Actual Value in Decision Theory

Andrew Bacon*

March 29, 2022

Decision theory is founded on the principle that we ought to take the action that has the maximum expected value from among actions we are in a position to take. The idea traces back to seventeenth century French mathematicians such as Pascal and Fermat, and has been influential since. But prior to the notion of expected value is the notion of the *actual value* of that action: roughly, a measure of the good outcomes you would in fact procure if you were to take it.

Philosophically speaking, the notion of actual value is surprisingly nuanced. I will show that if the pretheoretic notion is in good standing, then there will be many cases where the actual value of an action is indeterminate and unknowable in principle. Nonetheless I believe the notion *is* in good standing and offer a definition of it in terms of counterfactuals. By contrast, decision theories that pay lip service to the principle that one should maximize expected value typically offer no analysis of actual value, and subsequently do not clearly conform to the principle of maximizing expected value. I show that a form of decision theory due to Stalnaker can be reformulated so as to be in line with the edict to maximize expected value.[1] By contrast, I will prove that there is no quantity that plays the role of actual value in the decision theory of Jeffrey — whether given by my proposed definition of actual value or otherwise — and so it cannot be similarly restated.[2] I suggest this serves as a way of directly motivating Stalnaker's theory from a founding principle of decision theory, rather than as a revision of it intended to solve Newcomb's paradox as it is sometimes presented.

*Thanks to Frank Hong and two anonymous referees for several helpful comments on earlier drafts. Special thanks are due to Snow Zhang, who provided the characterization of decision theories satisfying The Expectation Condition in the appendix.

[1]The canonical presentation of Stalnaker's theory is Gibbard and Harper (1978); it was first outlined in Stalnaker (1978).

[2]The canonical presentation of Jeffrey's decision theory may be found in his book Jeffrey (1965).

1 Actual and Expected Value

The distinction between actual value and expected value is best introduced by example. Suppose that a coin has been flipped but you do not know the outcome. You are offered a bet that costs \$1 and pays out \$6 if the coin has landed heads, and pays out nothing otherwise. As it happens the coin has landed tails. So the actual value of the bet is in fact -\$1: you must pay \$1 to participate but you will receive nothing because the coin has landed tails.

On the other hand, you do not know that the coin has landed tails. Given what you know, there is a 50% chance that the coin has landed heads, and in that case the actual value of the bet would be \$5, since you have paid \$1 and made \$6. The expected value is what you get by taking an average of the possible actual values of the bet, weighted by how likely you think those actual values are: $(0.5 \times \$-1) + (0.5 \times \$5) = \$2$. Thus the expected value of the bet is \$2, even though the actual value of the bet is \$-1. And indeed, given your state of knowledge, the bet seems to be a good one and this is indeed predicted by the available versions of decision theory out there.

The actual value of the proposition that you accept the bet is easy enough to figure out from the description of the scenario. \$-1 represents the amount of money you would have ended up with if you had accepted the bet. Of course, this counterfactual holds because the coin in fact landed tails: in a world where the coin landed heads, the actual value of accepting the bet would have been \$5 — this is the amount of money I would have been better off by if I had accepted the bet in this alternate world.

Observe also that the counterfactual in our explanation of actual value is not idle: the actual value of accepting the bet is *not* the amount of money I am in fact going to end up with after making my decision. It could be that, for whatever reason, I decide not to take the bet — in which case the actual amount of money I lose or gain is \$0. But the actual value of accepting the bet is still -\$1, because that is what I would have made if I *had* accepted the bet.[3] Thus:

> The actual value of an action is a measure of the good or bad outcomes that would result if you were to take that action.

[3] In formal treatments of decision theory, the counterfactual nature of actual value is often is hidden behind the formalism. For instance the framework of Savage (1954) a decision problem is usually represented by a decision table whose columns are states and whose rows are actions: the entry under a given state and action represents the outcome that would result if you were to take that action in that state. Formally this is represented by a function O which maps any pair of a state and an action to an outcome, $O(s, A)$, informally glossed as the outcome that *would have obtained* if you had performed action A in state s.

Let us try and turn this into a definition of actual value suitable for mathematical analysis. First some preliminaries. Let us assume, as usual, that an agent's desires can be represented by a utility function $u : W \to \mathbb{R}$ mapping possible worlds to real numbers. A possible world settles all matters and, *ipso facto*, any matter the agent could possibly care about: $u(w)$ can be understood as a measure of the total amount of good or bad things that happen to the agent in world w. Let an action proposition for an agent be a proposition which the agent is in a position to make true, such as the proposition that you accept or decline a bet. Note that the utility of a world is *not* the same as the actual value of the action A at that world. This is simply a restatement of a previous observation in new terminology, but it is worth reiterating: Assuming you reject the bet, and have nothing else going on in your life, we may set the utility of the actual world to 0 (pretending for the moment that utility scales linearly with money, and choosing the obvious units), since you will in fact neither lose nor gain anything. But the actual value of accepting the bet is still -1.

We are now in a position to offer an analysis of the actual value of an action proposition, A, based on our previous informal definition: it is the utility for the agent of the way things would have been if A had been true. Given a proposition A, we may write $f(A, @)$ for the way things would have been if A had been true, according to the actual world @. (f is thus a *selection function* in the sense of Stalnaker (1968).) And the actual value of that proposition is the utility of the way things would have been if A had been true:

Actual Value The actual value $v_A(@)$ of a proposition A is defined as the utility of the world that would have obtained if A had been true, $u(f(A, @))$.

We have observed above that the actual value of a proposition is highly contingent: the actual value v_A is really a function that has a world as an argument. In a world w where the coin landed heads instead of tails prior to the bet being offered, the actual value of the bet is \$5 not -\$1. The world that would have obtained if I'd accepted the bet there, $f(A, w)$, is a world where I pay \$1 and make \$6 (and so is distinct from the world $f(A, @)$). So we have a different number $v_A(w) := u(f(A, w))$ representing the actual value in the world w: the utility of the way things would have been if you had accepted the bet in a world where the coin landed heads. In technical parlance, v_A is a random variable: it represents a contingent value which depends on the state of the world.

2 Actual Value Can Be Indeterminate

By contrast with the mathematically sophisticated notion of expected value, the notion of actual value seems straightforward. However, consider a variant of our example. Suppose that the payout is exactly the same as before — it costs \$1 and pays out \$6 if the coin lands heads — except the coin will only be flipped if you take the bet. And as it turns out you do not take the bet, so the coin is not flipped. Indeed we can suppose that the coin is subsequently destroyed before it is ever flipped. What is the actual value of accepting the bet? The only reasonable answers are as before: -\$1 or \$5. But unlike the previous case we cannot inspect the coin to see which it is. Nor can we flip it later to see what the result would have been (assuming this would indeed settle the question of how it would have landed if it had been flipped earlier for the bet). And there seems to be no other way to figure out the answer to the question of what the actual value is.

What does our proposed analysis have to say about this? Note that we have taken sides on a contentious principle of conditional logic:

Conditional Excluded Middle $(A \mathrel{\Box\!\!\rightarrow} B) \lor (A \mathrel{\Box\!\!\rightarrow} \neg B)$

This is a consequence of our implicit assumption that there is a particular world representing the way things would have been if A had been true. And that must either be a world where B holds, or not, so one of the two counterfactuals $A \mathrel{\Box\!\!\rightarrow} B$ or $A \mathrel{\Box\!\!\rightarrow} \neg B$ holds.[4]

According to proponents of conditional excluded middle, even if you do not in fact take the bet and the coin is never flipped, there is still a truth about what would have happened if you had taken the bet and it had been flipped. Either the coin would have landed heads if it had been flipped, or it would have landed tails. So according to our definition, the actual value is still perfectly well defined — it is either -\$1 or \$5 depending on which counterfactual is true. But perhaps it is an inherently chancey or indeterminate matter which of the two counterfactuals is true, in which case it is a similarly chancey or indeterminate matter what the actual value is.

3 Expected Value

Let us set aside for the time being the question of whether our analysis of actual value is correct. Let us suppose more neutrally that we have some

[4]Or else A is counterfactually inconsistent, and there is no such world. In this case, everything would have been the case if A.

random variable $v_A : W \to \mathbb{R}$ that maps each world to the actual value of taking the action A at that world, so that v_A just represents *some* contingent value which depends on the state of the world. If we do not know what the state of the world is, we will not know what the actual value is, but we can calculate the *expectation* by taking a weighted average of the possible actual values at each world weighted by how likely that world is according to your degrees of belief. One's degrees of belief may be represented by a probability function P assigning probabilities to possible worlds in such a way that the sum of the probabilities of all worlds adds up to 1.[5] The expected value of A is then:

Expected Value The expected value of an action A is the expectation of that propositions value: $\sum_w P(w)v_A(w)$.

Some things worth emphasizing: First, nothing about our formula for expectation is specific to the quantity actual value. You can calculate the expectation of any quantity you like in exactly the same manner: the expectation of tomorrow's temperature should be a weighted sum of the possible temperatures weighted by how likely you presently think those temperatures are. Second, you use your present degrees of belief about what the actual value, temperature, etc will be to calculate the expectation — not your future degrees of belief, someone else's degrees of belief, or your degrees of belief conditional on some supposition. Last, there is another salient quantity you can calculate the expectation of: your utility function. The expectation of your utility function is a measure of how good things are generally expected to be for you. But it is not the expectation of the actual value of any particular action for as we noted earlier the utility of a world is not always the same as the actual value of an action at a world. Indeed, there is only one utility function, and there are many actions, so the expectation of your utility function alone couldn't possibly help in comparing two or more prospects.

[5] I am assuming, for simplicity, that there are countably many worlds, and also taking a stance on the question of countable additivity. The former of these assumptions could be dropped by replacing summation, in what follows, with a suitable sort of integration. The assumption of countable additivity, however, is more intimately entangled with the theory of expectations. It would take us too far afield to try to disentangle it from that theory, and I will make no attempt to do so here, even though I am generally sympathetic to many of the worries people have about countable additivity.

4 Expected Value as an Action Guiding Quantity

Decision theory was founded on the idea that the proposition you ought to make true is the one with maximum expected value. The fundamental principle behind this claim traces back to Pascal and Fermat: when you have the prospect of obtaining some good, the worth of that prospect should be directly proportional to the probability that you will obtain that good.[6] So, for instance, a $\frac{1}{2}$ probability of winning a free car should be half as good as receiving a free car with certainty.

There are plenty of other quantities that an action might maximize, such as actual value, or expected temperature, but these are not, on the view under consideration, action guiding in the same way. Our starting point is thus:[7]

Maximize Expected Value The action guiding quantity is identical to the expectation of actual value of an action.

I do not take this assumption to be incontrovertible, and I think there is philosophical value in bringing it into question once in a while.[8] I will not, however, be doing so here: Here I will take this assumption for granted, and see what upshots it has for two approaches to decision theory that offer real

[6]For the relevant history see, for instance, Ore (1960).

[7]I have stated the principle in terms of 'action guidingness' to make room for a variety of responses to the question of what to do when there is not a unique action with maximum expected value. Here is one possible answer in which expected value still plays an action guiding role: When there are several propositions with maximal expected value, they are all permissible. And when there is no proposition with maximal expected value, as can happen when infinitely many actions are available, any action above a certain threshold is permissible.

[8]One extremely natural alternative to Maximize Expected Value is the principle that one should instead maximize actual value. A common objection to this alternative is that this is not an edict one can follow: usually we do not know which action leads to the maximum actual value. However, given the fact that we are rarely completely introspective of our own mental states (famously argued, for instance, in Williamson (2000)), including our own evidence, degrees of belief and desires, there will be cases where we do not know which proposition has the maximum expected value either (Feldman (2006) makes a related point). Of course, this alternative rejects the starting point of Pascal and Fermat, that the value of a future gain is proportional to the probability of obtaining it, and so the Pascal-Fermat principle can be taken to be a completely seperate argument against maximizing actual value that does not rest of considerations of followability.

Another alternative, which also rejects our starting point, is that it is rational to be risk averse, in which case the thing you should maximize is not expected value but a modification which discounts for your tolerance to risk; see for instance the risk weighted expected utility of Buchak (2013).

valued quantities playing the action guiding role we have above ascribed to expected value. The quantities are given, respectively, by:[9]

Stalnaker's Formula $S(A) = \sum_w P(A \mathrel{\Box\!\!\rightarrow} w)u(w)$

Jeffrey's Formula $J(A) = \sum_w P(w \mid A)u(w)$

According to decision theories based on the former formula, it is the quantity S that you ought to maximize when you are deciding what to do, and according to decision theories based on the latter it is the quantity J that you ought to maximize.

It is striking, however, that neither of these formulas have the mathematical form of an expectation of another quantity. Stalnaker's formula tells you to take a weighted sum of utilities that is weighted not by your credences in different worlds, but your credences in certain counterfactuals. Jeffrey's formula is also a weighted sum of utilities, but it is not weighted by your degrees of belief — as an expectation should be — but rather by what your degrees of belief *would be* if you were to learn that you performed the relevant action. Moreover, both formulas are taking weighted sums of your utility function, which we have already seen to be different from actual value.

Despite this, neither formula is obviously inconsistent with Maximize Expected Value. For all we have said, $S(A)$ and $J(A)$ can be rewritten so as to be expectations of another quantity, v_A. I will not presuppose any particular analysis of v_A in what follows, but rather focus on the constraint that some such quantity exists whose expectation is J or S respectively. If such a quantity *does* exists, we have at least a partial foundation for the notion of actual value: we can "Ramsify" the theory, by treating talk of actual value as being about the quantity, whatever it might be, whose expectation is action guiding.

Let's begin with Stalnaker's formula. $S(A)$ is defined as the the sum, for each world w, of the probability of $A \mathrel{\Box\!\!\rightarrow} w$ multiplied by the utility of w. Now, the probability of $A \mathrel{\Box\!\!\rightarrow} w$ is the sum of the probability of the worlds where this counterfactual is true.[10] Since we previously abbreviated 'w is the world that would have obtained if A had been true at x' (i.e. $A \mathrel{\Box\!\!\rightarrow} w$ is true at x) with $f(A, x) = w$, we have that $P(A \mathrel{\Box\!\!\rightarrow} w) = \sum_{f(A,x)=w} P(x)$. So we

[9] Jeffrey's formula is stated this way for ease of comparison with Stalnaker's formula, and with the notion of expected value. Jeffrey has an equivalent version of the equation $J(A) = \sum_{i\in I} P(A_i \mid A)J(A_i)$, whenever $(A_i)_{i\in I}$ is a partition of A. In Jeffrey's framework the utility of a world is identified with its Jeffrey value so that the two formulations are equivalent.

[10] This is a consequence of countable additivity given our assumption that there are countably many worlds.

are looking at the sum, for worlds x and w such that $f(A, x) = w$, of the probability of x times the utility of w. Simplifying a little, this is the sum for all worlds x of the probability of x times the utility of $f(A, x)$, which has the form of an expectation.

$$\begin{aligned} S(A) = \sum_w P(A \mathbin{\Box\!\!\to} w)u(w) &= \sum_w \sum_{f(A,x)=w} P(x)u(w) \\ &= \sum_x P(x)u(f(A,x)) = \sum_x P(x)v_A(x) \end{aligned}$$

So we have shown that $S(A)$ my be rewritten as the expectation of some random variable. Indeed, not any random variable: pleasingly, it is the expectation of actual value according to our earlier proposed definition of actual value — $v_A(x)$ is the utility of the world that would have obtained if A had been true at x, $u(f(A, x))$.

Now let us turn to Jeffrey's formula: could there be a notion of actual value, v_A, whose expectation is always given by Jeffrey's quantity $J(A)$? I will make one assumption about actual value which I think follows from our motivating remarks about it: the actual value of an action (unlike its expected value) does not depend on your degrees of belief. In our running example, the actual value of the bet is -\$1 because the coin has in fact landed tails — this value does not depend on how likely you think it is that the coin landed heads or tails. Thus for any given utility function u and action proposition A there exists a quantity v_A such that for any probability function P for which $P(A) > 0$:[11]

$$\sum_w P(w \mid A)u(w) = \sum_w P(w)v_A(w)$$

The order of the quantifiers here reflecting the fact that v_A does not depend on P, but can potentially depend on u and A. (Note of course that our proposed definition of actual value does depend on both u and A, but does not depend on P.) But it may be shown that there is *no* quantity that satisfies this condition for every probability function, given the modest assumption that it is possible to rationally prefer at least one proposition over another and that there are at least three possible worlds. I have placed the proof in an appendix.

I suspect there is some room for improvement on this result. For one thing, it is clear that the quantification over probability functions in our statement that actual value is independent of probability could be restricted in various

[11] When $P(A) = 0$ Jeffrey's formula is undefined so the restriction to actions with positive probability is necessary. This is a well-known shortcoming of Jeffrey's formula, because it means we cannot apply decision theory in cases where we know what we're going to do.

ways without affecting the result, but unclear what the exact class of restrictions for which the result still holds is. Secondly, one might ask what decision theories *can* be formulated in a way where the action guiding quantity is an expectation of a quantity that is independent of your beliefs (in the sense spelled out above). In personal communication Snow Zhang has provided a fairly general answer to this question: among a very wide class of decision theories, the theory's action guiding quantity is an expectation of a quantity that is independent of one's degrees of belief if and only if it is a causal decision theory based on a slight generalization of the Stalnakerian theory of selection functions.[12] A statement of this theorem may also be found in the appendix. There is finally a question of generalizing this impossibility result for evidential decision theory to other kinds of causal decision theory: for instance Lewis (1981) and others have formulated versions of causal decision theory that don't presuppose conditional excluded middle. But the notion of actual value seemingly requires you to talk about the things, good or bad, that would have happened if you had taken an action; without conditional excluded middle this way of talking would be ill-defined.

Could one resist the assumption that actual value doesn't depend on one's degrees of belief? To give up on this, in my view, is to give up on the distinction between actual and expected value altogether. Once we have rejected the pretheoretic notion of actual value, defined in terms of claims about how good things would have been for you if you had taken certain courses of action, what is left of the distinction between expected value and actual value? We know the former quantity is the expectation of the latter quantity, and it is supposed to be the operation of expectation that takes your subjective uncertainty into account. This is already thin, and it is thinner still once we suppose that actual value is something that already takes into account your uncertainty about the world.

But some will no doubt maintain that this is the right moral to draw: the pretheoretic notion of actual value is not in good standing, especially in light of the widespread indeterminacy of actual value. John Broome, for instance, has suggested that the only kind of value is expected value — a quantity which is its own expectation, and thus plays the role of actual and expected value simultaneously.[13] This assumption is also rejected by Konek and Levinstein

[12] Personal communication, July 9, 2021

[13] See chapter 6 of Broome (1991). There are, however, some worries for this proposal: in the presence of higher-order uncertainty, the expectation of the expectation of a quantity can come apart from its expectation. For a quantity to be equal to its own expectation, it either has to be trivial (i.e. have a constant value across worlds), or rule out certain patterns of higher-order uncertainty which have the appearance of being perfectly rational. Suppose

(2019), where they instead define actual value for Jeffrey decision theory in terms of credences as $v_A(w) := \frac{P(w|A)}{P(w)}u(w)$. It is easily seen that $J(A)$ is indeed the expectation of this quantity. However, once we let in quantities that depend on one's credences, there are many other such quantities: $J(A)$ is also the expectation of the constant function $v_A(w) = J(A)$ (note that $J(A)$ is defined in terms of P)—this is in effect the Broomian proposal—but also infinitely many other cooked up quantities. This radical non-uniqueness just illustrates how undemanding the relationship between actual value and expected value is without the assumption that actual value is independent of credences. By contrast, if we restrict to credence independent quantities, then the notion of actual value is *uniquely* pinned down by its relationship to expected value: for a given action and utility function there is at most one quantity whose expectation relative to every probability function is identical to its action worthiness relative to that utility function (see theorem 2 below).

5 Conclusion

Decision theory in the tradition of Stalnaker (1978) is often labeled 'causal decision theory'. These presentations typically introduce Stalnaker's formula as a *revision* of naïve decision theory that is required to deal with Newcomb's paradox — a revision that complicates the notion of expected value with causal or counterfactual notions.[14]

The present way of motivating Stalnaker's formula, by contrast, does not invoke causation or Newcomb's paradox or anything like that: it is motivated by the founding principle of maximizing expected value. Counterfactuality enters the picture because the notion of actual value is inherently counterfactual.

Appendix

Let W, the 'worlds', be a countable set containing at least three members. Suppose that $V(P, A, u)$ ('action worthiness') is a real valued quantity that depends on a countably additive probability function on W, P, a set of worlds

v_A is non-trivial and has different values at the worlds x and y, α and β, and imagine that at world x you are 50/50 concerning whether world x or y obtains, but in world y you are certain that y obtains. The expectation of v_A at x and y is respectively $\frac{1}{2}\alpha + \frac{1}{2}\beta$ and β. So the expectation of the expectation at x is $\frac{1}{2}(\frac{1}{2}\alpha + \frac{1}{2}\beta) + \frac{1}{2}\beta = \frac{1}{4}\alpha + \frac{3}{4}\beta$ which is different from $\frac{1}{2}\alpha + \frac{1}{2}\beta$, the expectation at x, since $\alpha \neq \beta$.

[14]See for instance Joyce (1999), or Weirich (2009).

A (the 'action proposition'), and a utility function $u : W \to \mathbb{R}$. V is an expectation of a credence independent quantity iff:

The Expectation Condition For every utility function u and proposition A, there exists a quantity $v : W \to \mathbb{R}$ such that for every probability function P: $V(P, A, u) = \sum_{w \in W} P(w)v(w)$.

Theorem 1. *Jeffrey's quantity,* $J(P, A, u) = \sum_w P(w \mid A)u(w)$ *does not satisfy The Expectation Condition.*

Proof. Suppose that Jeffrey's theory can recommend one proposition over another. There is a rational credence utility function pair, P and u, and pair of incompatible propositions X and Y such that the resulting Jeffrey values are different: $J(X) \neq J(Y)$. We may assume too that such an X and Y may be chosen so that $X \cup Y$ is not the whole space. Let $A = X \cup Y$. Now let $S = \{P' \mid P'$ is a probability function such that $P'(Y) = 0$ and $P'(\cdot \mid X) = P(\cdot \mid X)\}$. For any $P' \in S$ we have that $J'(X) = J(X)$ where J' is the Jeffrey value with respect to P' and u (since $J'(X) = \sum_w P'(w \mid X)u(x) = \sum_w P(w \mid X)u(x) = J(X)$). So for every $P' \in S$, $\sum_w P'(w)v_A(w)$ has a fixed value $J(X)$, because $\sum_w P'(w)v_A(w) = \sum_w P'(w \mid A)u(w) = \sum_w P'(w \mid X)u(w) = J'(X) = J(X)$. The first identity follows from our hypothesis that the expectation (according to P') of actual value of A is identical to its Jeffrey value, and the second identity because $P'(Y) = 0$ and thus $P'(\cdot \mid A) = P'(\cdot \mid X \cup Y) = P'(\cdot \mid X)$. Now this can happen only if $v_A(w) = J(X)$ for every $w \notin A$. For if there are worlds such that $v_A(w) > J(X)$ you could choose a probability function $P' \in S$ that assigns lots of probability to the worlds $w \notin A$ that have actual value that exceeds $J(X)$ ($v_A(w) > J(X)$), and assigns little probability to the remaining worlds, so ensuring that $\sum_w P'(w)v_A(w) > J(X)$ contradicting our previously established claim.[15] For similar reasons, there cannot be worlds

[15] This can be broken up into two claims. Suppose that X and Y are as above, and Z and U partition $W \setminus A$ into two non-empty sets. First, for any positive α and β with $\alpha + \beta < 1$ we can construct a probability function P' such that $P'(Z) = \alpha$, $P'(U) = \beta$, $P(Y) = 0$ and is such that $P'(\cdot \mid X) = P(\cdot \mid X)$ (the last two conditions simply ensure that $P' \in S$). Namely assign probabilities to worlds in Z that add up to α, probabilities to worlds in U that add up to β, assign 0 probability to worlds in Y, and for worlds $x \in X$ set $P'(x) = \frac{(1-\alpha-\beta).P(x)}{P(X)}$. Now if $v_A(w) > J(X)$ for some w, we may pick some $\epsilon > 0$ and let $Z = \{w \mid v_A(w) \geq J(X) + \epsilon\}$; and it is just a matter of choosing α to be big enough that the expectation of the Z part of logical space swamps the rest. For instance we could set $\beta = 0$ and $\alpha > \frac{J(X)-\gamma}{J(X)-\gamma+\epsilon}$ where $\gamma = \sum_x P(x \mid X)v_A(x)$. γ may be proved to be finite since $\gamma = \sum_x P(x \mid X)v_A(x) = \sum_x P(x \mid X)u(x)$ by plugging the probability function $P(\cdot \mid X)$ into our relationship between Jeffrey value and expected value. But the last value is just $J(X)$ which we know to be finite.

such that $v_A(w) < J(X)$. But then by parallel reasoning — switching X and Y in all of the above — we can see that $v_A(w) = J(Y)$ for every $w \notin A$, so we have a contradiction since $J(X) \neq J(Y)$ and we assumed that $W \setminus A$ was non-empty.[16] □

Fix an action A and a utility function u. Say that a quantity $v : W \to \mathbb{R}$ satisfies the actual value role, with respect to action worthiness, $V(P, A, u)$, if and only if $V(P, A, u)$ is the expectation of v for every probability function P: for every P, $V(P, A, u) = \sum_w P(w)v(w)$.

Theorem 2. *At most one quantity satisfies the actual value role with respect to a given measure of action worthiness, $V(P, A, u)$.*

Proof. Suppose v and v' satisfy the actual value role with respect to $V(P, A, u)$. Let w be an arbitrary world, and let P_w be the probability distribution that is certain in w. Then $V(P, A, u) = \sum_x P_w(x)v(x) = v(w)$ and $V(P, A, u) = \sum_x P_w(x)v'(x) = v'(w)$, and so $v(w) = v'(w)$. Since w was arbitrary $v = v'$. □

Assume for simplicity that W is now finite. Let a supposition procedure be a function s that maps a probability distribution P and a proposition A to another probability distribution, $s(P, A)$ such that $s(P, A)(A) = 1$. Bayesian conditioning, given by $s(P, A) = P(\cdot \mid A)$, and imaging, given by $s(P, A) = P(A \,\Box\!\!\rightarrow \cdot)$, are two examples of supposition procedures. Supposition procedures give rise to a very general class of decision theories where the action guiding quantity is defined by taking the expectation of u relative to the supposition that you have taken action A:

$$V_s(P, A, u) := \sum_w s(P, A)(w)u(w)$$

[16] Since submitting this article I have learned that Ahmed and Spencer (2020) have provided a different argument that, in certain special cases, Jeffrey value is not an expectation. The sorts of cases their arguments concern are somewhat unusual because they involve situations where learning that you are going to take a course of action provides evidence about the actual value of that action. Now usually, learning what you are going to do provides no evidence about that actions actual value: whether you accept the bet I outlined in the introduction has no evidential bearing on how the coin will land, and thus on whether you will win the bet. For better or for worse, many decision theorists, following Savage (1954), have built this independence assumption into their framework. Whether or not this is a good idea, it is clear that these decision theories are applicable in most cases. Crucially, the argument presented above is not restricted to these unusual cases: it shows that Jeffrey value cannot be an expectation even if you make the sorts of idealizations common among decision theorists working in the tradition of Savage.

Among this general class of decision theories is Jeffrey's decision theory (where s is Bayesian conditioning) and certain 'causal decision theories' where s is determined by a generalized kind of selection function. A general imaging function (see Gardenfors (1982)) is a function f that maps a proposition A and a world w to a probability distribution $f(A, w)$ that assigns A probability 1 ($f(A, w)(A) = 1$) (a Stalnakerian selection function may be thought of as the special case where $f(A, w)$ assigns some world probability 1). The supposition procedure corresponding to a general imaging function is $s(P, A)(B) = \sum_w f(A, w)(B)P(w)$, and the corresponding version of causal decision theory is given by the action guiding quantity $V_s(P, A, u)$ defined above. Say that a *general decision theory* is one whose action guiding quantity has the form $V_s(P, A, u)$ for some supposition procedure s. A general decision theory satisfies The Expectation Condition if $V_s(P, A, u)$ does, and is a *Stalnakerian causal decision theory* if s is determined by a general imaging function.

Theorem 3 (Zhang)**.** *A general decision theory satisfies The Expectation Condition if and only if it is a Stalnakerian causal decision theory.*

References

Arif Ahmed and Jack Spencer. Objective value is always newcombizable. *Mind*, 129(516):1157–1192, 2020. doi: 10.1093/mind/fzz070.

John Broome. *Weighing Goods: Equality, Uncertainty and Time.* Wiley-Blackwell, 1991.

Lara Buchak. *Risk and Rationality.* Oxford University Press, 2013.

Fred Feldman. Actual utility, the objection from impracticality, and the move to expected utility. *Philosophical Studies*, 129(1):49–79, 2006. doi: 10.1007/s11098-005-3021-y.

Peter Gardenfors. Imaging and conditionalization. *Journal of Philosophy*, 79 (12):747–760, 1982. doi: 10.2307/2026039.

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

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

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

Jason Konek and Benjamin A. Levinstein. The foundations of epistemic decision theory. *Mind*, 128(509):69–107, 2019. doi: 10.1093/mind/fzw044.

David Lewis. Causal decision theory. *Australasian Journal of Philosophy*, 59 (1):5–30, 1981. doi: 10.1080/00048408112340011.

Oystein Ore. Pascal and the invention of probability theory. *The American Mathematical Monthly*, 67(5):409–419, 1960. ISSN 00029890, 19300972. URL http://www.jstor.org/stable/2309286.

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

Robert Stalnaker. Letter to david lewis. In Robert C. Stalnaker William L. Harper and Glenn Pearce, editors, *Ifs: conditionals, belief, decision, chance, and time*, pages 151–152. D. Reidel, 1978.

Robert C. Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, pages 98–112. Oxford: Blackwell, 1968.

Paul Weirich. Causal decision theory. In *Stanford Encyclopedia of Philosophy*. 2009.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, 2000.