Comparativism About Instrumental Value

Abstract: Nihilists about instrumental value deny that choices have any objective, instrumental values. They say that there's no fact-of-the-matter about what a choice did to help you achieve your ends. Absolutists about instrumental value say that a choice has an objective, instrumental value, and this value is independent of which alternatives you could have selected instead. Here, I introduce and explore a third option, which I call 'comparativism about instrumental value'. According to the comparativist, a choice is objectively instrumentally valuable to the extent that it leaves the world better than an alternative would. Because this varies, depending upon which alternative we consider, the only facts about objective instrumental value are comparative facts. I give two reasons to take comparativism seriously. In the first place, it better fits with natural ways of thinking about instrumental value than either nihilism or absolutism. In the second place, it affords us theories of instrumental rationality which avoid problems faced by evidential and causal decision theory.

1 | INTRODUCTION

You value some things for their own sakes. Others you value because of how they help you achieve the things you value for their own sakes. Say that the first things have *non-instrumental*, or *final*, value, and that the second things have *instrumental* value. A choice is instrumentally valuable to the extent that it helps you achieve final value; it is instrumentally disvaluable to the extent that it hinders your pursuit of final value.

Some say that there is no such thing as *objective* instrumental value. They say there's no fact-of-the-matter about whether a choice of yours was helpful or harmful, nor any objective measure of the extent to which that choice was beneficial or detrimental. They think anything worth calling 'instrumental value' is irreducibly subjective and perspective-dependent.[1] Call any view like this a form of *nihilism*

Word Count: 11,479
Draft of February 1, 2023; ✉: dmitri.gallow@gmail.com

1. For instance, Arif Ahmed defends this view in joint work with Jack Spencer (Ahmed & Spencer, 2020).

about instrumental value. Others think that choices have objective instrumental values, and—moreover—that each choice has exactly one objective instrumental value, a value which is independent of the alternative choices you could have made instead.[2] Call any view like this a form of *absolutism* about instrumental value.

In this paper, I will introduce and explore a third option, which I'll call 'comparativism about instrumental value'. Unlike the nihilist, the comparativist says that choices have objective instrumental values. Unlike the absolutist, the comparativist says that a choice has *many* instrumental values—one for each alternative choice you could have made instead. According to the comparativist, a choice is objectively instrumentally valuable to the extent that it leaves the world better than an alternative would. Because this varies, depending upon which alternative we consider, the only facts about objective instrumental value are comparative facts. We can say that your choice left the world better than A would have—so that, compared to A, it had positive instrumental value. And we can say that your choice left things worse than B would have—so that, compared to B, it had negative instrumental value. But we cannot say whether your choice left the world better or worse *full stop*. So we cannot say how instrumentally valuable your choice was *full stop*. We can only say how instrumentally valuable it was compared to each of the available alternatives.

I think there are two reasons for taking comparativism seriously. In the first place, natural ways of thinking about whether choices were helpful or harmful are better captured by comparativism than its extant rivals. In the second place, comparativism affords us novel theories of instrumental rationality which avoid some problems faced by evidential and causal decision theory—at least, many of us regard them as problems, though, of course, opinion varies.

In §2 below, I introduce comparativism's extant rivals, absolutism and nihilism, and discuss how these views relate to evidential and causal decision theory. In §3, I introduce comparativism and explain why it better aligns with natural ways of thinking about the degree to which a choice was objectively beneficial or detrimental than either nihilism or absolutism. In §4, I'll explore how we should think about instrumental *rationality*, if the comparativist is right about instrumental *value*. We'll see that comparativism affords us theories of rational choice that have nice properties and issue plausible judgements in decisions where many of us think that causal or

2. Most recently, Andrew Bacon, Boris Kment, Jack Spencer, and Ian Wells have taken this position (Spencer & Wells, 2019, Bacon, forthcoming, Kment, forthcoming). Their terminology differs from my own. What I call 'objective instrumental value', they call 'actual value', 'real value', and 'objective value'. I take us all to be talking about the same thing—though see §2 below for a caveat. See also Joyce, 1999 and Konek & Levinstein, 2019 for related discussion.

evidential decision theory gets things wrong.

2 | ABSOLUTISM AND NIHILISM

In §2.1, I'll say a bit more about final value. In §2.2, I'll say a bit more about instrumental value, nihilism, and absolutism, and I'll discuss the relationship between nihilism and absolutism, on the one hand, and causal and evidential decision theory, on the other.

2.1 | *Final Value*

If you, like me, think that some things are non-instrumentally valuable, independent of whether anyone non-instrumentally values them, then you will want to distinguish between a subjective and an objective understanding of final value. The non-instrumentally valuable things which no one values have final value in the objective sense, but lack it in the subjective sense. The things which are non-instrumentally valued in spite of their not being non-instrumentally valuable have final value in the subjective sense, but not in the objective sense. Here, I'm not going to bother distinguishing between these two understandings of final value. I'm just going to assume that you non-instrumentally value all and only the things which are non-instrumentally valuable, and that you value them to exactly the degree that they are valuable. So I'll allow myself to switch back and forth between subjective questions like how well satisfied your desires are, and objective questions like how good things are overall. Your moral perfection allows us to treat these two question as one.

I'll take it for granted that the degree to which you value things non-instrumentally can be measured on at least an interval scale. An interval scale is one for which the order of values and the relative distances between values are significant, but absolute values and distances are not. Nor is it significant whether values are positive or negative.[3] I will assume that the objects of final value are propositions. It will be enough for our purposes to assign final values to a mutually exclusive and jointly exhaustive collection of propositions which are detailed enough to specify every outcome you care about. I'll call one of those propositions a 'world', and I'll write the final value of a world, W, as '$V(W)$'.

3. That is: if V and V' are interval scales, α is any real number greater than 0, β any real number, and $V' = \alpha \cdot V + \beta$, then V and V' agree—they both contain the same information. If there is no $\alpha > 0$ and β such that $V' = \alpha \cdot V + \beta$, then V' and V disagree. See Suppes & Zinnes, 1963.

2.2 | *Instrumental Value*

If you're facing a decision, you'll have some collection of options available to you. I'll suppose that these options are various *acts* that you can choose to perform or not. An act's *objective instrumental value*, if it exists, is the degree to which the act helped or hindered you in your pursuit of final value—the degree to which the act was beneficial or detrimental. I don't mean the degree to which you *expected* it to help, or the degree to which you *believed* it would help. I mean the degree to which it objectively did help, when all is said and done.

Nihilists deny that there is any such thing as objective instrumental value. According to them, the only thing deserving the name 'instrumental value' is a subjective, perspective-dependent quantity. They distinguish between acts which were rational and those which were irrational, but they countenance no distinction between those which were beneficial and those which were detrimental. Absolutists, on the other hand, recognise objective instrumental value. They distinguish between acts which turn out to be effective means to your ends and those which don't. They think, moreover, that the instrumental value of an act is singular, and can be specified independently of which alternative acts you could have performed instead.

Causal decision theorists are sometimes understood as endorsing absolutism.[4] On this understanding, the causalist holds that the objective instrumental value of the act A, at a world W, is an objective quantity which I'll call 'utility' and notate '$U(A, W)$'. The utility of A is given by the final value of the world which would result, were you to perform A. That is: it is given by the final value of the world, W^*, such that the counterfactual 'were you to choose A, it would be that W^*' is true.[5]

4. For this understanding of causal decision theory, see in particular Joyce, 1999, especially §§5.4–5.5, Konek & Levinstein, 2019, Spencer & Wells, 2019, Ahmed & Spencer, 2020, Spencer, 2021, and Bacon, forthcoming. For more on causal decision theory, see Stalnaker, 1981, Gibbard & Harper, 1978, Lewis, 1981, Rabinowicz, 1982, Skyrms, 1982, 1990, Sobel, 1994, Rabinowicz, 2009, and Joyce, 2012, 2018, among many others.

5. In the body, I'll take for granted that there is a unique world which would result, were you to perform A. But we might think that, in some situations, there are multiple worlds which might result, were you to choose A. (That is: we might reject Stalnaker's *uniqueness assumption*. See Lewis, 1973, and Stalnaker, 1980.) If so, then our definition of utility will have to be modified. In general, for any world W and any act A, let W^A be a probability distribution over worlds, with the interpretation that the probability W^A gives to the world W^* measures the causal tendency that A has, at W, to bring about the world W^*. (See Lewis, 1981, Sobel, 1994, and Joyce, 1999.) Then, the absolutist causalist holds that the objective, instrumental value of an act, A, at a world W, is the expected final value which would result, were you to choose A at W, where this expectation is taken with respect to the probability function W^A. Assuming the number of worlds is countable, we have that $U(A, W) = \sum_{W^*} V(W^*) \cdot W^A(W^*)$. Throughout, I will assume that the number of worlds is countable for the sake of expository convenience. If there is an uncountable number of worlds, the notion of expectation

Counterfactuals are notoriously context-sensitive. When we evaluate a counterfactual, we hold some things fixed and allow others to swing free, and English affords us a great deal of leeway in which to hold fixed. We can get true readings of both 'if you'd gone to law school, you would have hated it' and 'if you'd gone to law school, you'd have been interested in law, so you'd have loved it'. So, in our definition of 'utility', it matters how we evaluate the counterfactual 'were you to choose A, it would be that W^{*}'. Causalists will want to evaluate it by holding fixed factors which are causally upstream of your choice, and only allowing to swing free your choice and the factors which are causally downstream of your choice (that's what makes their decision theory 'causal').

Given this definition of utility, it's natural to understand the causal decision theorist as saying that the objective, instrumental value of an act is given by its utility.

Causal Absolutism The objective instrumental value of an act is its utility.

Causal decision theory says that, if you know the utility of every option, then you should choose one of the options with the greatest utility. If you are unsure what each option's utility is, then you should choose an act which has the greatest possible *expected* utility, where the expectation is taken with respect to your own subjective probabilities, or credences. So causal decision theory follows straightfowardly from causal absolutism, together with the thesis I'll call 'expectationism',[6]

Expectationism If options have objective, absolute instrumental values, then it is rational to choose A iff your expectation of the instrumental value of A is no less than your expectation of the instrumental value of every other option.

Your expectation of a quantity gives your best estimate of that quantity's value. So expectationism says: if options have objective instrumental values, and if your best estimate of A's instrumental value is no less than your best estimate of every other options' instrumental value, then it is rational to choose A. As a matter of notation, I will remove the variable 'W' to indicate that an expectation is being taken. Thus, '$U(A)$' will stand for the *expected* utility of the act A.[7] Then, causal decision theory says that it is rational to choose an option, A, iff, for every alternative option B, you expect the utility of A to be no lower than the utility of B, $U(A) \geqslant U(B)$. So it is

---

will have to be generalised in the usual ways.

6. Cf. Ahmed & Spencer, 2020.

7. That is, $U(A) = \sum_W U(A, W) \cdot C(W)$, where C is your credence function.

very natural to understand causal decision theorists as being committed to causal absolutism and expectationism—though of course you could be a causal decision theorist while rejecting both of these theses.

Jack Spencer and arch-evidential decision theorist Arif Ahmed teach that evidential decision theorists should deny that there is any such thing as objective instrumental value (see Ahmed & Spencer, 2020). Evidential decision theorists maintain that it is rational to choose an option iff that option gives you the best possible news about final value. In general, you won't know what the final value of the actual world is, but your credences will allow you to estimate the world's final value by taking an expectation of the world's final value. An option gives you good news about final value to the extent that learning that you have chosen that option raises your expectation of the world's V-value. So if your expectation of the world's V-value after conditioning your credences on the proposition that you've chosen A is no lower than your expectation of the world's V-value after conditioning your credences on the proposition that you've chosen B, for any alternative B, then it is rational to choose A. As a matter of notation, I will remove the variable 'W' from '$V(W)$' and subscript with 'A' to indicate that an expectation is being taken, conditional on A. (I'll sometimes use 'A' for an act, and sometimes for the proposition that you've performed that act; context will disambiguate.) Then, 'V_A' is your expectation of final value, conditional on you choosing A,[8] and the evidential decision theorist says that it is rational to choose A iff, for every alternative option B, A gives news about final value which is no worse than the news given by B, $V_A \geqslant V_B$.

Ahmed & Spencer present an argument that evidentialists should be nihilists (though they do not ultimately endorse it; they have other reasons for thinking evidentialists should be nihilists). This argument appeals to the thesis I called 'expectationism' above. If we accept both evidential decision theory and expectationism, then we must reject the claim that options have objective instrumental values. The issue is that there is no *objective* quantity Q such that A gives better news about final value than B does, $V_A > V_B$, iff A has more of Q than B does, in expectation, $Q(A) > Q(B)$. Of course, there *is* a quantity, Q, such that $V_A > V_B$ iff $Q(A) > Q(B)$. This is the quantity $C(A \mid W) \cdot V(W)/C(A)$, where '$C$' is your subjective probability, or credence, function.[9] But this quantity is not objective. It makes ineliminable reference to your own credences. As you get less confident that you'll choose A,

8. That is, $V_A = \sum_W V(W) \cdot C(W \mid A)$, where C is your credence function.

9. $V_A = \sum_W V(W) \cdot C(W \mid A) = \sum_W V(W) \cdot [C(A \mid W)/C(A) \cdot C(W)] = \sum_W [C(A \mid W) \cdot V(W)/C(A)] \cdot C(W)$, which is your expectation of $[C(A \mid W) \cdot V(W)/C(A)]$.

the quantity gets larger, and as you get more confident that you'll choose A, it gets smaller. As your credence that you'll choose A goes to zero, it diverges to infinity. Even if in *some* cases an act's objective instrumental value depends upon how likely you think you are to select it, this shouldn't be so in *every* case. This quantity is too subjective and perspective-dependent to deserve the name '*objective* instrumental value'. So the argument concludes: if you accept evidential decision theory, and if you accept expectationism, then you must deny that options have objective instrumental values. You must be a nihilist.[10]

Evidentialists could resist this argument by rejecting expectationism. They could insist that, when making a choice, you should only concern yourself with the instrumental value an act would have *if it were performed*. There are a variety of ways they could motivate this position, but let me mention one that I find particularly compelling: they could say that unperformed acts do not have any instrumental value at all. I don't mean: the instrumental value of unperformed acts is *zero*. I mean: there is no such thing as the objective instrumental value of an act which was not performed. Two roads diverge in a wood, and you take the one less travelled by. We can ask about what difference this made—to what extent it left the world better or worse. But it is quite unnatural to ask about what difference your taking the road *more* travelled by made. It doesn't seem that taking the road more travelled by made any difference at all. It didn't leave things better, or worse, or just as good. It didn't do anything. It didn't happen. Of course, we can ask about what difference your taking the road more travelled by *would* have made, had you taken it. But that's instrumental value the act *would* have had, and not instrumental value it *in fact* has.

So I think it's pre-theoretically plausible that unperformed acts neither help nor harm. Insofar as we want our theory of instrumental value to align with pre-theoretic ways of thinking about which choices are helpful or harmful, we have reason to say

10. A result from Snow Zhang (cited in Bacon, forthcoming) teaches us something stronger: if we accept expectationism and we think that objective instrumental value is always defined and always credence-independent, in the sense that changing your credences does not change the objective instrumental value of any option at any world, then we must think that instrumental value is something with the same form as the causalist's utility. That is: we must think that the objective instrumental value of A at W takes the form of an expectation of final value, where the expectation is taken with respect to some probability function W^A, which is a function of the world W and the act A. (See footnote 5.) That is, given expectationism, the objective instrumental value of A at W must take the form $\sum_{W^*} V(W^*) \cdot W^A(W^*)$, for *some* probability distribution W^A, which can vary as a function of the world W and the act A. Of course, we could interpret the function W^A in a variety of ways, and there's nothing in Zhang's result requiring even that $W^A(A) = 1$. Note that this result assumes that acts have objective instrumental values even when they are not performed. So it poses no threat to the thesis I call 'evidential absolutism' below.

that unperformed acts have no degree of instrumental value. (For this reason, the version of comparativism I will develop in §3 below will say an act can have instrumental values only if it is performed.) But this raises problems for expectationism. For the expectation of a quantity is only well-defined so long as the quantity itself is certain to have some value or other. If the quantity may not have any value, you cannot take an expectation of it.[11] However, you *can* take a *conditional* expectation of the quantity. Conditional on what? On the quantity having a value. If we think that acts only have instrumental value when they are performed, this means that we can take an expectation of the act's instrumental value, conditional on the act being performed. And it's natural to think that rational choice is guided by your best estimate of an option's instrumental value, conditional on it having any instrumental value at all. So it's natural to accept

Conditional Expectationism If options have objective, absolute instrumental values, then it is rational to choose A iff your expectation of the instrumental value of A—conditional on A having an instrumental value—is no less than your expectation of the objective instrumental value of every alternative option—conditional on them having instrumental values.

There is no conflict between evidential decision theory and conditional expectationism. If we trade expectationism in for conditional expectationism, then the evidentialist can accept precisely the same theory of instrumental value as the causalist, and identify instrumental value with utility—*modulo* their assumption that unperformed acts have no instrumental value. That is, the evidentialist can accept:

Evidential Absolutism The objective instrumental value of a performed act is its utility. Unperformed acts have no instrumental value.

If you choose A, then the final value which *would* result, *were* you to choose A, is just

11. I mean to say that, if $Q(A, W)$ is only defined for W which entail A and if $C(A) < 1$, then (1) $\sum_W Q(A, W) \cdot C(W)$ will not be defined. You might suggest defining the expectation as (2) $\sum_q q \cdot C(Q(A) = q)$, where '$Q(A) = q$' is a proposition which is true at any world W such that $Q(A, W) = q$, and false at all other worlds. If $Q(A, W)$ is certain to be defined, then (1) and (2) will always be equal. But if $Q(A, W)$ is only defined for W which entail A, then (2) will be well-defined even when (1) is not. Even though (2) is well-defined, it is not well-behaved, and it is not plausibly the kind of thing that should guide action. If $Q(A, W)$ is only defined when W entails A, then you can get the value of (2) arbitrarily close to zero by simply setting your credence that you'll choose A low enough. Consider a choice between a guaranteed \$1 and a guaranteed \$100. If you're confident enough that you'll take the \$1, then the expected instrumental value of taking the \$1 will be higher than the expected instrumental value of taking the \$100—if those expectations are calculated with (2). So it doesn't look like you should be guided by the quantity (2).

the final value of the actual world.[12] So the utility of a performed act is just equal to the actual world's final value.[13] So taking an expectation of the utility of A, conditional on you choosing A, is equivalent to taking an expectation of final value, conditional on you choosing A. In the notation I introduced earlier, where removing the 'W' variable indicates taking an expectation, and subscripting with 'A' indicates that the expectation is conditional on A, we have that $U_A(A) = V_A$.[14] So we can understand evidentialists as accepting almost the same theory of objective instrumental value as causalists; they simply disagree about whether unperformed acts have any instrumental value.

Ahmed & Spencer give another argument that evidentialists must be nihilists. And Andrew Bacon (forthcoming) gives a third argument for this conclusion. But each of these arguments rely on the assumption that unperformed acts have instrumental value. If evidentialists reject this assumption, then they can resist each of these other arguments. Evidentialists could, of course, accept nihilism, just as causalists could accept nihilism. But they could instead accept the form of evidential absolutism I've introduced here, paired with conditional expectationism. The debate between evidentialists and causalists is independent of the debate between absolutists and nihilists.

2.3 | Why I'm Not a Nihilist

Nihilism is a stark departure from the way we ordinarily think. You face a decision: you can either purchase travel insurance or take the risk. You choose to purchase the insurance but tragedy doesn't strike, so you make no claim. Suppose that whether tragedy strikes is causally counterfactually independent of whether you purchase insurance. As we ordinarily think about things, even though it was rational for you to purchase the insurance, there was something objectively suboptimal about the choice. As I'd want to put it: from an objective point of view, purchasing the insurance did less to advance your ends than taking the risk would have. But if purchasing the insurance doesn't have an objective instrumental value, then it's difficult to make sense of this natural thought.

Some nihilists might try to accommodate these natural thoughts by considering

12. This follows from the principle called 'conjunction conditionalisation', according to which 'A and C' entails 'If A were to be the case, C would be the case'. For a defence of this principle, see Walters & Williams, 2013.

13. That is, if W entails A, then $U(A, W) = V(W)$.

14. That is, $\sum_W U(A, W) \cdot C(W \mid A) = \sum_W V(W) \cdot C(W \mid A)$. See Gibbard & Harper, 1978, §3.

decisions made in conditions of certainty. In conditions of certainty, you know for sure exactly which world would result, were you to choose each of your options. Then, nihilists could suggest that what I've misidentified as 'objective instrumental value'—the degree to which an act objectively helps or hinders you in the pursuit of your ends—isn't an objective notion at all. It is instead simply the subjective notion of an act being rational to perform in conditions of certainty. The only sense in which the insurance wasn't helpful, from an objective point of view, is this: were you to choose whether to buy insurance from an objective point of view—that is, were you to face that decision in conditions of certainty—it would be irrational to buy the insurance.[15] This subjective surrogate captures many natural thoughts about instrumental value, but not all. Suppose the gods smite humans who know too much, but leave the ignorant be. If you were to be making your decision in conditions of certainty, one thing you would know for sure is that the gods are readying their smite, so it would be rational for you to purchase the insurance. But this doesn't make the insurance any more objectively helpful in the actual world, where your ignorance means that they the gods pose no threat.[16]

There are other, more extravagant ways of understanding instrumental value in subjective terms. For instance, we could imagine an idealised version of you—call them 'you+'—who shares your values but lacks no conceptual or computational resources, and who spends their days meditating on a possibility exactly like yours in every detail. The nihilist could say that the insurance did less to help you achieve your ends than the risk would have in at least this sense: you+ would want you to take the risk.[17] An approach like this does allow us to characterise facts about instrumental value in subjective terms. In fact, an approach like this allows us to characterise *anything* in subjective terms. In general, for any claim '*p*', we can imagine an idealised version of you who spends their days meditating on a possibility exactly like yours, and most wants to correctly answer the question 'would '*p*' express a truth in the possibility I'm imagining?'. Then, '*p*' is true iff it would be rational for you+ to answer this question with 'yes'. When it comes to characterisations like this, I just find it overwhelmingly natural to think that it is rational for you+ to answer this question with 'yes' *because* '*p*' is true, and overwhelmingly unnatural to think that '*p*' is

15. Something like this proposal appears in Parfit, unpublished, where an act is said to be "*objectively* right if it would be what I ought to do if I knew all of the morally relevant facts".

16. Cf. Shope, 1978.

17. Cf. Railton, 1986 and Smith, 1994—though I don't take these authors to be discussing *instrumental* value in particular.

true *because* it is rational for you+ to answer this question with 'yes'. And I likewise find it overwhelmingly natural to think that you+ would want you to take the risk *because* taking the risk helps you achieve your ends, and overwhelmingly unnatural to think that taking the risk helps you achieve your ends *because* you+ would want you to take the risk.

There's another kind of nihilist who doesn't even attempt to accommodate natural thoughts about the insurance being objectively unhelpful. This nihilist is sceptical about the idea of causal influence. They agree with Russell (1912) that causation is "a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm".[18] This kind of nihilist may be happy to grant both the counterfactuals 'had you not bought the insurance, tragedy still wouldn't have struck' and 'had you not bought the insurance, you'd have been a more reckless person, and so you'd likely have needed it.' But they'll deny that there's any objective basis on which we can say that the former is a *causal* counterfactual. The distinction between factors which are causally downstream of your choice and those which are not is a fiction.

This nihilist may be happy to say, in some contexts, that purchasing the insurance hurt you—because, after all, had you taken the risk, tragedy still wouldn't have struck. But they will also be happy to say, in other contexts, that purchasing the insurance helped you—because, after all, had you taken the risk, you'd have been more risky, and so more likely to need the insurance. Without an objective distinction between causal and non-causal counterfactuals, the nihilist recognises no objective basis for favouring the first claim. So they deny that there's any objective, context-invariant sense in which the insurance was suboptimal in spite of being subjectively rational.

I disagree with this nihilist because I don't share their scepticism about causation. I think it's an objective matter whether some factor is causally downstream of your choice or not. I've tried to say a bit about how to understand the relation of causal influence elsewhere,[19] though I won't have the space to go into it here.

Others will see things differently, but I would rather retain the natural way of thinking that allows me to recognise an objective sense in which rational choices can nonetheless fail to provide effective means to your ends. So I would rather not be a nihilist.

18. See Ahmed, 2014, who lists scepticism about causation as one of his reasons for preferring evidential decision theory to causal decision theory.

19. See Gallow, 2016, forthcoming.

2.4 | Why I'm Not a Utility Absolutist

Causal and evidential absolutists both agree that the instrumental value of a performed act is its utility, which is equal to the final value of the world at which the act is performed. Call this shared commitment of theirs 'utility absolutism'.

Utility absolutism has a hard time accommodating natural ways of thinking about instrumental value. Paying for groceries at the market, you decide between the left-hand and the right-hand check-out aisle. There are no other options available; you must either exit through the left aisle or exit through the right aisle. Both aisles are open and equidistant, and we can stipulate that nothing of any final value depends upon which aisle you choose. You choose left. How much instrumental value did this choice have? It's very natural to think that it had *zero* instrumental value. Both alternatives would have left the world with just as much final value, so choosing to go left, rather than right, didn't make anything better or worse. It didn't do anything to help or hinder you in your pursuit of final value. You may disagree because you think that your choice at least had the instrumental value of getting you home with groceries. On my favoured view, this is instrumental value that your choice could have, compared to alternatives like 'do nothing' or 'leave without groceries'. In this example, I mean to artificially stipulate such alternatives away—your *only* options are paying through the left-hand aisle and paying through the right-hand aisle. But I'll have more to say about this suggestion below.

For the utility absolutist, the objective instrumental value of your choosing the left-hand aisle is given by the final value of the world at which you make that choice. But, while it seems that your choice accomplished nothing, and so had no instrumental value, it does not appear that the world lacks any final value. It is filled with seas of joy and suffering, triumph and defeat, love and loss. There are many respects in which your desires are satisfied, and many more in which they are frustrated. According to utility absolutism, all of this final value must go into the accounting of the instrumental value of your choice at the check-out aisle. And it appears very unlikely that all this final value balances out. If you are inclined to think that your choice had an instrumental value equal to the final value of the groceries, a similar objection applies: it appears very unlikely that the final value of the world overall balances out, in net, to equal the final value of the groceries.

Utility absolutists can attempt to accommodate the natural thought that choosing the left-hand aisle had zero instrumental value by reminding us of our assumption that final value is measured on an interval scale, so that it is arbitrary where we place the zero. Since the utility of a chosen act is just equal to the final value of the world at which it's chosen, utility is *also* measured on an interval scale, and so

it is also arbitrary where we place the zero on our scale of instrumental value. Perhaps, when we are considering a decision between two options, the context makes it natural to set the zero point equal to the level of final value the world would have had, had you chosen the alternative. Then, if your choice left things better than they would have been, had you chosen the alternative, we will say that your choice has positive instrumental value. If your choice made things worse than the alternative would have, we will say that your choice has negative instrumental value. And if your choice made them no better or worse than the alternative would have, we will say that your choice had zero instrumental value. If the alternative is going home without groceries, we will say that your choice had an instrumental value equal to the final value of the groceries.

In using an interval scale, we needn't deny that there's a difference between good and bad outcomes, any more than we deny that there's a difference between cold and hot temperatures when using Celsius or Fahrenheit. We just don't encode that information numerically in the scale we are using. We could, however, decide to reserve positive numbers for good outcomes and negative numbers for bad ones. If we did so, then we'd be representing your final values on a *ratio* scale. A ratio scale is just like an interval scale, except that the value zero is significant. For instance, dollars and euros are both ratio scales of wealth. The choice of what counts as one unit (a dollar or a euro) is arbitrary, but the choice of zero is non-arbitrary. A wealth of zero means that same thing in every currency.[20]

It's natural to think that there is a distinction between good and bad, so that final value can be measured on a ratio scale, even if the additional information provided by a ratio scale is not needed for a theory of rational choice. It is also natural to think that there is a distinction between acts which make things better and acts which make things worse, so that *instrumental* value can also be measured on a ratio scale—where the value zero corresponds to acts which make things no better and no worse. If we accept all of this, then it's natural to think that your choice at the checkout aisle had an instrumental value of zero *measured on a ratio scale*, even while the world overall has a non-zero degree of final value *measured on a ratio scale*. That is to say: it's natural to think that you choosing left over right made things no better

20. In general, if you take one ratio scale, and you multiply all of its values by the same positive number, what you get will be another ratio scale that contains all the same information you started out with. But any other way of changing the scale will give you different and incompatible information. That is, if V and V' are ratio scales, α is any positive real number and $V' = \alpha \cdot V$, then V and V' agree—they contain the same information. If V and V' are two ratio scales and there is no $\alpha > 0$ such that $V' = \alpha \cdot V$, then V' and V disagree. See Suppes & Zinnes, 1963.

or worse, even though the world overall is not neutral. But this outright contradicts utility absolutism, since according to utility absolutism, the instrumental value of that choice *just is* the final value of the world overall. If we measure both on ratio scales, then your act can have zero instrumental value iff the world overall has zero final value. If you are inclined to think that your choice had an instrumental value equal to the final value of the groceries, a similar objection applies: it seems that your choice could be helpful, and so have a positive degree of instrumental value *measured on a ratio scale*, while the world overall is bad, and so has a negative degree of final value, *measured on a ratio scale*.

Utility absolutism also makes it difficult to make non-trivial comparisons between the instrumental values of different acts chosen at the same world. You donate millions to effective charities, whereas I steal $20 from the girl-scouts. According to utility absolutism, both of our choices have precisely the same instrumental value: the final value of the actual world. So it doesn't look like utility absolutism will allow us to say that your choice was any more instrumentally valuable than mine. I'd rather not give up the natural way of thinking that allows me to say that some actually performed acts accomplish more good, and so are more instrumentally valuable, than others. So I'd rather not be a utility absolutist.

As an aside: in explaining why I'd rather not be a utility absolutist, I have been taking for granted a straightforward connection between instrumental value and pre-theoretic notions like how beneficial or detrimental a choice is. So I have understood utility absolutism as implying claims about how how much good a choice accomplishes. However, I suspect that some of the authors who have endorsed utility absolutism have instead been thinking of (what they call) 'objective value', 'real value', or 'actual value' as a theoretical posit. On this way of thinking, we begin by supposing that instrumental rationality is aimed at some objective property of our choices. We then introduce a term like 'actual value' to name this objective property. The meaning of this term is implicitly defined by the role it plays in our theories of instrumental rationality.[21] To mark this distinction, let's use 'actual value' for the theoretical posit, and reserve 'instrumental value' for the pre-theoretic notion of how helpful or harmful a choice was. We shouldn't fault a view about actual value for failing to align with pre-theoretic ideas about which choices were helpful or harmful. That is not its goal. Nonetheless, so long as more helpful choices can have lower actual values than less helpful choices, there are potential issues to beware of. For

21. See Lewis, 1970.

instance, in §4, I'll show that, if we accept the comparativist theory of instrumental value I'll develop in §3, then there are decisions in which you know for sure that, if you choose A, this will have an instrumental value of x; and you know for sure that, if you choose B, this will have an instrumental value of y, with $y < x$. Because causal and evidential absolutists aim at utility and not helpfulness,[22] they tell you to choose B over A. And this strikes me as a problem for causal and evidential absolutists, even if they don't say anything directly about which choices are helpful or harmful. Even if your theory of rational choice doesn't talk about the good which choices accomplish, it still shouldn't advise you to make a choice which you know for sure will accomplish less good, if chosen, than the alternative will accomplish, if you choose the alternative instead.

In general, there should be a link between the rationality of a choice and the good that choice accomplishes. Roughly, our theory of rational choice should lead us to make more beneficial choices, on average. If instrumental value were utility, then causal and evidential decision theory would both give us a link like this. For, if instrumental value were utility, then both causal and evidential absolutists would guide you towards helpful choices—either in expectation or in conditional expectation (conditional on the choice being made). But if the utility of a choice has nothing to do with the good that choice accomplishes, then it's not clear what link causal and evidential decision theory give us between a choice's rationality and the good it accomplishes.

Getting back on track: I'm inclined to reject utility absolutism. It conflates value the world has independent of your choice with the good your choice accomplishes. And it doesn't allow us to distinguish the instrumental values of actually performed acts. Of course, utility absolutism isn't the only form of absolutism. There may be another, more plausible, brand of absolutism out there. I'd be interested to see others explore the absolutist terrain further. But, having rejected nihilism and found no version of absolutism that satisfies me, I'll turn in the next section to considering an alternative to both nihilism and absolutism.

3 | COMPARATIVISM

If we think that choices have objective, instrumental values, it's natural to specify that value by considering counterfactual relationships between your choice and the

22. Causal absolutists aim at utility by maximising it in expectation; whereas evidential absolutists aim at utility by maximising it in conditional expectation (conditional on your choice).

world's final value. There are at least two ways we could try to do this. In the first place, we could consider the final value which your choice *counterfactually implies*. (A choice *counterfactually implies* ϕ just in case the counterfactual 'if you were to make that choice, then it would be that ϕ' is true.) In the second place, we could consider the final value which *counterfactually depends* upon your choice. The utility absolutist takes the first approach. In this section, I'll develop a view about instrumental value built around the second approach.

On this view, the instrumental value of taking the left-hand aisle is given by the difference between the final value the world has overall and the final value it *would* have had, had you chosen to take the right-hand aisle instead. Equivalently: the objective instrumental value of taking the left-hand aisle is the difference between the utility of taking the left-hand aisle and the utility of taking the right-hand aisle. More generally, in a decision between two options, we can say that the instrumental value of your choice is given by the difference between the utility of the choice you made and the utility of the alternative.

This works well in decisions between two options, but it will need to be generalised to decisions between more than two options. Suppose that, in addition to the left- and right-hand aisles, there is a centre aisle. If you go through the centre aisle, 50 billion tonnes of carbon dioxide will be emitted into the atmosphere. There are two natural thoughts about this new decision. Firstly: choosing the left-hand aisle *still* didn't make anything any better or worse, so choosing the left-hand aisle *still* had an instrumental value of zero. Secondly: choosing the left-hand aisle made things much better, since it kept 50 billion tonnes of carbon from being released into the atmosphere, so choosing the left-hand aisle had positive instrumental value. We can capture the first thought by comparing the choice of the left-hand aisle to the alternative of the right-hand aisle. Then, since the difference between the utility of the left-hand aisle and the utility of the right-hand aisle is zero, choosing the left-hand aisle will have an instrumental value of zero. And we can capture the second thought by comparing the choice of the left-hand aisle to the alternative of the centre aisle. Since the difference between the utility of the left-hand aisle and the utility of the centre aisle is positive, choosing the left-hand aisle will have a positive instrumental value.

You might think that there should be a *single* quantity which is the instrumental value of your choice. To that end, you might suggest that there's some privileged alternative option that you should be using to calculate instrumental value. Perhaps it is the choice you would have made, had you not made your actual choice? This makes the objective instrumental value of your choice depend upon your disposi-

tions in interesting ways. You and Barb both take the left-hand aisle. Had the left aisle been closed, you would have taken the right. But Barb thinks climate change is a Chinese hoax, and she's indifferent to releasing 50 billion tonnes of carbon. So, had the left aisle been closed, Barb would have taken the centre aisle. Then, even though you and Barb faced exactly the same decision in exactly the same circumstances and made exactly the same choice, Barb's choice was astronomically more instrumentally valuable than yours. If we think that you should be credited for helpful choices, this means that Barb is due an enormous amount of credit for her choice, whereas you are due no credit at all.

Perhaps the privileged alternative is the 'do nothing' option? Suppose that you and Buridan face the same decision: each of you can either choose the left-hand aisle, the right-hand aisle, or do nothing. Both aisles are equi-distant with the same wait time. What happens if either of you do nothing depends upon a prediction I made yesterday. If I predicted that you'd go left, then I set things up so that 50 billion tonnes of carbon would be released into the atmosphere if you did nothing. If I predicted that you'd either go right or do nothing, then I set things up so that *100* billion tonnes of carbon would be released if you did nothing. You're both sure that my predictions are accurate. Suppose that you go left, Buridan goes right, and I correctly predicted both of these choices. So no carbon is released, and you both get home at the same time with the same groceries. On the proposed view, there's no sense in which we get to say that your choice and Buridan's choice accomplished the very same thing and so had the very same instrumental value. When we turn to instrumental rationality, this view leads naturally to the conclusion that your choice was irrational. For, on this view, you are sure that, if you go left, your choice will have an instrumental value worth 50 billion fewer tonnes of carbon. And you are sure that, if you go right, your choice will have an instrumental value worth 100 billion fewer tonnes of carbon. Insofar as instrumental rationality is a matter of trying to perform acts with high instrumental value, it looks as though going left should count as irrational. This seems wrong to me; it feels much more natural to say that, while doing nothing is irrational, taking the left aisle is just as rational as taking the right aisle.

The view I prefer says that instrumental value is both comparative and plural. For each available alternative, there is some degree to which your choice made things better or worse than that alternative would have. This is the instrumental value of your choice, compared to that alternative. So your choice has as many instrumental values as there are alternatives available. If some particular alternative is made salient, we might prefer to talk about the instrumental value of your choice com-

pared to this salient alternative. So, if we are considering different alternative actions for you and Barb, then it might seem more natural to say that your choice accomplished nothing and had no objective instrumental value, whereas Barb's choice did something and had positive instrumental value. But we can bring out the plurality of instrumental values with explicit 'rather than' clauses. For instance, we can say both 'choosing the left aisle, rather than the centre, kept the world from getting much worse, and so had positive instrumental value' and 'choosing the left aisle, rather than the right, didn't make any difference, and so had zero instrumental value.' In my view, both of these claims are true. And they are made true by two different quantities of instrumental value that your choice has: one it has in comparison to the alternative of taking the centre aisle, and another it has in comparison to the the alternative of taking the right aisle.[23]

This is the thesis I'll call 'comparativism':

Comparativism Only acts which are performed have instrumental values. The instrumental value of a performed act, A, compared to an alternative, B, is the difference between the world's actual final value and the final value the world would have had, had you performed B instead.

The comparativist agrees with the evidential absolutist that unperformed acts do nothing to make anything better or worse, and so do not have any degree of instrumental value. (Recall the discussion from page 7.) The comparativist says: if you in fact choose A, then the objective instrumental value of A, compared to B, is the difference between the utility of A and the utility of B. I'll write the instrumental value of A, compared to B, at the world W, as '$I(A, B, W)$'. Then, the comparativist says that $I(A, B, W) = U(A, W) - U(B, W)$—so long, that is, as you choose A at W; if you don't choose A at W, then $I(A, B, W)$ is undefined.

There are two reasons I think comparativism is worth taking seriously. The first is that it seems to do a better job of capturing the way we're naturally inclined to think about instrumental value than either nihilism or utility absolutism. Comparativism allows us to recognise that, while buying insurance was the rational choice, it did less to advance your ends, and so had less instrumental value, than not buying the insurance would have. According to comparativism, we can measure final value and instrumental value on different ratio scales. So we can distinguish both between good and bad outcomes and between choices which make things better and worse—and we don't have to treat these distinctions as one-and-the-same. And comparativism

23. Cf. Schaffer, 2005*a*,*b*.

allows us to distinguish between the instrumental values of different choices made at the same world.

The second reason to take comparativism seriously is that it motivates theories of instrumental rationality which disagree with causal and evidential decision theory in decisions where those theories give verdicts many of us find counter-intuitive. In the next section, I'll give a sample comparativist theory of instrumental rationality, and argue that this theory gives more plausible verdicts than either evidential or causal decision theory.

4 | COMPARATIVISM AND INSTRUMENTAL RATIONALITY

You could accept comparativism about instrumental *value* and go on to accept a theory of instrumental *rationality* which has nothing at all to do with instrumental value—nothing at all to do with pursuing effective means to your ends. I find this way of thinking about things rather unnatural. Suppose you're facing a decision in conditions of certainty, so that you know for sure precisely which world would result, were you to choose each of the available options. Then, there will be at least one option which you know will have non-negative instrumental value compared to every alternative, if you choose it. An option like this is objectively most helpful. I find it overwhelming natural to think that, in circumstances like these, the objectively most helpful option is rational. So, at least in decisions made in conditions of certainty, I'm inclined to say that instrumental rationality aims at instrumental value. And I'm inclined to think that introducing uncertainty doesn't make it rational to pursue something other than instrumental value. It just complicates the pursuit.[24] So I'll take it for granted that, if we are comparativists, then we think that instrumental rationality is the rational pursuit of comparative instrumental value.

In conditions of certainty, rational choice is easy. You should choose an option which you know will leave things no worse than any alternative would. But when you don't know what the world is like, you won't know how much better or worse each of your choices would leave things than each of the alternatives. So you'll have to be guided by your best estimates of the quantities $I(A, B, W)$, for each pair of options A and B. You can't take an unconditional expectation of these quantities, since they may not be well-defined—if you don't choose A, then A won't leave things any

24. Being uncertain about which options are objectively most helpful may make it rational to settle for an option which you know to *not* be objectively most helpful, because each of the options which might be objectively most helpful are too risky. Cf. Regan, 1980, Parfit, unpublished, Jackson, 1991, and Muñoz & Spencer, 2021, among many others.

better or worse than B would have, so A won't have any instrumental value, compared to B. But you can take a *conditional* expectation of these quantities. Conditional on what? On the quantities being well-defined. As mentioned above, I'll remove the variable 'W' from a function to indicate that I'm taking an expectation of that function's value (with respect to your credences). And I'll additionally subscript a function with 'A' to indicate that this expectation is conditional on A. Then, $I_A(A, B)$ is the degree to which you will expect A to leave the world better than B would, if you choose A. Note that this will be equal to the difference between the expected utility of A, conditional on you choosing A, and the expected utility of B, conditional on you choosing A: $I_A(A, B) = U_A(A) - U_A(B)$.[25] Comparativists should be guided by these quantities.

There are a great many ways that you could be guided by the expected comparative instrumental values $I_A(A, B)$. So comparativists aren't forced to accept any particular decision theory. They could, for instance, accept orthodox causal decision theory. Consider the weighted sum $\sum_B I_B(B, A) \cdot C(B)$ (where C is your credence function). An option will minimise this weighted sum iff it maximises expected utility.[26] $I_B(B, A)$ is the degree to which you'll think B leaves the world better than A would have, if you choose B. So this weighted sum gives your expectation of how much better you think your choice will leave things than A would. So causal decision theorists could be understood as endorsing comparativism, and enjoining you to choose an option in comparison to which you expect your actual choice to have the least instrumental value.

So, in this section, I won't be arguing that any particular theory of instrumental rationality *follows from* comparativism. Instead, I want to emphasise that comparativism *affords us* theories of instrumental rationality which deliver plausible verdicts in decisions where many of us think causal and evidential decision theory get things wrong. These theories will be fairly natural for comparativists, but appear unmotivated and *ad hoc* for utility absolutists. For utility absolutism can make the choice between evidential and causal decision theory seem forced. If the instrumental value of an act is its utility, then it is difficult to object to the idea that instrumental rationality is a matter of maximising expected utility. There's just the question of whether

25. This follows because the expectation of $X - Y$ is equal to the expectation of X minus the expectation of Y (whether the expectation is conditional or not).

26. To appreciate this, note that $f(A) \equiv \sum_B I_B(B, A) \cdot C(B) = \sum_B [U_B(B) - U_B(A)] \cdot C(B)] = \sum_B U_B(B) \cdot C(B) - \sum_B U_B(A) \cdot C(B)$. And $\sum_B U_B(A) \cdot C(B) = U(A)$, by the law of iterated expectations. So $f(A)$ is equal to $(\sum_B U_B(B) \cdot C(B)) - U(A)$. Since $\sum_B U_B(B) \cdot C(B)$ is a constant, independent of our choice of A, this function is minimised exactly when $U(A)$ is maximised.

you should maximise expected utility unconditionally or instead maximise expected utility conditional on the act being performed. The former choice leads to causal decision theory; the latter, to evidential decision theory. In contrast, for comparativists, evidential decision theory is not remotely plausible as a theory of rational choice—its verdicts have nothing at all to do with acts' instrumental values. And while comparativists are free to endorse causal decision theory, that theory is not as inevitable for comparativists as it is for causal absolutists. After all, comparativists don't think that the instrumental *dis*-value of A is the degree to which your actual choice leaves things better than A would. If they thought this, then they'd think that the instrumental dis-value of every actually performed act was the same, namely zero. So it's not clear why instrumental rationality should consist in pursuing options that minimise your expectation of this quantity. If we give up on the idea that instrumental value is utility, causal decision theory seems far less natural, and far less inevitable.[27]

Consider the following two decisions:

Easy Money You can either take a guaranteed dollar or leave it.

Newcomb's Problem You can either take a guaranteed dollar or leave it. Incidentally, yesterday I took a scan of your brain and made a prediction about how you would choose. If I predicted that you'd leave the dollar behind, then I deposited $1000 into your bank account. If I predicted that you'd take the dollar, I did nothing. My predictions are never wrong, so you should be sure that you will take the dollar iff I predicted you would take the dollar.[28]

In both of these decisions, taking the dollar will leave you one dollar richer than leaving it behind would. And leaving the dollar behind will leave you one dollar poorer than taking it would. So the expected comparative instrumental values of both options are the same in the two decisions. Let me introduce a way of visually displaying expected comparative instrumental values. For each option, A, and each alternative $B \neq A$, we can consider how much better you'll expect A to leave things than B would, if you choose A, $I_A(A, B)$. If this is positive, I'll draw an arrow from B to A. If it is negative, then I'll drawn an arrow from A to B. I'll colour-code the arrows so that we can keep track of which arrows are associated with each option.

27. In addition to the theory I'll introduce below, see Barnett, 2022, Gallow, 2020*b*, and Podgorski, 2022, each of whom could be interpreted as developing comparativist theories of instrumental rationality.

28. See Nozick, 1969.

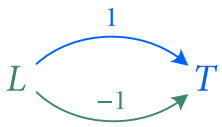

FIGURE 1: Expected comparative instrumental values for Easy Money and Newcomb's Problem. (L is the option to *l*eave the dollar behind, and T is the option to *t*ake it.)

And next to each arrow, I'll also write the associated value, $I_A(A, B)$. (That way, even without the colour-coding, you'll know which arrow is associated with each option: if the value is positive, then the arrow is associated with the option at the arrow's tip; whereas, if it is negative, then the arrow is associated with the option at the arrow's tail.) Then, the expected comparative instrumental values for both Easy Money and Newcomb's Problem are shown in figure 1.[29] Figure 1 tells us: if you choose L, your choice will leave things \$1 worse than T would. And, if you choose T, your choice will leave things \$1 better than L would.

Comparativists cannot distinguish between Easy Money and Newcomb's Problem. But evidential decision theory does. It tells you to take the dollar in Easy Money, but leave it behind in Newcomb's Problem. (The reason is that leaving the dollar behind gives you the news that there's likely \$1000 in your bank account, raising your expectation of world's final value.) So the evidentialist cannot be understood as advising you to pursue instrumental value as the comparativist understands it.

In Newcomb's Problem, there is some inclination to leave the dollar behind, and some inclination to take it. On reflection, many of us have decided that the inclination to leave the dollar behind is mistaken. On our view, what makes the decision confusing is that, while you have control over what to rationally *believe* about whether there's \$1000 in your bank account, you have no control over *whether* there's \$1000 in your bank account. Given my predictive powers, leaving the dollar behind gives you fantastic evidence that there is \$1000 in your account, and taking it gives you fantastic evidence that there's not \$1000 in your account. Since it's under your control whether you take the dollar or not, it's under your control what evidence you'll get. So it's under your control what to rationally believe about whether the \$1000 has been deposited. This can make it feel as though it is under your control *whether* the \$1000 has been deposited. But this is an illusion. By stipulation, at this point you have no control over whether the \$1000 is there or not. On reflection,

29. I've assumed that you value dollars linearly, and chosen a scale for final value on which the value of each additional dollar is one unit. I'll do this in any decision where the relevant final goods are specified in terms of dollars.

many of us have decided that the inclination to leave the dollar behind is due to this illusion. We have decided that there is no important difference between Newcomb's Problem and Easy Money. So we should welcome comparativism's inability to distinguish the two decisions.

Causalists agree that you should take the dollar in Newcomb's Problem. But, if they are causal absolutists, they do so for an unusual reason. In this decision, there are four relevant possible worlds: the world where you take the dollar and I predicted you'd take the dollar, W_{TT}, the world where you leave the dollar and I predicted you'd leave it, W_{LL}, the world where you take the dollar and I predicted you'd leave it, W_{TL}, and the world where you leave the dollar and I predicted you'd take it, W_{LT}. The utility of your chosen act, in each of these possibilities, is shown below.

World:	W_{TT}	W_{LL}	W_{TL}	W_{LT}
Utility of chosen act:	1	1000	1001	0

Because you know for sure that my prediction is accurate, you know for sure that you are in one of the first two columns—you're either in the world W_{TT} or you are in the world W_{LL}. So you know both of the following indicative conditionals for sure: (a) if you take the dollar, this choice will have a utility worth \$1; and (b) if you leave the dollar behind, this choice will have a utility worth \$1000. If utility is instrumental value, this means that you know for sure that the *instrumental value* of taking the dollar will be \$1, if you take the dollar, and you know for sure that the *instrumental value* of leaving the dollar behind will be \$1000, if you leave the dollar behind. In spite of this, causal absolutists recommend taking the dollar. The reason is that their decision theory doesn't only attend to the epistemically possible instrumental value of a choice. It *also* attends to instrumental value that you know for sure your choice won't have. So it is moved by the facts that (a) in W_{TT}, if you *were* to leave the dollar behind, you'd be at W_{LT}, and your choice *would* have a utility worth \$0; and (b) in W_{LL}, if you *were* to take the dollar, you'd be at W_{TL}, and your choice *would* have a utility worth \$1001.

Horgan (1981) objects to causal decision theory on the grounds that you are "virtually certain that our actual world will turn out to be either $[W_{TT}]$ or $[W_{LL}]$, rather than either $[W_{LT}]$ or $[W_{TL}]$. The worlds $[W_{LT}]$ and $[W_{TL}]$ therefore ought to be regarded as essentially irrelevant, for purposes of practical decision making".[30] I think we should agree with Horgan this far: instrumental value you know your choice won't have isn't worth pursuing. So *if* instrumental value is utility, then you should

30. Horgan, 1981, p. 338–9. Similar objections show up in Seidenfeld, 1984.

leave the dollar behind. But we should reject the antecedent, not accept the consequent. Comparativists should agree that the *instrumental* value of your choice at the worlds W_{TL} and W_{LT} is irrelevant to rational choice without conceding that the *final* value of those worlds is irrelevant to rational choice. For, even though you know for sure that these worlds aren't actual, the instrumental value of your *actual* choice depends upon what things are like at these worlds. At W_{TT}, your choice leaves things \$1 better than they would have been, had you left the dollar behind—and to see that, you have to consider what things are like at W_{LT}. Likewise, at W_{LL}, your choice leaves things \$1 worse than they would have been, had you taken the dollar—and to see that, you have to consider what things are like at W_{TL}.

World:	W_{TT}	W_{LL}
Instrumental value of chosen act:	1	-1

In my view, it is an advantage of comparativism that it allows us to say that taking the dollar is instrumentally rational *without* saying that instrumental rationality requires you to pursue instrumental value you know for sure is not in the offing.

In both Easy Money and Newcomb's Problem, comparativists should say that taking the dollar is more rational than leaving it behind. This suggests the following natural constraint on a comparativist theory of rational choice: if $I_A(A, B) > 0 > I_B(B, A)$, then A is a more rational choice than B is. That is: if you'll expect A to leave the world better than B would, if you choose A, and you'll expect B to leave the world worse than A would, if you choose B, then A is a more rational choice than B is. Unfortunately, this constraint is a bit too strong. Consider the following decision.

Cycle You must decide between three boxes, labelled '*P*', '*Q*', and '*R*'. Yesterday, I made a prediction about which box you would choose. If I predicted that you'd choose *P*, I left *P* empty, I put \$1 in *Q*, and I put an invoice for \$1 in *R*. If I predicted that you'd choose *Q*, then I left *Q* empty, I put \$1 in *R*, and I put an invoice for \$1 in *P*. And if I predicted that you'd choose *R*, then I left *R* empty, I put \$1 in *P*, and I left an invoice for \$1 in *Q*. My predictions are never wrong, so you should be certain that I correctly predicted your choice.

Expected comparative instrumental values for this decision are shown in figure 2. Figure 2 tells us that, if you choose *P*, you'll expect your choice to leave the world \$1 better than *R* would but \$1 worse than *Q* would. If you choose *Q*, you'll expect your choice to leave things \$1 better than *P* would but \$1 worse than *R* would. And if you choose *R*, you'll expect your choice to make things \$1 better than *Q* would, but \$1 worse than *P* would.

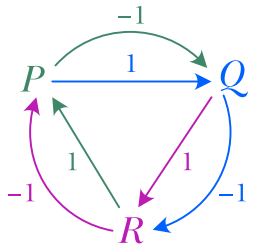

FIGURE 2: Expected comparative instrumental values for Cycle

The comparative instrumental values for any pair of options in Cycle are exactly the same as the comparative instrumental values for your options in Easy Money. So the natural constraint that A is more rational than B whenever $I_A(A, B) > 0 > I_B(B, A)$ tells us that, in Cycle, P is a more rational choice than Q, Q is a more rational choice than R, and R is a more rational choice than P. By the transitivity of 'is a more rational choice than', we have that P is a more rational choice than itself, which contradicts the irreflexivity of 'is a more rational choice than'. Some have wanted to deny that relations like 'is a more rational choice than' are transitive,[31] but I do not. So I think that the natural constraint is a bit too strong.

Comparing Easy Money and Cycle highlights the difficulty of defining a binary relation over options which is reflexive, transitive, total, and reasonable well-behaved.[32] In Easy Money, we wish to say that taking the dollar is strictly more rational than leaving it behind. But we do not wish to say this about any pair of options in Cycle. By the symmetry of the decision, we wish to say that any option is as rational as any other. So it looks like we can't say whether A is more, less, or just as rational as B by just considering the quantities $I_A(A, B)$ and $I_B(B, A)$ in isolation.

So I'll take a more indirect route. In general, if we're given a total but non-transitive relation '$\trianglerighteq$', we can *extend* this to a total and transitive ordering, $\geq$, as follows. If there is a sequence of options such that each option in the sequence bears $\trianglerighteq$ to its successor, say that there's a '$\trianglerighteq$-chain' leading from the option at the sequence's start to the option at the sequence's end. Now, for any options A and B, if there's a $\trianglerighteq$-chain leading from A to B, then say that $A \geq B$. So long as $\trianglerighteq$ was total, $\geq$ will be a total and transitive ordering.[33] (Since $\geq$ is total, it is also reflexive, since totality

31. See, for instance, Schwartz, 1972, Temkin, 1996, Rachels, 1998, and Barnett, 2022, among others.

32. For an argument that 'is at least as rational as' is total, see Dorr *et al.*, forthcoming.

33. To be explicit: we define $\geq$ to be the transitive closure of $\trianglerighteq$, $\trianglerighteq^+$. If either $A \trianglerighteq B$ or $B \trianglerighteq A$, then it must be that either $A \trianglerighteq^+ B$ or $B \trianglerighteq^+ A$. So it must be that either $A \geq B$ or $B \geq A$. So the totality of $\trianglerighteq$ implies the totality of $\geq$. Suppose that $A \geq B$ and $B \geq C$. Then, $A \trianglerighteq^+ B$ and $B \trianglerighteq^+ C$. So $A \trianglerighteq^+ C$. So

implies, for every A, $A \succeq A \vee A \succeq A$, which is equivalent to $A \succeq A$.)

Now, let us say that $A \trianglerighteq B$ iff $I_A(A, B) \geqslant I_B(B, A)$—that is, iff the degree to which you'll expect A to leave the world better than B would, if you choose A, is no less than the degree to which you'll expect B to leave the world better than A would, if you choose B. This relation will be total. As Cycle demonstrates, it can fail to be transitive. But the procedure from the previous paragraph allows us to extend $\trianglerighteq$ to a transitive and total relation, $\succeq$. Then, the comparativist theory of rational choice I'll consider says that A is as rational a choice as B iff $A \succeq B$. So, when there are only finitely many options, it says that A is rational iff $A \succeq B$, for every alternative B. Given a reflexive, transitive, and total ordering $\succeq$ ('is as rational as'), we can define an irreflexive and transitive ordering $\succ$ ('is more rational than') by stipulating that $A \succ B$ iff $A \succeq B$ and $\neg B \succeq A$. And we can define a reflexive, symmetric, and transitive ordering $\approx$ ('is just as rational as') by stipulating that $A \approx B$ iff $A \succeq B$ and $B \succeq A$.

(To be clear: my goal here is not to advance any particular theory of rational choice, but rather to showcase the benefits of thinking about instrumental rationality in terms of the pursuit of comparative instrumental value. There are many unexplored choice-points for a comparativist theory of rational choice. But this theory suffices to show that, once you start thinking about rational choice in terms of the pursuit of comparative instrumental value, it becomes easy to capture a collection of judgements and principles that neither causal nor evidential decision theory can.)

In Easy Money, we have that $T \trianglerighteq L$ and $\neg L \trianglerighteq T$. So there is a $\trianglerighteq$-chain leading from T to L and no $\trianglerighteq$-chain leading from L back to T. So the theory tells us that $T \succ L$. In contrast, in Cycle, we have that $P \trianglerighteq Q \trianglerighteq R \trianglerighteq P$. So there is a $\trianglerighteq$-chain leading from any option to any other. So we'll have that $P \approx Q \approx R \approx P$, and the theory will say that any choice is a rational one.

Consider the following decision.

Frustrator You may either take the left box, L, the right box, R, or the envelope, E. In the envelope is a guaranteed \$49. Yesterday, I took a scan of your brain and made a prediction about how you'd choose. If I predicted that you'd take the left box, then I put nothing in the left box, and put \$100 in the right box. If I predicted that you'd take the right box, then I put nothing in the right box, and put \$100 in the left box. If I predicted that you'd take the envelope, then I put \$50 in both boxes. My predictions are never wrong, so you should be certain

---

$A \succeq C$. So $\succeq$ is transitive.

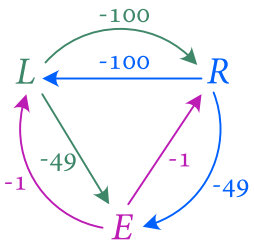

FIGURE 3: Expected comparative instrumental values for Frustrator

that I correctly predicted your choice.[34]

Expected comparative instrumental values for this decision are shown in figure 3. Figure 3 says: if you choose L, you'll expect your choice to get you \$49 less than E would, and \$100 less than R would. Symmetrically, if you choose R, you'll expect your choice to get you \$49 less than E would, and \$100 less than L would. Finally, if you choose E, you'll expect your choice to get you \$1 less than either L or R would.

In Frustrator, causal decision theory says that taking the envelope is irrational. No matter how likely you think you are to choose any of the options, taking the envelope will have a lower expected utility than one of the other options. And, if you think you're just as likely to choose the left box as the right, then the expected utility of either box will be \$50, exceeding the guaranteed \$49 in the envelope.[35]

Some are willing to call you irrational for taking the envelope,[36] but many of us continue to find this deeply counter-intuitive, even on reflection. If you take the left box, you will know that it is empty, that you could get \$49 by taking the envelope instead, and you could get \$100 by taking the right box instead. So, if you take the left box, you'll know that your choice leaves things \$100 worse than taking the right box would leave them, and \$49 worse than taking the envelope would leave them. Likewise, if you take the right box, you will know that your choice leaves things \$100 worse than taking the left box would leave them, and \$49 worse than taking the envelope would leave them. Finally, if you take the envelope, you'll know that there's

34. See Ahmed, 2014 and Spencer & Wells, 2019. Related decisions are discussed in Richter, 1984, Weirich, 1985, Harper, 1986, Egan, 2007, Arntzenius, 2008, Greaves, 2013, Joyce, 2012, Briggs, 2010, and Hare & Hedden, 2016, among many others.

35. If λ is your credence that you'll take the left box and ρ is your credence that you'll take the right box, then $U(L) = 50(1 + \rho - \lambda)$, $U(R) = 50(1 + \lambda - \rho)$, and $U(E) = 49$. If $\lambda \geqslant \rho$, then $1 + \lambda - \rho \geqslant 1$, so $U(R) \geqslant 50 > U(E)$. If $\rho \geqslant \lambda$, then $1 + \rho - \lambda \geqslant 1$, so $U(L) \geqslant 50 > U(E)$. Either way, one of the boxes has a greater expected utility than the envelope. If $\lambda = \rho$, then $U(L) = U(R) = 50 > U(E)$.

36. See, for instance, the valiant defence of this claim given by Joyce (2018). I explain why I'm unpersuaded by the defence in Gallow, 2020*a*.

$50 in the boxes, so that your choice leaves things $1 worse than the alternatives would.

It's not that I think throwing away a dollar is instrumentally valuable—it's not. Throwing away the extra dollar in the boxes is instrumentally dis-valuable. It makes things worse. But in this decision, you know in advance that your choice is going to make things worse than every other alternative would. I don't mean: you know that each of your options would make things worse. That's false. No matter what the world is like, one of the options would get you the most money possible. That option is the one that's objectively most helpful. You don't know what that option is, but you do know for sure that you won't choose it. For each of your options, you know that, if you choose it, it has negative instrumental value compared to every alternative. And, given this, I have a hard time seeing why the known instrumental dis-value of taking the envelope should count against it.

In Frustrator, $E \trianglerighteq L$, $E \trianglerighteq R$, $L \trianglerighteq R$, and $R \trianglerighteq L$. So there is a $\trianglerighteq$-chain leading from E to L, but no $\trianglerighteq$-chain leading from L back to E. So our theory will tell us that $E \succ L \approx R$, and that taking the envelope is the only rational choice.

Notice that, if either box is removed from the menu of options, causal decision theory will say that it is rational to take the envelope. In a binary decision between the left box and the envelope, the envelope will have a higher expected utility than the left box whenever your credence that you'll take the left box is greater than 2%.[37] So, in a binary decision between the left box and the envelope, causal decision theory says that it is rational to take the envelope. Likewise, in a binary decision between the right box and the envelope, the envelope will have a higher expected utility than the right box so long as your credence that you'll take the right box is greater than 2%. So, again, in a binary decision between the right box and the envelope, causal decision theory says that the envelope is rational.

Here are two plausible principles about rational choice:

Downwards Preservation Suppose that A is more rational than B. Then, if every option other than A and B were removed from the menu of options, A would be more rational than B.

Upwards Preservation Suppose that, for every option $B \neq A$, A would be more rational than B in a binary decision between the two. Then, choosing from the full menu of options, A is more rational than any other option.

37. If λ is your credence that you'll take the left box, then $U(L) = 50 \cdot (1 - \lambda)$ and $U(E) = 49$. So long as $\lambda > 1/50$, $U(E) > U(L)$.

These principles relate rational choice in a decision between three or more options to rational choice in its binary sub-decisions.[38] (I'm picturing the binary subdecisions as sitting below the decision on the full menu, whence the 'upwards' and 'downwards' names.) The first principle says that a 'more rational than' relation is preserved when you descend to the binary subdecision. (It then follows by contraposition that an 'as rational as' relation is preserved when you ascend from the binary subdecision to the decision on the full menu.) You might initially want a principle stronger than this. You might think that the 'as rational as' relation should also be preserved as you descend to the binary subdecision. But Cycle shows us that it is not. P is as rational as Q on the full menu, but P is not as rational as Q in a binary decision between the two. The second principle says that, if an option bears a 'more rational than' relation to the alternative in *every* binary subdecision in which it features, then all of those relations are preserved when you ascend to the full menu of options.

Causal decision theory violates both of these principles in Frustrator. It says that the left box is more rational than the envelope, but, if the right box were taken away, the left box would no longer be more rational than the envelope. So it violates Downwards Preservation. And, if you think you're just as likely to choose any of the options, then causal decision theory will say that the envelope would be more rational than either box in a binary subdecision between the two. But it says that the envelope is strictly less rational than either box on the full menu. So it violates Upwards Preservation. In contrast, the comparativist theory I introduced above satisfies this principle.[39] (To reiterate: my goal here is not to defend any particular theory of rational choice. What this theory highlights for us is just how easy it is to vindicate these plausible principles once we give up the idea that instrumental value is utility.)

Utility absolutism seems to force a choice between evidential and causal decision theory. But both of those theories seem to have problems. In Newcomb's Problem,

38. What I'm here calling 'the binary sub-decision between A and B' is a decision just like the original except that every option other than A and B has been removed from menu, and your credences have been conditioned on the disjunction $A \vee B$.

39. To show that the theory satisfies the principles, we need only appeal to the fact that $\trianglerighteq$ is a total relation and that whether one option bears it to another does not depend upon which other options are on the menu. (Notation: '$A \succ B \mid_{\mathbf{M}}$' says that A is more rational than B in the decision between the options in the menu $\mathbf{M}$.)

Start with Downwards Preservation. Suppose that $A \succ B \mid_{\mathbf{M}}$. Then, $\neg B \trianglerighteq^{+} A$. So $\neg B \trianglerighteq A$. By the totality of $\trianglerighteq$, $A \trianglerighteq B$. So $A \trianglerighteq B$ and $\neg B \trianglerighteq A$. So $A \succ B \mid_{\{A,B\}}$.

Next consider Upwards Preservation. Suppose that for every B, $A \succ B \mid_{\{A,B\}}$. Then, for every B, $A \trianglerighteq B$ and $\neg B \trianglerighteq A$. Then, on the full menu of options, there is a $\trianglerighteq$-chain leading from A to every other option, and no $\trianglerighteq$-chain leading from any other option back to A. So, for every option $B \neq A$, $A \trianglerighteq^{+} B$ and $\neg B \trianglerighteq^{+} A$. So $A \succ B \mid_{\mathbf{M}}$.

evidential decision theory tells you to leave behind a guaranteed dollar because doing so gives you good news about matters over which you exercise no control. Though it's not obvious at first, on reflection, many of us have decided that this is irrational behaviour. In the same decision, the causal absolutist's advice is sensitive to instrumental value that you know for sure is not in the offing. In Frustrator, causal decision theory advises you to leave behind a guaranteed $49 and instead take a box that you know for sure will be empty, if you take it. And it leads causal decision theory to violate natural structural principles like Downwards and Upwards Preservation. In contrast, comparativism affords us theories—like the one I developed here, though there are others—that don't face these problems. They don't tell you to manage the news about matters over which you exercise no control. They don't tell you to pursue instrumental value that you know for sure is not in the offing. They vindicate the strong intuition that it is rational to take the envelope in Frustrator. And they validate plausible structural principles like Downwards and Upwards Preservation. To my mind, this gives us reason to take comparativism about instrumental value seriously, and to further explore comparativist theories of instrumental rationality.

REFERENCES

Ahmed, Arif. 2014. "Dicing with Death." In *Analysis*, **74** (4): 587–592. [11], [27]

Ahmed, Arif & Spencer, Jack. 2020. "Objective Value is Always Newcombizable." In *Mind*, **129** (516): 1157–1192. [1], [4], [5], [6], [9]

Arntzenius, Frank. 2008. "No regrets, or: Edith Piaf revamps decision theory." In *Erkenntnis*, **68** (2): 277–297. [27]

Bacon, Andrew. forthcoming. "Actual Value in Decision Theory." In *Analysis*. [2], [4], [7], [9]

Barnett, David James. 2022. "Graded Ratifiability." In *The Journal of Philosophy*, **119** (8): 57–88. [21], [25]

Briggs, R. A. 2010. "Decision-Theoretic Paradoxes as Voting Paradoxes." In *The Philosophical Review*, **119** (1): 1–30.

Dorr, Cian, Nebel, Jacob M., & Zuehl, Jake. forthcoming. "The case for comparability." In *Noûs*. doi:10.1111/nous.12407. [25]

Egan, Andy. 2007. "Some Counterexamples to Causal Decision Theory." In *The Philosophical Review*, **116** (1): 93–114. [27]

Gallow, J. Dmitri. 2016. "A Theory of Structural Determination." In *Philosophical Studies*, **173** (1): 159–186. [11]

Gallow, J. Dmitri. 2020*a*. "Review of *Newcomb's Problem*, edited by Arif Ahmed." In *Economics & Philosophy*, **36** (1): 171–176. [27]

Gallow, J. Dmitri. 2020*b*. "The Causal Decision Theorist's Guide to Managing the News." In *The Journal of Philosophy*, **117** (3): 117–149. [21]

Gallow, J. Dmitri. forthcoming. "Causal Counterfactuals without Miracles or Backtracking." In *Philosophy and Phenomenological Research*. [11]

Gibbard, Allan & Harper, William L. 1978. "Counterfactuals and Two Kinds of Expected Utility." In *Foundations and Applications of Decision Theory*, edited by A. Hooker, J.J. Leach, & E.F. McClennan, Dordrecht: D. Reidel, 125–162. [4], [9]

Greaves, Hilary. 2013. "Epistemic Utility Theory." In *Mind*, **122** (488): 915–952. [27]

Hare, Caspar & Hedden, Brian. 2016. "Self-Reinforcing and Self-Frustrating Decisions." In *Noûs*, **50** (3): 604–628. [27]

Harper, William. 1986. "Mixed Strategies and Ratifiability in Causal Decision Theory." In *Erkenntnis*, **24**: 25–36. [27]

Horgan, Terence. 1981. "Counterfactuals and Newcomb's Problem." In *The Journal of Philosophy*, **78** (6): 331–356. [23]

Jackson, Frank. 1991. "Decision-theoretic consequentialism and the nearest and dearest objection." In *Ethics*, **101** (3): 461–482. [19]

Joyce, James M. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press. [2], [4]

Joyce, James M. 2012. "Regret and instability in causal decision theory." In *Synthese*, **187** (1): 123–145. [4], [27]

Joyce, James M. 2018. "Deliberation and Stability in Newcomb Problems and Pseudo-Newcomb Problems." In *Newcomb's Problem*, edited by Arif Ahmed, Oxford: Oxford University Press. [4], [27]

Kment, Boris. forthcoming. "Decision, Causality, and Pre-Determination." In *Philosophy and Phenomenological Research*. [2]

Konek, Jason & Levinstein, Benjamin A. 2019. "The Foundations of Epistemic Decision Theory." In *Mind*, **128** (509): 69–107. [2], [4]

Lewis, David K. 1970. "How to Define Theoretical Terms." In *The Journal of Philosophy*, **67** (13): 427–446. [14]

Lewis, David K. 1973. *Counterfactuals*. Malden, MA: Blackwell Publishers. [4]

Lewis, David K. 1981. "Causal Decision Theory." In *Australasian Journal of Philosophy*, **59** (1): 5–30. [4]

Muñoz, Daniel & Spencer, Jack. 2021. "Knowledge of Objective 'Oughts': Monotonicity and the New Miners Puzzle." In *Philosophy and Phenomenological Research*, **103** (1): 77–91. [19]

Nozick, Robert. 1969. "Newcomb's Problem and Two Principles of Choice." In *Essays in Honor of Carl G. Hempel*, edited by Nicholas Rescher, Dordrecht: D. Reidel, 114–146. [21]

Parfit, Derek. unpublished. "What We Together Do." [10], [19]

Podgorski, Abelard. 2022. "Tournament Decision Theory." In *Noûs*, **56** (1): 176–203. [21]

Rabinowicz, Wlodek. 2009. "Letters from Long Ago: On Causal Decision Theory and Centered Chances." In *Logic, Ethics, and All That Jazz—Essays in Honour of Jordan Howard Sobel*, edited by Lars-Göran Johansson, Uppsala: Uppsala Philosophical Studies, volume 56, 247–273. [4]

Rabinowicz, Włodzimierz. 1982. "Two Causal Decision Theories: Lewis vs Sobel." In *Philosophical Essays Dedicated to Lennart Åqvist on His Fiftieth Birthday*, edited by Tom Pauli, Uppsala: Uppsala Philosophical Studies, volume 34, 299–321. [4]

Rachels, Stuart. 1998. "Counterexamples to the transitivity of *better than*." In *Australasian Journal of Philosophy*, **76** (1): 71–83. doi:10.1080/00048409812348201. [25]

Railton, Peter. 1986. "Moral Realism." In *Philosophical Review*, **95** (2): 163–207. [10]

Regan, Donald. 1980. *Utilitarianism and Cooperation*. Oxford: Oxford University Press. [19]

Richter, Reed. 1984. "Rationality Revisited." In *Australasian Journal of Philosophy*, **62** (4): 392–403. [27]

Russell, Bertrand. 1912. "On the Notion of Cause." In *Proceedings of the Aristotelians Society*, **13**: 1–26. [11]

Schaffer, Jonathan. 2005*a*. "Contrastive Causation." In *The Philosophical Review*, **114** (3): 297–328. [18]

Schaffer, Jonathan. 2005*b*. "Contrastive Knowledge." In *Oxford Studies in Epistemology 1*, edited by Tamar Szabo Gendler & John Hawthorne, Oxford University Press, 235. [18]

Schwartz, Thomas. 1972. "Rationality and the Myth of the Maximum." In *Noûs*, **6** (2): 97–117. [25]

Seidenfeld, Teddy. 1984. "Comments on Causal Decision Theory." In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 201–212. [23]

Shope, Robert K. 1978. "The Conditional Fallacy in Contemporary Philosophy." In *Journal of Philosophy*, **75** (8): 397–413. doi:10.2307/2025564. [10]

Skyrms, Brian. 1982. "Causal Decision Theory." In *Journal of Philosophy*, **79** (11): 695–711. [4]

Skyrms, Brian. 1990. *The Dynamics of Rational Deliberation*. Cambridge, MA: Harvard University Press. [4]

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell. [10]

Sobel, Jordan Howard. 1994. *Taking Chances: Essays on Rational Choice*. Cambridge: Cambridge University Press. [4]

Spencer, Jack. 2021. "An Argument Against Causal Decision Theory." In *Analysis*, **81** (1): 52–61. [4]

Spencer, Jack & Wells, Ian. 2019. "Why Take Both Boxes?" In *Philosophy and Phenomenological Research*, **99** (1): 27–48. [2], [4], [27]

Stalnaker, Robert C. 1980. "A Defense of Conditional Excluded Middle." In *Ifs*, edited by W. L. Harper, R. Stalnaker, & G. Pearce, Dordrecht: D. Reidel, 87–104. [4]

Stalnaker, Robert C. 1981. "Letter to David Lewis." In *Ifs*, edited by William Harper, Robert Stalnaker, & Glenn Pearce, Dordrecht: D. Reidel Publishing Company, 151–152. [4]

Suppes, Patrick & Zinnes, Joseph. 1963. "Basic Measurement Theory." In *Handbook of mathematical psychology, Volume I*, edited by D. Luce & Robert Bush, John Wiley & Sons., 1–76. [3], [13]

Temkin, Larry S. 1996. "A Continuum Argument for Intransitivity." In *Philosophy and Public Affairs*, **25** (3): 175–210. doi:10.1111/j.1088-4963.1996.tb00039.x. [25]

Walters, Lee & Williams, J. Robert G. 2013. "An Argument for Conjunction Conditionalization." In *The Review of Symbolic Logic*, **6** (4): 573–588. doi:10.1017/S1755020313000191. [9]

Weirich, Paul. 1985. "Decision Instability." In *Australasian Journal of Philosophy*, **63** (4): 465–478. [27]