, that $\mathcal{CV}_@(\varphi, \psi) = V_@(\varphi) - V_@(\psi)$, then $\mathcal{CV}_@(\varphi, \psi) > \mathcal{CV}_@(\psi, \varphi)$ just in case $V_@(\varphi) - V_@(\psi) > V_@(\psi) - V_@(\varphi)$, which holds just in case $2 \cdot V_@(\varphi) > 2 \cdot V_@(\psi)$, which holds just in case $V_@(\varphi) > V_@(\psi)$.

rationality is about taking the best means to your ends. If you're certain that the actual value of φ exceeds the actual value of ψ, then you're also certain that, given how the world actually is, φ does a better job than ψ at securing your ends. So, on this picture of decision-theoretic instrumental rationality — what I've been calling **The Actual Value Conception** — you should prefer φ to ψ.

[Principle of Actual Value][21]

If you are certain that the actual value of φ exceeds the actual value of ψ, then you should prefer φ to ψ.

$$\text{If } Cr\Big(\mathcal{CV}_{@}(\varphi, \psi) > 0\Big) = 1, \text{ then } \varphi \succ \psi$$

If you are certain that the actual value of φ doesn't exceed the actual value of ψ, then you shouldn't prefer φ to ψ.

$$\text{If } Cr\Big(\mathcal{CV}_{@}(\varphi, \psi) > 0\Big) = 0, \text{ then } \varphi \not\succ \psi$$

This is the central idea behind the Actual Value Conception. You should strive, as best you can, to align your preferences over your options with the facts concerning how the actual values of those options compare. If you don't know how the actual

[21] This principle is more-or-less equivalent to a principle Schoenfield [2014] calls "Link." It says (where p is your rational credence function, and $\ulcorner V(X) > V(Y) \urcorner$ says that *the outcome that would actually result from choosing X is better than the outcome that would actually result from choosing Y*.):

> If $p\left(V(\varphi) > V(\psi)\right) = 0$ & $p\left(V(\psi) > V(\varphi)\right) = 0$, then $EV^{p}\left(\varphi\right) \not\succ EV^{p}\left(\psi\right)$ & $EV^{p}\left(\psi\right) \not\succ EV^{p}\left(\varphi\right)$.

In words: if you are rationally certain that the value of φing doesn't exceed the value of ψing (and *vice versa*), then neither should have higher expected value than the other.

Schoenfield [2014] defends Link by arguing that "if Link is rejected, expected value theory cannot play the role that it was intended to play: namely, providing agents with limited information guidance concerning how to make choices in circumstances in which value-based considerations are all that matter." (pg. 268). Schoenfield [2014] claims that it's central to the role we want expected value theory to play that it's recommendations not conflict with what you know about how the actual values of your options compare. However, it's not true that every version of expected value theory satisfies Link. The Newcomb Problem brings out that *evidential* decision theory violates the constraint. And, at least offhand, evidential decision theory is a satisfactory account of expected value. Schoenfield [2014]'s argument is persuasive only if we limit our attention to accounts of expected value, like causal decision theory, that are supported by the Actual Value Conception.

values of your options compare, it might be rational to prefer φ to ψ even if it turns out that φ's actual value doesn't exceed ψ's (in the same way that it can be rational to hold a false, but justified, belief). But if you *do* know how the actual values of your option compare, it ought to be reflected in your preferences over those options in the manner described. Otherwise, you will be in a position to know that your preferences violate the ideal.

The Principle of Actual Value supports *Two-Boxing* in the Newcomb Problem. Given the setup of the case, you're in a position to reason as follows:

Reasoning By Cases: Newcomb

P1 The opaque box either contains $\$M$ or $\$0$.

P2 If the opaque box contains $\$M$, then the actual value of *Two-Boxing* exceeds the actual value of *One-Boxing*.

P3 If the opaque box contains $\$0$, then the actual value of *Two-Boxing* exceeds the actual value of *One-Boxing*.

C The actual value of *Two-Boxing* exceeds the actual value of *One-Boxing*.

This is a valid argument.[22] And you are in a position to know each of the premises. You are, therefore, in a position to know the conclusion: *that the actual value of Two-Boxing exceeds the actual value of One-Boxing.* And so it's epistemically rational to

[22] Not all arguments of this form — i.e., reasoning by cases with indicative conditionals — are valid (as the much-discussed Miners Puzzle makes clear [Kolodny and McFarlane, 2010]). The reasoning leads us astray when the consequents of the indicative conditionals are "information-sensitive." But the As-a-Matter-of-Fact value of some option doesn't depend on what information you have; it depends only on which prizes are, as a matter of fact, in which box. The reasoning here is analogous to following non-puzzling Miners argument: "Either the miners are in shaft A or they are in shaft B; if they are in shaft A, then blocking neither shaft saves fewer lives than something else I could do; if they are in shaft B, then blocking neither shaft saves fewer lives than something else I could do; therefore, blocking neither shaft saves fewer lives than something else I could do." That's a fine argument. It would be a mistake, however, to take the conclusion to be a decisive reason to not block either shaft.

have the following credence:

$$Cr\Big(\mathcal{CV}_{@}(\textit{Two-Box}, \textit{One-Box}) > 0\Big) = 1$$

Therefore, by the Principle of Actual Value, you ought to prefer *Two-Boxing* to *One-Boxing*.

A similar argument shows that the Actual Value Principle supports *Either* in Vacation Boxes. Suppose that as a matter of fact (but, of course, unbeknownst to you) the Larger box contains A^+ and the Regular box contains B. The actual outcome that would result, then, from choosing the Larger box is the one in which you get A^+ and the actual outcome that would result from choosing the Regular box is the one in which you get B^+. If the world is as just described, then the actual value of taking the Larger box is equal to the value you assign to A^+ and the actual value of taking the Regular box is equal to the value you assign to B; and so the actual value of taking the Larger box doesn't exceed the actual value of taking the Regular box. Suppose instead that the coin had landed the other way. Analogous reasoning gets us to the same conclusion: the actual value of taking the Larger box doesn't exceed the actual value of taking the Regular box. You are in a position to know all of this, and are able to reason as follows:

Reasoning By Cases: Vacation Boxes

P1 The coin has landed either Heads or Tails.

P2 If the coin has landed Heads, then the actual value of taking the Larger box does not exceed the actual value of taking the Regular box.

P3 If the coin has landed Tails, then the actual value of taking the Larger box does not exceed the actual value of taking the Regular box.

C The actual value of taking the Larger box does not exceed the actual value of taking the Regular box.

Again, this is a valid argument, and you are in a position to know each of the premises. And so you should be certain that the actual value of taking the Larger box doesn't exceed the actual value of taking the Regular box. By the Principle of Actual Value, you shouldn't think that rationality requires you to prefer taking the Larger box to taking the Regular box.

In both the Newcomb Problem and Vacation Boxes, you are in a position to know something about how the actual values of your options compare. In the former case, you can be certain that the actual value of *Two-Boxing* exceeds the actual value of *One-Boxing*; in the latter, you can be certain that the actual value of taking the Larger box *doesn't* exceed the actual value of taking the Regular box. According to the Actual Value Conception, if you know how the actual values of your options compare, you should align your preferences over those options to these comparisons. The same underlying conception of rationality — the Actual Value Conception — supports *Two-Boxing* in the Newcomb Problem and *Either* in Vacation Boxes.

3.2 Reflection, Deference, & Dominance

The Actual Value Conception helps explain why the Reflection, Deference, and Dominance arguments seem appealing in some cases (the Newcomb Problem, Vacation Boxes) but not in others (the Big Test). The claim is this. These arguments seem appealing when they are deployed to cases in which you are in a position to know how the actual values of your options compare. The arguments are unappealing when that connection to actual value is lost.

Recall the lesson we drew from the Big Test: the Reflection, Deference, and Dominance principles offer bad advice when applied to decision-problems in which your options fail to be independent of the states. We can recast the decision you face in the Big Test, using dependency hypotheses, so as to ensure independence:

	$\overbrace{S \mathbin{\square\!\!\rightarrow} \textsc{Pass}}^{K_1}$	$\overbrace{S \mathbin{\square\!\!\rightarrow} \textsc{Fail}}^{K_2}$	$\overbrace{S \mathbin{\square\!\!\rightarrow} \textsc{Pass}}^{K_3}$	$\overbrace{S \mathbin{\square\!\!\rightarrow} \textsc{Fail}}^{K_4}$
	$G \mathbin{\square\!\!\rightarrow} \textsc{Pass}$	$G \mathbin{\square\!\!\rightarrow} \textsc{Pass}$	$G \mathbin{\square\!\!\rightarrow} \textsc{Fail}$	$G \mathbin{\square\!\!\rightarrow} \textsc{Fail}$
Study	20	0	20	0
Goof	25	25	5	5

When reformulated in this way, your options $\{Study, Goof\}$ no longer fail to be independent (either causally or evidentially) of the states $\{K_1, K_2, K_3, K_4\}$. But it is also no longer true that *Goofing* dominates *Studying*: in particular, K_3 — the dependency hypothesis according to which studying would result in passing and goofing would result in failing — is a state in which studying does better than goofing. It's also no longer tempting to think that Grandma (or future-you) should count as *fully-informed*: she (or future-you) would need to learn which of the four dependency hypotheses is the actual one, not merely whether you passed or failed.[23]

But why are these principles only applicable in cases where the states of the world are independent of your options? The Actual Value Conception can help explain. Here's the idea. When these principles are applied properly — that is, when we apply Dominance only to partitions of dependency hypotheses, and we understand "fully-informed" in the Deference and Reflection principles to mean "knows which dependency hypothesis is actual" — you will be in a position to know something about how the actual values of your options compare. To see why this is, let's look at each argument in turn:[24]

- **Reflection.** You are in a position to know that you will prefer having φed to having ψed. This is because you know that, after making your decision, the actual values of your options will be revealed to you. Furthermore, assuming that future-you will be rational and will value things in exactly the same way that you do now, you are in a position to infer from the fact that future-you will prefer having φed to having ψed that the actual value of φing exceeds the actual value of ψing. So, you are now in a position to know that the actual

[23] It's true that by learning whether you passed or failed, Grandma (or future-you) is *better-informed* than you are now. For example, learning that you passed the test is equivalent to learning $\big((Study \wedge K_1) \vee (Goof \wedge K_1) \vee (Goof \wedge K_2) \vee (Study \wedge K_3)\big)$, which is more than can be said for you. And it's true that, out of these possibilities, the *G*-worlds (the worlds in which you opted to goof around) are better than the *S*-worlds (the worlds in which you opted to study). But, because you expected your choice to causally influence whether Grandma (or future-you) learns that you passed or that you failed, it's not clear that you expect Grandma (or future-you) to be in a better epistemic position concerning the extent to which the actual value of studying exceeds the actual value of goofing. And so it's not clear that Grandma (or future-you) should count as better-informed in the relevant sense.

[24] The arguments concerning the cases in which you're in a position to know that you will *not* prefer having φed to having ψed proceed analogously.

value of φing exceeds the actual value of ψing.

- **Deference.** You are in a position to know that your expert advisor, in virtue of being fully-informed, is aware of your options' actual values. Given that your expert advisor knows the options' actual values, she should prefer one to the other if and only if the former has more actual value than the later. Because your expert advisor wants you to have φed rather than ψed, you can infer that φing has more actual value than ψing.

- **Dominance.** Partition the states of the world into dependency hypotheses. If ψ is dominated by φ, then you are in a position to know that the actual value of ψ doesn't exceeded the actual value of φ. Here's why. You know that $V_@(\psi) = V(\psi \wedge K_@)$ and $V_@(\varphi) = V(\varphi \wedge K_@)$. If ψ is dominated by φ, then you are in a position to know, for each dependency hypotheses K, that $V(\psi \wedge K) \leq V(\varphi \wedge K)$. And so, even though you might not know which dependency hypothesis is actual (i.e., for each K, you don't know if $K = K_@$), you are in a position to know that, whichever it is, the value of ψ's outcome in that state doesn't exceed the value of φ's. But the values of these outcomes correspond to their respective option's actual values. (In other words, you can think of each dependency hypothesis as corresponding to a hypothesis about how the actual values of φ and ψ might compare. If φ dominates ψ, then every such hypothesis is one according to which the actual value of φ exceeds the actual value of ψ.) Therefore, you are in a position to know that the actual value of φ doesn't exceed the actual value of φ.

Each argument dramatizes the fact that you are in a position to know that the actual value of φing exceeds the actual value of ψing. If you know that the actual value of some option exceeds the actual value of another, then, by the Principle of Actual Value, rationality requires you to prefer it. (And *mutatis mutandis* for those cases in which you're in a position to know that the actual value of φing *doesn't* exceed the actual of ψing).

However, when these principles are misapplied, the connection to actual value is lost. For example, from the fact that *Goofing* dominates *Studying* relative to the partition $\{\textsc{Pass}, \textsc{Fail}\}$, you cannot infer anything of interest about how the actual

values of your options compare. It would be wrong to conclude, for example, that *Goofing* has more actual value than *Studying* — if K_3 describes the way the world actually is, then *Studying* has more actual value than *Goofing*. The states {PASS, FAIL}, unlike the states $\{K_1, K_2, K_3, K_4\}$, do not correspond to hypotheses concerning how the actual values of your options might compare.

So long as these principles are applied properly, whenever their antecedents hold, you will be in a position to be rationally certain about how the actual values of your options compare. When you're rationally certain about how the actual values of your options compare, the Actual Value Conception recommends aligning your preferences over those options with what you know about those comparisons. The Reflection, Deference, and Dominance arguments are all ways of dramatizing that you are in a position to know something relevant about how your options' actual values compare.

4 Actual Value Decision Theory, Part I

I've been arguing that there's an affinity between *Two-Boxing* in the Newcomb Problem and *Either* in Vacation Boxes in that the best argument for the former position is also the best argument for the latter position. In both cases, you're in a position to know something about how the actual values of your options compare. In the former, it's that the actual value of *Two-Boxing* exceeds the actual value of *One-Boxing*; in the latter, it's that the actual value of taking the Larger box *does not* exceed the actual value of taking the Regular box. I motivated a principle, which I called "the Principle of Actual Value," that says: if you're in a position to know that the actual value of one option exceeds the actual value of another, then you ought to prefer it; and if you're in a position to know that the actual value of an option doesn't exceed the actual value of another, you aren't rationally required to prefer it.

When you're not in a position to be certain about how the actual values of your options compare, what should you do? The Principle of Actual Value doesn't say. I will present one way of developing the idea underlying the Principle of Actual Value into a full decision theory. Let's call it **Actual Value Decision Theory**.

I'll proceed in two steps. First, in this section, I will ignore decision-problems

like Vacation Boxes by focusing exclusively on the special case where your ends can be represented with a utility-function. I will show that Actual Value Decision Theory entails the Principle of Actual Value and is, thus, a generalization of it. I will then show that, in this special case, Actual Value Decision Theory is equivalent to causal decision theory. In the next section, I will sketch a way of generalizing the view to decision-problems, like Vacation Boxes, in which your ends *cannot* be represented with a utility-function.

4.1 Actual Value Estimates & Causal Decision Theory

The idea underlying the Principle of Actual Value is that, ideally, you would align your preferences over your options to the facts concerning how the actual values of those options compare. If you know that the actual value of an option exceeds the actual value of another, then, you should prefer it. How should you evaluate your options when you don't know how their actual values compare?

Here's the suggestion. Roughly: you should align your preferences over your options to your best *estimates* of how the actual values of those options compare. Less roughly: in evaluating the respective merits of options φ and ψ, first, use your credences to estimate the extent to which the actual value of φ exceeds the actual value of ψ, and compare that estimate with your estimate of the extent to which the actual value of ψ exceeds the actual value of φ. You are instrumentally rational insofar as your instrumental preferences match, not the comparisons in actual value of one's options *themselves*, but your best *estimates* of those comparisons.

> **Estimate Comparisons of Actual Value Rule:** "Prefer option φ to option ψ when, and only when, your best estimate of the extent to which φ's actual value exceeds ψ's actual value is greater than your best estimate of the extent to which ψ's actual value exceeds φ's actual value."

$$\varphi \succ \psi \quad \text{when, and only when,} \quad \textsc{Estimate}\Big[\mathcal{CV}_{@}(\varphi, \psi)\Big] > \textsc{Estimate}\Big[\mathcal{CV}_{@}(\psi, \varphi)\Big]$$

$$\varphi \approx \psi \quad \text{when, and only when,} \quad \textsc{Estimate}\Big[\mathcal{CV}_{@}(\varphi, \psi)\Big] = \textsc{Estimate}\Big[\mathcal{CV}_{@}(\psi, \varphi)\Big]$$

Supposing you're rational, what should your *best estimates* of these actual value

comparisons look like? There's room for disagreement here, but for the purposes of this paper I will assume that the best estimate of $\mathcal{CV}_@(\varphi, \psi)$ is the weighted average of how much the actual value of φ might exceed the actual value of ψ, where the weights correspond to your *unconditional* credences in hypotheses about how the actual values of these options might compare.[25]

[Estimating Comparisons of Actual Value]

The best estimate of the extent to which the actual value of φ exceeds the actual value of ψ:

$$\textsc{Estimate}\Big[\mathcal{CV}_@(\varphi, \psi)\Big] = \sum_v Cr\big(\mathcal{CV}_@(\varphi, \psi) = v\big) \cdot v$$

In the rest of this section, I'll show how Actual Value Decision Theory relates to causal and evidential decision theory. I'll then prove that Actual Value Decision Theory entails the Principle of Actual Value.

Actual Value & Causal Decision Theory. If the Actual Value Decision Theory sounds familiar, it should: it's one particular way of spelling out the central idea motivating causal decision theory. One way of describing what causal decision theory says is as follows: when facing a decision, first partition the states into dependency hypotheses (which are maximally specific propositions about how the things you care about depend causally on your options); then, for each of these dependency

[25] Why *unconditional* as opposed to *conditional* credence? That is, why not take your best estimate of of $\mathcal{CV}_@(\varphi, \psi)$ to be the weighted average of how much the actual value of φ might exceed the actual value of ψ, where the weights correspond, roughly, to the credences you *would have* in the hypotheses about how the actual values of these options might compare *were you to learn that you φed*? I think there are good reasons to think that unconditional estimates are epistemically better than conditional estimates, but a full defense of this claim is outside the scope of the paper. (See the accuracy-dominance argument in [Pettigrew, 2015] for an indication of such a defense might go). In any case, on either proposal, Actual Value Decision Theory is incompatible with evidential decision theory. I'll show that, if your estimates reflect your unconditional credences, then the Actual Value Decision Theory is equivalent to causal decision theory. (In fact, the causal expected utility of an option *just is* your best unconditional estimate of that option's actual value). On the other hand, if your estimates reflect your *conditional* credences instead, the view entails Wedgwood [2013]'s benchmark decision theory. I focus on the former case in the main text for the sake of presentation, but a discussion of the latter proposal can be found in the appendix.

hypotheses, find the values your option has if that dependency hypothesis is true; the value of your option is the weighted average of these values, where the weights correspond to your credence in each dependency hypothesis being the one that actually holds.[26]

[Causal Expected Utility]

The causal expected utility of an option, φ, is the weighted average of the values you assign, for each dependency hypothesis, to the outcome that would result from φ-ing if that dependency hypothesis is true, and where the weights correspond to your unconditional credences in the dependency hypotheses.

$$U(\varphi) = \sum_K Cr(K) \cdot V(\varphi \wedge K)$$

The causal utility of φ is greater than the causal utility of ψ if and only if your unconditional estimate of the extent to which the actual value of φ exceeds the actual value of ψ is greater than zero. Actual Value Decision Theory, therefore, underlies causal decision theory.

I'll present only a sketch of the idea here. A fuller statement of the proof can be found in the appendix.

Recall that your unconditional estimate of $\mathcal{CV}_@(\varphi, \psi)$ is the weighted average of all the ways the actual value of φ might exceed the actual value ψ, where the weights correspond to your unconditional credences in the various hypotheses about the

[26] Two quick clarifications. First, I will, following Lewis [1981], use U to denote an option's *causal* expected value and V to denote the evidential expected value of a proposition: $V(X) = \sum_Z Cr(Z \mid X) \cdot V(X \wedge Z)$. The evidential expected value (or "news value") of a proposition measures how good you would expect the actual world to be were you to learn that it's true. Within a dependency hypothesis, a proposition's value is its evidential expected value. Second, there are several other versions of decision theory which don't make reference to dependency hypotheses. Some versions, like [Sobel, 1978] and [Joyce, 1999], define expected value using *imaging*. Other versions, like [Gibbard and Harper, 1978] and [Stalnaker, 1981], appeal to probabilities of subjunctive conditionals. However, as Lewis [1981] convincingly argues, given various plausible assumptions, these other versions of causal decision theory are notational variants of each other. What I say here could just as well, although perhaps less perspicuously, be formulated using one of these other versions.

ways the actual values of φ and ψ might compare:

$$\textsc{Estimate}\Big[\mathcal{CV}_@\,(\varphi, \psi)\Big] = \sum_v Cr\big(\mathcal{CV}_@\,(\varphi, \psi) = v\big) \cdot v$$

Given how we've characterized actual value in terms of dependency hypotheses, the proposition that $\mathcal{CV}_@\,(\varphi, \psi) = v$ is equivalent to the following disjunction of conjunctions:

$$\bigvee_{K_i}\Big(\mathcal{CV}_{K_i}\,(\varphi, \psi) = v \wedge K_i\Big)$$

Because dependency hypotheses are mutually exclusive, your credence in the hypothesis that $\mathcal{CV}_@\,(\varphi, \psi) = v$ can be expressed as the sum of your credences in each of the disjuncts.

$$Cr\big(\mathcal{CV}_@\,(\varphi, \psi) = v\big) = \sum_K Cr\Big(\mathcal{CV}_K\,(\varphi, \psi) = v \wedge K\Big)$$

And because each dependency hypothesis determines a way that the actual values of your options might compare, your credence in the hypothesis that $\mathcal{CV}_@\,(\varphi, \psi) = v$ equals the sum of your credences in those dependency hypotheses according to which, if it is actual, then the actual value of φ exceeds the actual value of ψ by the amount v. In other words, consider all dependency hypotheses, K, such that $\mathcal{CV}_K(\varphi, \psi) = v$; your credence that $\mathcal{CV}_@\,(\varphi, \psi) = v$ equals the sum of your unconditional credences in each of these Ks. Because each dependency hypothesis corresponds to exactly one hypothesis concerning how the actual values of your options might compare, it follows that:

$$\textsc{Estimate}\Big[\mathcal{CV}_@\,(\varphi, \psi)\Big] = \sum_K Cr\,(K) \cdot \mathcal{CV}_K(\varphi, \psi)$$

Therefore, according to Actual Value Decision Theory, you should prefer φ to ψ when, and only when, $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) > \sum_K Cr(K) \cdot \mathcal{CV}_K(\psi, \varphi)$. And, because, for each K, $\mathcal{CV}_K(\varphi, \psi) = V(\varphi \wedge K) - V(\psi \wedge K)$, that inequality holds just

in case:

$$\sum_K Cr(K) \cdot V(\varphi \wedge K) > \sum_K Cr(K) \cdot V(\psi \wedge K)$$

Which is to say: *just in case the causal expected utility of φ is greater than the causal expected utility of ψ*. Therefore, Actual Value Decision Theory entails causal decision theory.

Actual Value & **Evidential Decision Theory.** Evidential decision theory says that you should prefer one option to another if and only if the expected *evidential* value of the former exceeds that of the latter, where the expected evidential value of an option is, roughly, your estimate of how good the actual world would be were you to learn that you performed that option.

> [Evidential Decision Theory]
>
> You should prefer φ to ψ if and only if the evidential expected value of φ exceeds the evidential expected value of ψ.
>
> $$V(\varphi) = \sum_Z Cr(Z \mid \varphi) \cdot V(\varphi \wedge Z)$$

Evidential decision theory, as one might expect, is incompatible with Actual Value Decision Theory. Evidential decision theory will sometimes recommend preferring one option to another even when you're certain that the former has less actual value than the latter. The Newcomb Problem serves as an example.

The Principle of Actual Value. Actual Value Decision Theory entails the Principle of Actual Value.[27] According to Actual Value Decision Theory, you should prefer φ to ψ if and only if $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) > 0$. In order to show that the

[27] Both causal and benchmark decision theory, as one might suspect, entail the Actual Value Principle. In the main text, I show that the version of Actual Value Decision Theory that entails causal decision theory also entails the principle. The proof that benchmark decision theory entails the principle can be found in the Appendix.

principle follows, we'll assume that $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$ and, then, show that $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$.

Here's a sketch of the idea (see the appendix for a fuller presentation of the proof). If $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then, for every dependency hypothesis K, either: (1) the value of φ's outcome in K *doesn't* exceed the value of ψ's outcome in K, or (2) you are certain that K is not true, or (3) both. Therefore, $Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \leq 0$, for every K. And thus, $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$.

Therefore, if $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$. And, so, according to Actual Value Decision Theory, you should not prefer option φ to option ψ.

In the next section, I will generalize Actual Value Decision Theory to cases, like Vacation Boxes, in which your ends cannot be represented with a utility-function because you regard the possible outcomes of your decision to be on a par.

5 Actual Value Decision Theory, Part II

Classic formal models of instrumental rationality require that one's preference be *complete* (or, *trichotomous*): for any things, X and Y, that you regard as comparable, either you prefer X to Y, or you prefer Y to X, or you are indifferent between the two.[28] There are, however, a growing number of philosophers and economist who argue that practical rationality requires no such thing.[29] When your preferences over outcomes are incomplete, it's not clear that the possible consequences of a decision can be said to have an unequivocal value — and, so, it's also unclear how your (incomplete) preferences over outcomes are to constrain your preferences over your options.

Here's what we'll do. First, I'll present three desiderata that, in my view, any adequate decision theory for agents with incomplete preferences must satisfy in

[28] See, for example, [Savage, 1954], [von Neumann and Morgenstern, 1944], and [Anscombe and Aumann, 1963].

[29] See, for example, [Chang, 2002, 2005] [Dubra et al., 2004], [Evren and Ok, 2011], [Galaabaatar and Karni, 2013], [Hare, 2010], [Herzberger, 1973], [Joyce, 1999], [Levi, 1986, 1999, 2006], [Nau, 2006], [Ok et al., 2012], [Raz, 1985], [Seidenfeld et al., 1990, 1995], [Sen, 2004]. Even the developers of the classic models, for example [Aumann, 1962] and [Savage, 1954], express doubts that the Completeness Axiom is an honest-to-goodness constraint imposed by rationality.

order to count as respecting the Actual Value Conception. Then, I'll present an idealized version of the decision theory (one that makes an unrealistic assumption about your value-structure: namely, that, whenever two goods, X and Y, are on a par, there is always a precise amount of value that can be added to X such that it is the least amount of value that needs to be added in order for X plus it to be preferred to Y). Next, I'll show that the view satisfies the three desideratum. Finally, I'll sketch how the view can be weakened to handle cases in which this unrealistic assumption fails to hold.

5.1 The Three Desiderata

Think of a decision theory like a helpful advisor: you give your advisor information about how you take the the world to be, and information about how you value outcomes, and your advisor issues recommendations about what you rationally ought to do. Standard decision theories require a great deal of information about how you value outcomes in order to issue recommendations: they require you to have complete preferences over all possible outcomes, and that, for any two outcomes, there be a determinate fact about the precise degree to which you prefer the one to the other. Without this information, standard decision theories remain *silent* — they are unable to offer any recommendations; they have nothing to say about what rationality requires or permits you to do. As we've seen, the Actual Value Conception requires slightly less of you: your advisor only needs information about how the values of the outcomes in the same dependency hypothesis compare. In particular, it requires, for any options, φ and ψ, and for each dependency hypothesis K, that there be some real number r such that $\mathcal{CV}_K(\varphi, \psi) = r$.[30] Even this, I think, requires more from you than is needed.

Any adequate decision theory for agents with incomplete preferences should be *robust:* it should require less input than the standard views in order to issue recommendations. In order to respect the Actual Value Conception, the proposal should be a generalization of a standard decision theory that's supported by the conception. Furthermore, in cases like Vacation Boxes, where you're certain that

[30] More carefully, it requires, for any two options, φ and ψ, and for each K, that, *given a conventionally chosen zero-point and scale*, there be a real number r such that $\mathcal{CV}_K(\varphi, \psi) = r$.

the actual outcomes of your decision are on a par, the proposal should avoid recommending that you prefer either option to the other. But, in order to be robust, the proposal shouldn't be rendered completely silent in all but the most trivial cases of parity. In other words, any adequate decision theory for agents with incomplete preferences that respects the Actual Value Conception should meet the following three desiderata:

First, the view should be a generalization of a version of Expected Utility Theory that is supported by the Actual Value Conception. The view developed here is a generalization of causal decision theory: if you had complete preferences, the view should give the same recommendations as causal decision theory.[31]

Second, in Vacation Boxes, the view shouldn't recommend preferring the Larger box to the Regular box. More generally, the view should entail the Principle of Actual Value: if, when considering two options, you are position to be rationally certain that the actual value of the former doesn't exceed the actual value of the latter, you shouldn't prefer the former to the latter.

Lastly, if you are sufficiently confident that the actual value of φ exceeds the actual value of ψ to a significant extent, then, even though it *might* be the case that the outcome that would result from performing φ is on a par with the outcome that would result from performing ψ, the view should recommend preferring φ to ψ. In other words, the view should be capable of offering non-trivial recommendations in the face of parity. Here's an example.

> **Probabilistic Sweetening.** There are two boxes in front of you: the Larger box and the Regular box. A *biased* coin has been tossed. If it landed heads, then $\$1,000,000$ has been place in the Larger box and $\$0$ has been placed in the Regular box. If the biased coin landed tails, then a *fair* coin was tossed. If the fair coin landed heads, then *A* has been placed in the Larger box and *B* has been placed in the Regular box. If the coin landed tails, then *B* is in the Larger box and *A* is the Regular box.

[31] It will not be difficult to see how the proposal can be amended in order to generalize benchmark decision theory instead. For the sake of presentational perspicuity, however, I will only focus on the version that generalizes causal decision theory.

	K_1	K_2	K_3
L	A	B	\$1,000,000
R	B	A	\$0

Even though you regard A and B as on a par, if your credence that the biased coin landed heads is sufficiently great, then you ought to prefer L to R. However, if your credence in the biased coin landing heads is sufficiently low, then you shouldn't be rationally required to prefer L to R. We want a decision theory that, in cases like these, offers recommendations that are sensitive to your credence in receiving the money if you take L. We also want the decision theory's recommendations, in cases of this sort, to be sensitive to how much money you might win if you take L, and how valuable you take receiving that sum of money to be.

5.2 The "Elasticity" of Parity

Recall that, according to Actual Value Decision Theory, you should align your preferences over your options with your best estimates of how the actual values of those options compare.[32]

Can the view be generalized to cases in which the outcomes of your options might be on a par? If you give some credence to dependency hypothesis K being actual, and you regard outcome $(\varphi \wedge K)$ as on a par with outcome $(\psi \wedge K)$, how should this be reflected in your estimation of the extent to which the actual value of φ might exceed the actual value of ψ?

Here's a sketch of the proposal. You want to estimate the extent to which the actual value of an option φ exceeds the actual value of an option ψ. Partition the ways the world might be into dependency hypotheses. Each dependency hypothesis determines a possible way the actual values of your options, for all you know, might compare. Speaking somewhat-metaphorically: the dependency hypotheses

[32] And, also recall, that the idea behind the Actual Value Conception is that, ideally, you should match your preference-like attitudes over your options to the facts concerning how the actual values of those options compare. So, for example, if you prefer outcome $(\varphi \wedge K)$ to outcome $(\psi \wedge K)$, then, if K is the way the world actually is, the actual value of φ exceeds the actual value of ψ. Similarly, I think that if you regard $(\varphi \wedge K)$ as on a par with $(\psi \wedge K)$, then, if K is the way the world actually is, the actual values of φ and ψ are on a par.

according to which φ does better than ψ (if there are any) contribute a *positive* amount to your estimate of the extent to which φ has more actual value than ψ; the ones according to which ψing does better (if there are any) contribute a *negative* amount to your estimate; the dependency hypotheses according to which φ and ψ are equally good (if there are any) contributes *nothing*, positive or negative, to your estimate.

What about the dependency hypotheses (if there are any) according to which φ and ψ are on a par? If outcomes $(\varphi \wedge K)$ and $(\psi \wedge K)$ are on a par, then, if K is the way the world actually is, the actual value of φ doesn't exceed the actual value of ψ, and *vice versa*; so, hypothesis K neither contributes a positive nor a negative amount to your estimate. Does it contribute *nothing*? No. Parity is *elastic*: if two outcomes are on a par, small improvements (in either direction) won't break the parity. On my proposal, hypotheses according to which your options are on a par contribute some "elasticity" to your estimate of the extent to which an option's actual value exceeds another.

What do I mean by *"elasticity"*? If A and B are on a par, there will be a range of improvements to A, and a range of diminishments to A, that will also be on a par with B; and likewise for B: there will be a range of improvements, and diminishments, that will be on a par with A. The parity between A and B is *maximally elastic* if no matter how much we improve (or diminish) one of them, you still regard them as on a par. Although there might be cases of parity which are maximally elastic, it's implausible that all cases are. (Surely you'd prefer the alpine ski vacation plus *a trillion dollars* to the beach vacation, for example!) In many cases, there are limits to the elasticity — we can place upper and lower bounds on the extent to which the outcomes are on a par. In particular, suppose that you strictly prefer prize A plus $\$y$ to prize B, and suppose that you strictly prefer prize B plus $\$x$ to prize A. Then, the extent to which A and B are on a par is bounded by the values you assign to those sums of money. We can interpret these bounds as placing limits on the extent to which the value of A fails to exceed the value of B (and *vice versa*).

$$-V(\$y) < \mathcal{CV}(A, B) < V(\$x)$$
$$-V(\$x) < \mathcal{CV}(B, A) < V(\$y)$$

If you would prefer B plus $\$x$ to A, then the extent to which A is on a par with B can't exceed $V(\$x)$; and if you would prefer A plus $\$y$ to B, then the extent to which B is on a par with A can't exceed $V(\$y)$. When A and B are on a par, your assessment of the extent to which A's value exceeds B's value reflects the extent to which the two are on a par by being *unsharp:* there is no precise amount such that the value of A exceeds, or falls short of, the value of B; rather, there is a range, or interval, of values capturing the extent to which the two are on a par.

Let $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$ be the *least upper bound* on the extent to which the value of outcome $(\varphi \wedge K)$ exceeds the value of outcome $(\psi \wedge K)$, and let $\lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor$ be the *greatest lower bound*.[33]

To assume that $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$ and $\lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor$ are well-defined is an unrealistic idealization. If you regard two goods to be on a par, we shouldn't expect there to be a precise amount of value such that it is the least amount that needs to be added to the one in order for it to be preferred to the other. I'll provisionally assume that these quantities *are* well-defined in order more easily state the proposal, and then show how this unrealistic idealization can be relaxed.

5.3 Actual Value Decision Theory, Generalized

You should align your preferences over outcomes with your estimate of how the actual values of those options compare. Your best estimate of the extent to which the actual value of φ exceeds the actual value of ψ is the weighted average of all the various ways their actual values might compare, where the weights correspond to your credences in the hypotheses about those comparisons. However, if the actual values of your options might be on a par — so that there is no precise fact of the matter about the extent to which the actual value of one of the options exceeds, or falls short of, the actual value of the other — then your *estimate* might also fail to be precise.

Because we've placed upper and lower bounds on $\mathcal{CV}_K(\varphi, \psi)$, we can likewise

[33] If these outcomes are on a par (and the parity isn't maximally elastic), is it guaranteed that there will be a *least* upper bound and a *greatest* lower bound? I don't think so. There needn't be a precise amount of value such that were A improved by exactly that amount, nothing more and nothing less, you will strictly prefer it to B. This is a limitation of the way we're modeling parity, and it's a limitation that will be inherited by the decision theory proposed in the next section.

place bounds on your estimates in the following way:

$$\textsc{Estimate}\Big[\lceil \mathcal{CV}_@(\varphi,\psi)\rceil\Big] = \sum_K Cr(K)\cdot\lceil \mathcal{CV}_K(\varphi,\psi)\rceil$$

$$\textsc{Estimate}\Big[\lfloor \mathcal{CV}_@(\varphi,\psi)\rfloor\Big] = \sum_K Cr(K)\cdot\lfloor \mathcal{CV}_K(\varphi,\psi)\rfloor$$

You should prefer φ to ψ when, and only when, your estimate of the extent to which the actual value of φ exceeds the actual value of ψ is greater than zero. When your actual value estimate is interval-valued, you should prefer φ to ψ when, and only when, the *lower bound* of your estimate is greater than zero.[34] You should regard your options to be on par when, and only when, the lower bound of your estimate is less than zero and the upper bound is greater than zero. If both bounds are the same, and equal to zero, you ought to be indifferent between the two.

> **Actual Value Decision Theory:** "Prefer option φ to option ψ when, and only when, the lower bound on your estimate of the extent to which the actual value of φ exceeds the actual value of ψ is greater than zero; the options are on a par ($\varphi \bowtie \psi$) when, and only when, the lower bound on your estimate is less than zero and the upper bound is greater than zero."

$$\varphi \succ \psi \text{ when, and only when } \textsc{Estimate}\Big[\lfloor \mathcal{CV}_@(\varphi,\psi)\rfloor\Big] > 0$$

$$\varphi \approx \psi \text{ when, and only when } \textsc{Estimate}\Big[\lfloor \mathcal{CV}_@(\varphi,\psi)\rfloor\Big] = 0 = \textsc{Estimate}\Big[\lceil \mathcal{CV}_@(\varphi,\psi)\rceil\Big]$$

$$\varphi \bowtie \psi \text{ when, and only when } \textsc{Estimate}\Big[\lfloor \mathcal{CV}_@(\varphi,\psi)\rfloor\Big] < 0 < \textsc{Estimate}\Big[\lceil \mathcal{CV}_@(\varphi,\psi)\rceil\Big]$$

As we've seen, even if it is more likely than not that the actual value of ψ exceeds the actual value of φ, if there's a sufficiently large enough chance that the actual value of φ might exceed the actual value of ψ to a great extent, then you should prefer φ to ψ.

[34] Notice that $\lfloor \mathcal{CV}_K(\varphi,\psi)\rfloor = -\lceil \mathcal{CV}_K(\psi,\varphi)\rceil$, and $\lceil \mathcal{CV}_K(\varphi,\psi)\rceil = -\lfloor \mathcal{CV}_K(\psi,\varphi)\rfloor$. So, this is equivalent to saying that you should prefer φ to ψ when, and only when, $\textsc{Estimate}\big[\lceil \mathcal{CV}_@(\psi,\varphi)\rceil\big] < 0$.

In estimating the actual value comparison between φ and ψ, the "losses" in some states can be outweighed by the "gains" in others. If outcome $(\varphi \wedge K)$ and outcome $(\psi \wedge K)$ are on a par, then it's not the case that receiving either outcome constitutes a "loss" or a "gain." But, if we can place bounds on the extent to which these outcomes are on a par — that is, if we can say for each outcome, how much value would need to be added (or subtracted) in order for you to prefer (or disprefer) it to the other — then we can, in effect, also place bounds on how large, or small, the "gains" (or "losses") in the other states must be in order to outweigh the parity between outcomes in others. Just as the parity between two prizes can be overcome if one of the prizes is sufficiently improved, the "elasticity" of your actual value estimate, inherited from the parity of those options' outcomes, can be overcome if there's a sufficient chance that the actual value of one of your options might exceed the actual value of the other by a significant enough extent.[35]

5.4 Actual Value Decision Theory Satisfies the Three Desiderata

Actual Value Decision Theory satisfies the desiderata mentioned above: (1) it is a generalization of causal decision theory; (2) in Vacation Boxes, the proposal says that you should regard taking the Larger box and taking the Regular box as on a par; and (3) there are non-trivial cases in which the proposal *does* recommend preferring one option to another even though you regard some (but, crucially, not all) of the outcomes in the same states of the world to be on par.

---

[35] Here's one way to motivate part of what the decision theory says. Suppose you are deciding between option φ and ψ, which might, for all you know, be on a par. Now, let's introduce the following *"virtual"* option: φ^*, which is just like φ except that, in each state, it's outcome is augment by $\lceil \mathcal{CV}_K(\psi, \varphi) \rceil$ (that is to say, if $\lceil \mathcal{CV}_K(\psi, \varphi) \rceil > 0$, the outcome is improved by exactly that amount, and if $\lceil \mathcal{CV}_K(\psi, \varphi) \rceil < 0$, it is diminished by exactly that amount). You ought to (weakly) prefer φ^* to ψ. Here's why. In those states in which the outcomes are *not* on a par, $\mathcal{CV}_K(\varphi^*, \psi) = 0$ because the value of φ^*'s outcomes have been adjusted so as to equal the value of ψ's outcomes. In those states in which the outcomes *are* on a par, $\mathcal{CV}_K(\varphi^*, \psi) > 0$ because (i) $\lceil \mathcal{CV}_K(\psi, \varphi) \rceil$ is the smallest amount that $(\varphi \wedge K)$ needs to be improved in order to be preferred to ψ, and (ii) $(\varphi^* \wedge K)$ is just like $(\varphi \wedge K)$ except that it's been improved by exactly that amount. So, you ought to prefer φ^* to ψ. But, also, if $\sum_K Cr(K) \cdot \lceil \mathcal{CV}_K(\psi, \varphi) \rceil < 0$, you ought to prefer φ to φ^*. Our recipe for defining φ^* ensures that its outcomes are comparable to φ's in the same states. And, in particular, for each K, $\mathcal{CV}_K(\varphi^*, \varphi) = \lceil \mathcal{CV}_K(\psi, \varphi) \rceil$. So, $\textsc{Estimate}\big[\mathcal{CV}_{@}(\varphi^*, \varphi)\big] = \sum_K Cr(K) \cdot \lceil \mathcal{CV}_K(\psi, \varphi) \rceil < 0$. So, you ought to prefer φ to φ^*. Because your preferences ought to be transitive, you ought to prefer φ to ψ. This shows that, in general, if $\textsc{Estimate}\,[\lfloor \mathcal{CV}_{@}(\varphi, \psi) \rfloor] > 0$, then you ought to prefer φ to ψ.

A Generalization of Causal Decision Theory. If your ends can be represented with a utility-function, then, for every dependency hypothesis K, $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil = \lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor = \mathcal{CV}_K(\varphi, \psi)$.

Therefore, $\textsc{Estimate}\,[\lfloor \mathcal{CV}_@(\varphi, \psi) \rfloor] > 0$ just in case $\textsc{Estimate}\,[\mathcal{CV}_@(\varphi, \psi)] > 0$. And, as we saw, $\textsc{Estimate}\,[\mathcal{CV}_@(\varphi, \psi)] > 0$ holds just in case the causal expected utility of φ exceeds the causal expected utility of ψ. Thus, Actual Value Decision Theory is a generalization of causal decision theory.

Parity in Vacation Boxes. According to Actual Value Decision Theory, you should regard the Larger box and the Regular box as on a par. Because A^+ is on a par with B and B^+ is on a par with A, the lower bound on your estimate of the comparison in actual values between the Larger box and the Regular box is less the zero, and the upper bound on your estimate is greater than zero.

$$\textsc{Estimate}\left[\lfloor \mathcal{CV}_@(L, R) \rfloor\right] < 0 < \textsc{Estimate}\left[\lceil \mathcal{CV}_@(L, R) \rceil\right]$$

$$\sum_K Cr(K) \cdot \lfloor \mathcal{CV}_K(L, R) \rfloor < 0 < \sum_K Cr(K) \cdot \lceil \mathcal{CV}_K(L, R) \rceil$$

Suppose that $\$y$ is the smallest amount that A^+ must be improved in order for it to be preferred to B; and suppose that $\$x$ is the smallest amount that B^+ needs to be improved to be preferred to A. It follows from this that $\$(y + 2)$ is the smallest amount needed by A to be preferred to B^+, and that $\$(x+2)$ is the smallest amount needed by B to be preferred to A^+. These facts place bounds on the extent to which these outcomes are on a par:

$$\begin{array}{lll}
\text{Because} & (A^+ + \$y) \succ B, & \lceil \mathcal{CV}_H(R, L) \rceil = V(\$y) \\
\text{Because} & (B^+ + \$x) \succ A, & \lceil \mathcal{CV}_T(R, L) \rceil = V(\$x) \\
\text{Because} & (A + \$(y + 2)) \succ B^+, & \lceil \mathcal{CV}_T(L, R) \rceil = V(\$(y + 2)) \\
\text{Because} & (B + \$(x + 2)) \succ A^+, & \lceil \mathcal{CV}_H(L, R) \rceil = V(\$(x + 2))
\end{array}$$

We can use these quantities to arrive at the lower and upper bounds of your estimates.

$$\sum_K Cr(K) \cdot \lfloor \mathcal{CV}_K(L,R) \rfloor = \frac{-(V(\$x) + V(\$y))}{2} < 0$$

$$\sum_K Cr(K) \cdot \lceil \mathcal{CV}_K(L,R) \rceil = \frac{V(\$(x+2)) + V(\$(y+2))}{2} > 0$$

Because $V(\$y)$ and $V(\$x)$ are both greater than zero, the lower bound on your estimate is less than zero and the upper bound on your estimate is greater than zero. According to Actual Value Decision Theory, then, you shouldn't prefer taking the Larger box over the Regular box; the two options are on a par.

Going Beyond Parity. Consider the following decision-problem. There are two boxes: the Larger box and the Regular box. There is some chance, p, that L contains $\$z$ while R contains $\$0$; otherwise, a coin was flipped to determine whether prize A was placed in L and prize B in R or *vice versa*.

	K_1	K_2	K_3
L	A	B	$\$z$
R	B	A	$\$0$

If the chance, p, and the prize money, $\$z$, are large enough, then, according to Actual Value Decision Theory, rationality requires you to take the Larger box. Suppose that $\$y$ is the smallest improvement to A to render it preferred to B, and that $\$x$ is the smallest improvement to B to render it preferred to A. These facts place bounds on the extent to which your outcomes are on a par.

$$\begin{aligned}
&\text{Because} \quad (A + \$y) \succ B, \quad \lceil \mathcal{CV}_{K_1}(R,L) \rceil = \lceil \mathcal{CV}_{K_2}(L,R) \rceil = V(\$y) \\
&\text{Because} \quad (B + \$x) \succ A, \quad \lceil \mathcal{CV}_{K_2}(R,L) \rceil = \lceil \mathcal{CV}_{K_1}(L,R) \rceil = V(\$x) \\
&\text{Because} \quad \$z \succ \$0, \qquad \left\lceil \mathcal{CV}_{K_3}(L,R) \right\rceil = V(\$z)
\end{aligned}$$

Your credence in K_3 is p, and your credences in K_1 and K_2 are both $\frac{1-p}{2}$. We can use these quantities to arrive at the lower bound on your estimate of the extent to

which L has more actual value than R.

$$\sum_K Cr(K) \cdot \lfloor \mathcal{CV}_K(L,R) \rfloor = \frac{1-p}{2} \cdot -V(\$y) + \frac{1-p}{2} \cdot -V(\$x) + p \cdot V(\$z)$$
$$= p \cdot V(\$z) - \frac{1-p}{2} \cdot \Big(V(\$x) + V(\$y) \Big)$$

According to Actual Value Decision Theory, if $p \cdot V(\$z) - \frac{1-p}{2} \cdot (V(\$x) + V(\$y)) > 0$, then you are rationally required to prefer L to R.

$$p \cdot V(\$z) - \frac{1-p}{2} \cdot \Big(V(\$x) + V(\$y) \Big) > 0$$
$$p \cdot V(\$z) > \frac{1-p}{2} \cdot \Big(V(\$x) + V(\$y) \Big)$$
$$\frac{p}{1-p} > \frac{V(\$x) + V(\$y)}{2 \cdot V(\$z)}$$
$$\frac{p}{1-p} > \frac{V(\$x) + V(\$y)}{V(\$z) + V(\$z)}$$

If p and $V(\$z)$ are large enough, then this inequality will hold. For example, suppose that $p = \frac{1}{2}$ and that you prefer $\$z$ to both A and B. The lower bound on your estimate of the extent to which the actual value of L exceeds the actual value of R will be greater than zero just in case $2 \cdot V(\$z) > V(\$x) + V(\$y)$. And, because you prefer $\$z$, by itself, to B, $V(\$z) \geq V(\$y)$; and, because you prefer $\$z$, by itself, to A, $V(\$z) \geq V(\$x)$. So, $2 \cdot V(\$z) > V(\$x) + V(\$y)$. Therefore, according to Actual Value Decision Theory, you should prefer option L to option R.

Although it's just as likely that your options are on a par as it is that the actual value of L exceeds the actual value of R, the extent to which the actual value of L might exceed the actual value of R is large enough to outweigh the elasticity of the parity between the outcomes in the other states.

$p = ?$	*verdict*
0	$L \bowtie R$
$\frac{1}{2}$	$L \succ R$
1	$L \succ R$

The interesting cases are when $0 < p < \frac{1}{2}$. For that range of credences, whether you are rationally required to take L over R depends on the extent to which your preference for $\$z$ over $\$0$ is greater than the extent to which you regard prizes A and B as on a par. And we needn't expect there to be a precise fact of the matter about this.

5.5 Relaxing the Unrealistic Assumption

As mentioned above, it's unrealistic to assume that $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$ and $\lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor$ are well-defined. There needn't be a *least* upper bound, nor a *greatest* lower bound, on the extent to which the value of outcome $(\varphi \wedge K)$ exceeds the value of outcome $(\psi \wedge K)$ if you regard these outcomes as on a par.

It's *not* unrealistic to assume, however, that you can place *some* (upper and lower) bounds on $\mathcal{CV}_K(\varphi, \psi)$. In order to relax the unrealistic assumption, then, replace $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$ in the formulation above with *some* upper bound, and replace $\lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor$ with *some* lower bound. If your estimate of the actual value comparisons between φ and ψ *using these bounds* is greater than zero, then rationality requires you to prefer φ to ψ. Why? If this estimate is greater then zero, then there's some lower bound on the estimate of the extent to which φ's actual value might exceed ψ's which is greater than zero. But if this lower bound is greater than zero, then, the *greatest* lower bound on your estimate of the extent to which φ's actual value might exceed ψ's — if it were to exist — would, also, be greater than zero.[36] In this manner, we can relax the unrealistic assumption that there are precise bounds on the extent to which outcomes are on par, while retaining sufficient conditions for when rationality requires you to prefer one option to another.

> *Sufficient Conditions for Preferring φ to ψ*
>
> If your lower bound estimate of the extent to which φ's actual value might exceed φ's actual value is greater than zero, then rationality requires you to

[36] Alternatively, because $\lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor = -\lceil \mathcal{CV}_K(\psi, \varphi) \rceil$, if your estimate of the extent to which the actual value of ψ might exceed the actual value of φ using some upper bounds is *less* than zero, then you should prefer φ to ψ. This is because, if there's an upper bound on your estimate of the extent to which ψ's actual value might exceed φ's that's less than zero, then the *least* upper bound — if it were to exist — would, also, be less than zero.

prefer φ to ψ.

We can work toward providing necessary conditions for preferring one option to another (as well as sufficient conditions for regarding two options as on a par) by looking at the upper and lower bounds on $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$ and $\lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor$ themselves. If you prefer outcome $(\psi \wedge K)$ sweetened by $\$u$ to outcome $(\varphi \wedge K)$, then $V(\$u)$ is an upper bound on $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$. And, if you regard $(\psi \wedge K)$ sweetened by $\$l$ as on par with $(\varphi \wedge K)$, then $V(\$l)$ is a lower bound on $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$. In other words, the least upper bound on the extent to which the value of $(\varphi \wedge K)$ exceeds the value of $(\psi \wedge K)$ — were it to exist — can be approached from above or below.

Necessary Conditions for Preferring φ to ψ

Rationality requires you to prefer φ to ψ only if your lower-bound-on-the-least-upper-bound estimate of the extent to which ψ's actual value might exceed φ's actual value is less than zero.

If the least upper bound on your estimate of the extent to which ψ's actual value might exceed φ's actual value — were it to exist — is less than zero, then your lower-bound-on-the-least-upper-bound estimate must also be less than zero. Because, according to Actual Value Decision Theory, you should prefer φ to ψ only when $\textsc{Estimate}\big[\lceil \mathcal{CV}_{@}(\psi, \varphi) \rceil\big] < 0$, it follows that you should prefer φ to ψ only if the lower-bound-on-the-least-upper-bound estimate is less than zero. In a similar fashion — by making use of the upper and lower bounds that you can place on $\lceil \mathcal{CV}_K(\varphi, \psi) \rceil$ and $\lfloor \mathcal{CV}_K(\varphi, \psi) \rfloor$ — we can provide a sufficient condition for regarding your options as on a par.

Sufficient Conditions for Parity Between φ and ψ

If (i) your upper-bound-on-the-greatest-lower-bound estimate of the extent to which φ's actual value might exceed ψ's actual value is less than zero, and (ii) your lower-bound-on-the-least-upper-bound estimate of the extent to which φ's actual value might exceed ψ's actual value is greater than zero, then you ought to regard φ and ψ as on a par.

We have, then, sufficient conditions for preferring one option to another and sufficient conditions for regarding the two as on a par. But there's no guarantee that these sufficient conditions will be met in all cases — it will depend on the (upper and lower, and upper-on-lower and lower-on-upper) bounds that you place on the value-comparisons between the outcomes of your options. If none of the sufficient conditions are met, then Actual Value Decision Theory is silent: it says nothing at all about what rationality requires or permits you to do.

What do I mean by *"silent"*? Again, think of a decision theory like a helpful advisor who, given information about your perspective and your aims, issues recommendations about what you rationally ought to do. The advisor might say, of some particular option, that you are rationally required to take it. Or, she might say that there are several options, each of which it is rationally permissible to take. But, also, she might remain silent: she might say nothing at all about what rationality requires, or permits, you to do. If your advisor isn't given enough information about your perspective and your aims, we shouldn't expect her to be able to help you. If you are unable to place informative enough (upper and lower, and upper-on-lower and lower-on-upper, etc.) bounds on the extent to which the outcomes of your options are on a par, Actual Value Decision Theory, like the helpful advisor, can't help you. There's simply no fact of the matter about what rationality requires of you. And that, I think, is exactly right.

References

F. Anscombe and R. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, pages 199–205, 1963.

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

R. Aumann. Utility theory without the completeness axiom. *Econometrica*, 30(3): 445–462, 1962.

A. Bales, D. Cohen, and Toby Handfield. Decision theory of agents with incomplete preferences. *Australasian Journal of Philosophy*, 92(3):453–470, 2014.

Rachel Briggs. Normative theories of rational choice: Expected utility, Aug 2014. URL http://plato.stanford.edu/entries/rationality-normative-utility/.

John Broome. Is incommensurability vagueness? In Ruth Chang, editor, *Incommensurability, Incomparability, and Practical Reason*. Harvard University Press, 1997.

Ruth Chang. The possibility of parity. *Ethics*, 112(4):659–688, 2002.

Ruth Chang. Parity, interval value, and choice. *Ethics*, 115(2):331–350, 2005.

David Christensen. Clever bookies and coherent beliefs. *Philosophical Review*, 100: 229–47, 1991.

J. Dubra, F. Maccheroni, and E. Ok. Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115(1):118–133, 2004.

Ozgur Evren and Efe Ok. On the multi-utility representation of preference relations. *Journal of Mathematical Economics*, 47:554–563, 2011.

H. Gaifman. A theory of higher order probabilities. In B. Skyrms and W. Harper, editors, *Causation, Chance, and Credence*. Dordrecht: Kluwer, 1988.

Tsogbadral Galaabaatar and Edi Karni. Subjective expected utility theory with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.

J. Gert. Value and parity. *Ethics*, 114(3):492–510, 2004.

Allan Gibbard and William Harper. Counterfactuals and two kinds of expected utility. In JJ. Leach C.A. Hooker and E.F. McClennen, editors, *Foundations and Applications of Decision Theory, Vol I*. Dordrecht, 1978.

I.J. Good. Rational decisions. *Journal of the Royal Statistical Society*, B(14):107–114, 1952.

Caspar Hare. Take the sugar. *Analysis*, 70(2):237–247, 2010.

H. Herzberger. Ordinal preference and rational choice. *Econometrica*, 41:187–237, 1973.

Richard Jeffrey. The sure thing principle. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2:719–720, 1982.

Richard Jeffrey. *The Logic of Decision*. University of Chicago Press, 1983.

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

James Joyce. Are newcomb problems really decisions? *Synthese*, 156:537–562, 2007.

James Joyce. A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, 24(1):281–323, 2010.

Nico Kolodny and John McFarlane. Ifs and oughts. *The Journal of Philosophy*, 108 (3):115–143, 2010.

Isaac Levi. Imprecision and indeterminacy in probability judgment. *Philosophy of Science*, 36:331–340, 1985.

Isaac Levi. *Hard Choices*. Cambridge University Press, Cambridge, 1986.

Isaac Levi. Value commitmnts, value conflict, and the separability of belief and value. *Philosophy of Science*, 66:509–533, December 1999.

Isaac Levi. Symposium on cognitive rationality: Part i minimal rationality. *Mind and Society*, 5:199–211, August 2006.

Isaac Levi. Why rational agents should not be liberal maximizers. *Canadian Journal of Philosophy*, 38(Supplementary Vol 34):1–17, 2008.

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

Robert Nau. The shape of incomplete preferences. *The Annals of Statistics*, pages 2430–2448, 2006.

Robert Nozick. Newcomb's problem and two principles of choice. In N. Rescher et al., editor, *Essays in Honor of Carl G. Hempel*. Reidel, 1969.

E. Ok, P. Ortoleva, and Gil Riella. Incomplete preferences under uncertainty: Indecisiveness in beliefs versus tastes. *Econometrica*, 80(4):1791–1808, 2012.

Richard Pettigrew. Risk, rationality and expected utility theory. *Canadian Journal of Philosophy*, 45(5-6):798–826, 2015.

Wlodek Rabinowicz. Incommensurability meets risk. unpublished, July 2016.

J. Raz. Value incommensurability: Some preliminaries. *Proceedings in the Aristotelian Society*, 86:117–134, 1985.

Michael D. Resnik. *Choices: an Introduction to Decision Theory*. University of Minnesota Press, 1987.

Susanna Rinard. A decision theory for imprecise credences. *Philosopher's Imprint*, 15(7):1–16, 2015.

Leonard J Savage. *The Foundations of Statistics*. Wiley, 1954.

George Schlesigner. The unpredictability of free choices. *The British Journal for the Philosophy of Science*, 25(3):209–221, Sep 1974.

George Schlesigner. Unpredictability: A reply to cargule and to benditt and ross. *The British Journal for the Philosophy of Science*, 27(3):267–274, Sep 1976.

Miriam Schoenfield. Decision making in the face of parity. *Philosophical Perspectives*, 28(1):263–277, 2014.

Teddy Seidenfeld, M.J. Schervish, and J.B. Kadane. Decisions without ordering. In W. Sieg, editor, *Acting and Reflecting: the interdisciplinary turn*, pages 143–170. Kluwer Academic Publishing, 1990.

Teddy Seidenfeld, M.J. Schervish, and J.B. Kadane. A representation of partially ordered preferences. *The Annals of Statistics*, 23(6):2168–2217, 1995.

Amartya Sen. Incompleteness and reasoned choice. *Synthese*, 140(1/2):43–59, May 2004.

Jordan Howard Sobel. Probability, chance and choice: A theory of rational agency. unpublished, May 1978.

Jordan Howard Sobel. Circumstances and dominance in a causal decision theory. *Synthese*, 63:167–202, 1985.

Jack Spencer and Ian Wells. An argument for two-boxing. manuscript, June 2016.

Robert Stalnaker. Letter to david lewis. In Robert Stalnaker William Harper and Glenn Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht, 1981.

W.J. Talbott. Two principles of bayesian epistemology. *Philosophical Studies*, 62: 135–50, 1991.

Bas C. van Fraassen. Belief and the will. *The Journal of Philosophy*, 81:236–56, 1984.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

Brian Weatherson. Decision making with imprecise probabilities. manuscript, November 2008.

Ralph Wedgwood. Gandalf's solution to the newcomb problem. *Synthese*, 190(14): 2643–75, 2013.

A Causal Decision Theory

In this section, I prove the following claim:

> *Your unconditional estimate of the extent to which the actual value of φ exceeds the actual value of ψ is greater than your unconditional estimate of the extent to which the actual value of ψ exceeds the actual value of φ* if and only if *the causal expected utility of φ is greater than the causal expected utility of ψ.*

Recall our notion of *actual value:*

$$V_@(\varphi) = V(\varphi \wedge K_@)$$

The proposition that $V_@(\varphi) = v$, then, is equivalent to the proposition that $V(\varphi \wedge K_@) = v$. In turn, *that* proposition is equivalent to the following disjunction of conjunctions:

$$\bigvee_{K_i}\Big(V(\varphi \wedge K_i) = v \wedge K_i\Big)$$

Similarly, the proposition that $\mathcal{CV}_@(\varphi, \psi) = v^*$ is equivalent to the following disjunction of conjunctions:

$$\bigvee_{K_i}\Big(\mathcal{CV}_{K_i}(\varphi, \psi) = v^* \wedge K_i\Big)$$

Because the dependency hypotheses, in virtue of being a partition, are mutually exclusive and mutually exhaustive, exactly one such K holds (which we've been calling $K_@$). Furthermore, because the dependency hypotheses are mutually exclusive, each of this disjunction's disjuncts are mutually exclusive. Consequently, your credence that $\mathcal{CV}_@(\varphi, \psi) = v^*$ can be expressed as a sum of your credences in each of the disjuncts.

$$Cr\Big(\mathcal{CV}_@(\varphi, \psi) = v^*\Big) = \sum_K Cr\Big(\mathcal{CV}_K(\varphi, \psi) = v^* \wedge K\Big)$$

Furthermore, assume (as we've implicitly been doing) that you are *self-aware:* you know how the values you assign to the various possible outcomes compare to one another. In other words, we take your credences in propositions of the form $\mathcal{CV}_K(\varphi, \psi) = v^*$ to be maximally opinionated and accurate:

$$Cr\Big(\mathcal{CV}_K(\varphi, \psi) = v^*\Big) = \begin{cases} 1 & \text{if } \mathcal{CV}_K(\varphi, \psi) = v^* \\ 0 & \text{otherwise} \end{cases}$$

Therefore, your credence that *both* $\mathcal{CV}_K(\varphi, \psi) = v^*$ *and* K should, likewise, be zero when $\mathcal{CV}_K(\varphi, \psi) \neq v^*$ but equal $Cr(K)$ when $\mathcal{CV}_K(\varphi, \psi) = v^*$.

$$Cr\Big(\mathcal{CV}_K(\varphi, \psi) = v^* \wedge K\Big) = \begin{cases} Cr(K) & \text{if } \mathcal{CV}_K(\varphi, \psi) = v^* \\ 0 & \text{otherwise} \end{cases}$$

Let $1[q]$ be an *indicator function* that returns 1 if q is true and 0 if q is false.

$$1[q] = \begin{cases} 1 & \text{if } q \\ 0 & \text{otherwise} \end{cases}$$

Using this indicator function, we can express your credences in propositions about the actual values of your options in terms of your credences in dependency hypotheses. In particular,

$$Cr\Big(\mathcal{CV}_@(\varphi, \psi) = v^*\Big) = \sum_K 1\big[\mathcal{CV}_K(\varphi, \psi) = v^*\big] \cdot Cr(K)$$

In other words, your credence that the extent to which the actual value of φ exceeds the actual value of ψ is v^* should be equal to the sum of your credences in the dependency hypotheses in which the difference in value between the outcome of φ and the outcome of ψ is v^*.

This allows us to rewrite Actual Value Estimate in terms of your credences in dependency hypotheses, as follows:

$$\begin{aligned}\textsc{Estimate}\Big[\mathcal{CV}_{@}(\varphi,\psi)\Big] &= \sum_{v^*} Cr\big(\mathcal{CV}_{@}(\varphi,\psi)=v^*\big)\cdot v^* \\ &= \sum_{v^*}\Bigg(\sum_{K}\mathbf{1}\big[\mathcal{CV}_K(\varphi,\psi)=v^*\big]\cdot Cr(K)\Bigg)\cdot v^*\end{aligned}$$

For each possible value v^*, the term $\sum_K \mathbf{1}\big[\mathcal{CV}_K(\varphi,\psi) = v^*\big]\cdot Cr(K)\cdot v^*$ equals $\sum_K Cr(K)\cdot\mathcal{CV}_K(\varphi,\psi)$ if $\mathcal{CV}_K(\varphi,\psi)=v^*$ and, otherwise, it equals zero. And, so,

$$\begin{aligned}\textsc{Estimate}\Big[\mathcal{CV}_{@}(\varphi,\psi)\Big] &= \sum_{v^*}\sum_{K}\mathbf{1}\big[\mathcal{CV}_K(\varphi,\psi)=v^*\big]\cdot Cr(K)\cdot\mathcal{CV}_K(\varphi,\psi) \\ &= \sum_{K} Cr(K)\cdot\mathcal{CV}_K(\varphi,\psi)\cdot\sum_{v^*}\mathbf{1}\big[\mathcal{CV}_K(\varphi,\psi)=v^*\big]\end{aligned}$$

Furthermore, because, for each dependency hypothesis K, there is *exactly one* possible value v^* such that $\mathcal{CV}_K(\varphi,\psi)=v^*$,

$$\sum_{v}\mathbf{1}\big[\mathcal{CV}_K(\varphi,\psi)=v^*\big]=1$$

Therefore, the (unconditional) estimate of the extent to which the actual value of φ exceeds the actual value of ψ is equal to the difference between φ's and ψ's causal expected utilities:[37]

$$\begin{aligned}\textsc{Estimate}\Big[\mathcal{CV}_{@}(\varphi,\psi)\Big] &= \sum_{v^*} Cr\big(\mathcal{CV}_{@}(\varphi,\psi)=v^*\big)\cdot v^* \\ &= \sum_{K} Cr(K)\cdot\mathcal{CV}_K(\varphi,\psi) \\ &= \sum_{K} Cr(K)\cdot\big(V(\varphi\wedge K)-V(\psi\wedge K)\big) \\ &= \sum_{K} Cr(K)\cdot V(\varphi\wedge K)-\sum_{K} Cr(K)\cdot V(\psi\wedge K)\end{aligned}$$

[37] Or, rather, this equivalence holds when the actual values of your options are well-defined, so that $\mathcal{CV}_K(\varphi,\psi)=V(\varphi\wedge K)-V(\psi\wedge K)$.

According to the Actual Value Conception, you should prefer option φ to option ψ if and only if $\textsc{Estimate}\left[\mathcal{CV}_{@}\left(\varphi, \psi\right)\right] > \textsc{Estimate}\left[\mathcal{CV}_{@}\left(\psi, \varphi\right)\right]$.

$$\textsc{Estimate}\left[\mathcal{CV}_{@}\left(\varphi, \psi\right)\right] > \textsc{Estimate}\left[\mathcal{CV}_{@}\left(\psi, \varphi\right)\right]$$

$$\sum_{K} Cr(K) \cdot V(\varphi \wedge K) - \sum_{K} Cr(K) \cdot V(\psi \wedge K) > \sum_{K} Cr(K) \cdot V(\psi \wedge K) - \sum_{K} Cr(K) \cdot V(\varphi \wedge K)$$

$$2 \cdot \sum_{K} Cr(K) \cdot V(\varphi \wedge K) > 2 \cdot \sum_{K} Cr(K) \cdot V(\psi \wedge K)$$

$$\sum_{K} Cr(K) \cdot V(\varphi \wedge K) > \sum_{K} Cr(K) \cdot V(\psi \wedge K)$$

$$U(\varphi) > U(\psi)$$

So, the Actual Value Conception entails causal decision theory: you should prefer φ to ψ if and only if the causal expected utility of φ is greater than the causal expected utility of ψ.

B Benchmark Decision Theory

Ralph Wedgwood [Wedgwood, 2013] defends a decision theory that, much like evidential decision theory, uses *conditional probabilities* but that, much unlike evidential decision theory, conforms to the Actual Value Conception by measuring the value of an option in a state *comparatively*.

Benchmark Decision Theory

You should prefer an option φ to an option ψ just in case

$$\sum_K Cr(K \mid \varphi) \cdot \big(V(\varphi \wedge K) - b_K\big) > \sum_K Cr(K \mid \psi) \cdot \big(V(\psi \wedge K) - b_K\big)$$

Where b_K is a "benchmark" value in state K. Wedgwood suggests that, when there are only two options under consideration, we can take b_K to be the *average* of the values of the outcomes of φ and ψ in K:

$$b_K = \frac{V(\varphi \wedge K) + V(\psi \wedge K)}{2}$$

Let's write $V_B(\varphi)$ to denote the "benchmark" expected value of φ. (That is: $V_B(\varphi) = \sum_K Cr(K \mid \varphi) \cdot \big(V(\varphi \wedge K) - b_K\big)$.) When you choosing only between two options, Wedgwood [2013] recommends identifying b_K, the benchmark value in state K, with the average of the values of the outcomes in K. However, when there are three or more options under consideration, a more complicated procedure is necessary to generate an appropriate "benchmark." Wedgwood [2013] offers a couple suggestions for how this procedure might go. The argument in this section, however, pertains only to what benchmark decision theory says in the simple two-option case.

According to Benchmark Decision Theory, you should prefer φ to ψ (when those are the only two options under consideration) just in case:

$$\sum_K Cr(K \mid \varphi) \cdot \Big(V(\varphi \wedge K) - avg\big(V_K(\varphi), V_K(\psi)\big)\Big) > \sum_K Cr(K \mid \psi) \cdot \Big(V(\psi \wedge K) - avg\big(V_K(\varphi), V_K(\psi)\big)\Big)$$

$$\sum_K Cr(K \mid \varphi) \cdot \left(\frac{V(\varphi \wedge K) - V(\psi \wedge K)}{2}\right) > \sum_K Cr(K \mid \psi) \cdot \left(\frac{V(\psi \wedge K) - V(\varphi \wedge K)}{2}\right)$$

$$\sum_K Cr(K \mid \varphi) \cdot \Big(V(\varphi \wedge K) - V(\psi \wedge K)\Big) > \sum_K Cr(K \mid \psi) \cdot \Big(V(\psi \wedge K) - V(\varphi \wedge K)\Big)$$

$$\sum_K Cr(K \mid \varphi) \cdot \mathcal{CV}_K(\varphi, \psi) > \sum_K Cr(K \mid \psi) \cdot \mathcal{CV}_K(\psi, \varphi)$$

Furthermore, $\sum_K Cr(K \mid \varphi) \cdot \mathcal{CV}_K(\varphi, \psi)$ is equivalent to your *conditional estimate* of the extent which the actual value of option φ exceeds the actual value of ψ.[38] And so, if your estimates are conditional estimates, the "benchmark value" of an option (at least when there are only two options under consideration) equals your estimate of the extent to which that option's actual value exceeds the actual value of the other option under consideration.

Wedgwood's benchmark decision theory conforms to the Actual Value Conception. It uses conditional, rather than unconditional, estimates. Causal decision theory, as we've seen, also conforms to the Actual Value Conception. Evidential decision theory, on the other hand, does not.

C The Principle of Actual Value

C.1 Causal Decision Theory entails the Principle of Actual Value

According to the unconditional version of the Actual Value Conception, you should prefer φ to ψ if and only if $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) > 0$. In order to show that the principle follows, we'll assume that $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$ and, then, show that $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$.

Assume that $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$.

$$Cr\big(V_@(\varphi) > V_@(\psi)\big) = \sum_K 1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] \cdot Cr(K)$$

So, if $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then $\sum_K 1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] \cdot Cr(K) = 0$.

And, $\sum_K 1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] \cdot Cr(K) = 0$ just in case, for each dependency hypothesis K, either:

("αenumi) $1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] = 0$, or

("αenumi) $Cr(K) = 0$, (or both). For each K, if $1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] = 0$, then $\mathcal{CV}_K(\varphi, \psi) \not> 0$.

[38] The proof is analogous to the one presented in the previous section.

And so, $Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$. Also, for each K, if $Cr(K) = 0$, then $Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) = 0$.

Therefore, for every dependency hypothesis K, $Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$. And thus, $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$.

Therefore, if $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then $\sum_K Cr(K) \cdot \mathcal{CV}_K(\varphi, \psi) \not> 0$. And, so, according to the Actual Value Conception, you should not prefer option φ to option ψ.

C.2 Benchmark Decision Theory entails the Principle of Actual Value

I will show that if $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then Wedgwood's Benchmark Decision Theory will say that you shouldn't strictly prefer φ-ing to ψ-ing.

Claim: If $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then $V_B(\varphi) \not> V_B(\psi)$

First, recall that if $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then, for all dependency hypotheses K,

$$1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] \cdot Cr(K) = 0$$

If $1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] \cdot Cr(K) = 0$, then either $1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] = 0$, or $Cr(K) = 0$, or both.

(1) If $1\big[V(\varphi \wedge K) > V(\psi \wedge K)\big] = 0$, then $V(\varphi \wedge K) - V(\psi \wedge K) \leq 0$, or

(2) If $Cr(K) = 0$, then $Cr(K \mid \varphi) = Cr(K \mid \psi) = 0$.

Second, if we, following Wedgwood, take b_K to be the average of the values of the outcomes of φ and ψ in K, then the benchmark expected value of an option ψ can be rewritten as follows:[39]

[39] The benchmark value in K needn't be the unweighted average of the values of the outcomes in K in order for the proof to go through. Any weighted average — just so long as the same weights are used in every K — will work just as well.

$$\begin{aligned} V_B(\varphi) &= \sum_K Cr(K \mid \varphi) \cdot \big(V(\varphi \wedge K) - b_K\big) \\ &= \sum_K Cr(K \mid \varphi) \cdot \left(V(\varphi \wedge K) - \frac{V(\varphi \wedge K) + V(\psi \wedge K)}{2}\right) \\ &= \sum_K Cr(K \mid \varphi) \cdot \left(\frac{V(\varphi \wedge K) - V(\psi \wedge K)}{2}\right) \end{aligned}$$

Finally, $V_B(\varphi) > V_B(\psi)$ just in case $V_B(\varphi) - V_B(\psi) > 0$.

$$\sum_K Cr(K \mid \varphi) \cdot \left(\frac{V(\varphi \wedge K) - V(\psi \wedge K)}{2}\right) - \sum_K Cr(K \mid \psi) \cdot \left(\frac{V(\psi \wedge K) - V(\varphi \wedge K)}{2}\right) > 0$$

$$\sum_K \left(\frac{V(\varphi \wedge K) - V(\psi \wedge K)}{2}\right) \cdot \Big(Cr(K \mid \varphi) + Cr(K \mid \psi)\Big) > 0$$

$$\sum_K \Big(V(\varphi \wedge K) - V(\psi \wedge K)\Big) \cdot \Big(Cr(K \mid \varphi) + Cr(K \mid \psi)\Big) > 0$$

As established above, if $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then, for all K, either, $V(\varphi \wedge K) - V(\psi \wedge K) \leq 0$, or $Cr(K \mid \varphi) = Cr(K \mid \psi) = 0$, or both. This means that, for all K,

$$\Big(V(\varphi \wedge K) - V(\psi \wedge K)\Big) \cdot \Big(Cr(K \mid \varphi) + Cr(K \mid \psi)\Big) \leq 0$$

Therefore,

$$\sum_K \Big(V(\varphi \wedge K) - V(\psi \wedge K)\Big) \cdot \Big(Cr(K \mid \varphi) + Cr(K \mid \psi)\Big) \leq 0$$

So, if $Cr\big(V_@(\varphi) > V_@(\psi)\big) = 0$, then $V_B(\varphi) \not> V_B(\psi)$.