RELATIVE EXPECTATION THEORY*

The St. Petersburg game and its nearby neighbors present serious problems for decision theory.[1] The St. Petersburg game invokes an unbounded utility function to produce an infinite expectation for playing the game. The problem is usually presented as a clash between decision theory and intuition: most people are not prepared to pay a large finite sum to buy into this game, yet this is precisely what decision theory suggests we ought to do. But there is another problem associated with the St. Petersburg game. This second problem is that standard decision theory counsels us to be indifferent between any two acts that have infinite expected utility. So, for example, consider the decision problem of whether to play the St. Petersburg game or a game where every payoff is $1 higher. Let us call this second game the *Petrograd game* (it is the same as St. Petersburg but with a bit of twentieth-century inflation). Standard decision theory tells us to be indifferent between these two options.

It might be argued that any intuition that the Petrograd game is better than the St. Petersburg game is a result of misguided and naïve intuitions about infinity.[2] But this argument against the intuition in question is misguided. The Petrograd game is clearly better than the St. Petersburg game. And what is more, there is no confusion about infinity involved in thinking this. When the series of coin tosses comes to an end (and it comes to an end with probability 1), no matter how many tails precede the first head, the payoff for the Petrograd game is one dollar higher than the St. Petersburg game. Whatever the

*Thanks to Alan Hájek, Adam La Caze, Aidan Lyon, Tony Martin, Martin Rechenauer, Katie Steele, and Juliana Weingaertner. I am also grateful to audiences at the 2005 Stradbroke Island Philosophy and Linguistics Cogitamus, the 2005 Australasian Association of Philosophy (New Zealand) conference at the University of Otago in Dunedin, the Philosophy Department at the University of California, Los Angeles, and the Philosophy Department at the Ludwig-Maximilians University in Munich.

[1] Recall that the St. Petersburg game is a game where a fair coin is tossed repeatedly, if necessary, until the first head appears. The game is then over. The payoff for this game is $2 if the head appears on the first toss, $4 for heads on the second toss, $8 on the third, and so on. See Robert Martin, "The St. Petersburg Paradox," in Edward Zalta, ed., *The Stanford Encyclopedia of Philosophy (Fall 2001 Edition)*, URL = <http://plato.stanford.edu/archives/fall2001/entries/paradox-stpetersburg/>.

[2] These are the same intuitions that incorrectly suggest that $1 + \infty > \infty$, or so the story goes.

0022-362X/08/0501/37–44

outcome, you are better off playing the Petrograd game. Infinity has nothing to do with it. Indeed, a straightforward application of dominance reasoning backs up this line of reasoning.[3] Standard decision theory (in the form of expected utility maximization) cannot deliver this verdict, but so much the worse for standard decision theory. Dominance reasoning gets it right, and this is significant.

I have argued elsewhere[4] that there are a number of cases like this, where dominance reasoning delivers the correct result and expected utility maximization is either silent or indifferent. In light of such cases, it would seem that we have a kind of unsettling pluralism about decision rules on our hands: expected utility maximization and dominance reasoning are both required—neither will do on its own. So here is the first challenge: provide a generalization of decision theory that will unify the rule of expected utility maximization and the rule of dominance.

Next consider a variation on the Petrograd game. Like the Petrograd, this game has payoffs \$1 higher than the corresponding payoffs of the St. Petersburg game—except for one. The exception is a payoff for some very low probability state and this is \$1 *less* than the corresponding St. Petersburg payoff. Call this game the *Leningrad game*. Here expected utility theory suggests that we ought to be indifferent between the Leningrad game and the St. Petersburg game; dominance reasoning is not applicable and so is silent. But there is a very strong intuition that the Leningrad game is better than the St. Petersburg game. After all, the Leningrad game almost dominates the St. Petersburg game, and the probability of finding oneself in the nondominant state is, by construction, very low. So here is the second challenge: either find a decision rule that supports this intuition or explain away the intuition. But note how hard the latter will be. Any argument that explains away the intuition must not also explain away the intuition that the Petrograd game is better than the St. Petersburg game, for the latter intuition is certainly correct. So the real challenge here is to provide a decision-theoretic validation of the intuition that the Leningrad game is preferable to the St. Petersburg game.

In this paper I will present new rules of decision that meet both these challenges. That is, I will provide a generalization of decision

[3] Recall that dominance reasoning suggests that one ought to choose act A_1 over act A_2 if in every state the utilities associated with A_1 are never less than the corresponding utilities for A_2, and in at least one state the utility of A_1 is higher than the corresponding utility for A_2. This rule is only applicable when the states are independent of the acts.

[4] Colyvan, "No Expectations," *Mind*, CXV (2006): 695–702.

theory that unifies expected utility theory and dominance reasoning. Indeed, the theory I advance is a generalization in the sense that it is a conservative extension of both dominance reasoning and finite expected utility theory. Moreover, I will show how the new theory is also able to deal with problem cases such as choosing between the Leningrad game and the St. Petersburg game.

I. INTRODUCING RELATIVE EXPECTED UTILITY

Assume throughout that we have a decision problem where the states are independent of the acts. We define the *relative expected utility* of act A_k over A_l as

$$\mathrm{REU}(A_k,A_l) = \sum_{i=1}^{n} p_i(u_{ki} - u_{li})$$

where p_i is the probability associated with state S_i and u_{ji} are the utilities of the outcome resulting from act A_j in state S_i. For infinite state spaces we define relative expected utility similarly:

$$\mathrm{REU}(A_k,A_l) = \sum_{i=1}^{\infty} p_i(u_{ki} - u_{li})$$

where the right-hand side converges or diverges to infinity. The decision rule is then as follows:

> REU Decision Rules: Choose act A_k over act A_l iff $\mathrm{REU}(A_k,A_l) > 0$. If $\mathrm{REU}(A_k,A_l) = 0$ an agent should be indifferent between the two acts in question.

The resulting theory I call *relative expectation theory*. The basic idea here is clear. Relative expected utility is the expected gain or loss in utility between the two acts in question. In other words, it is the expected relative advantage in choosing one act over another.[5] It is also worth noting that there is a sense in which the expected utility can be thought of as a special case of relative expected utility: the expected utility of an act is the relative expected utility of that act over an act with zero payoffs for every state (the null act).[6] On the

[5] The basic idea of considering the expectations of differences was advanced independently in Alan Hájek, "In Memory of Richard Jeffrey: Some Reminiscences and Some Reflections on *The Logic of Decision*," *Philosophy of Science*, LXXIII (2006): 947–58.

[6] We need to be careful though. For the standard axioms for utility theory, the von Neumann-Morgenstern axioms, do not allow us to make sense of a distinguished

other hand, it might be argued that relative expectation theory is nothing new; economists may well have had something like this in mind when considering expected profits and losses, for example. After all, since utilities are random variables, their differences are also random variables, and relative expectation theory just considers the expectation of these.[7] While there is a sense in which this is right, it turns out that the advice relative expectation theory gives in the problem cases I opened with is quite different from that given by standard decision theory. Standard decision theory counsels us to calculate the expected utility of each act and choose the act that has the greatest expected utility (if there is such an act). Relative expectation theory gives no such advice. In fact, relative expectation theory does not even require the calculation of expectations for individual acts. In the next section I will have more to say about the relationship between standard decision theory and relative expectation theory. As we shall see, the two theories are closely related but the advice they give is quite different. I will thus continue to speak of relative expectation theory as a distinct theory of decision. So much for what relative expectation theory is, let me now say a little about what it is good for.

It is straightforward to show that the above decision rules support the intuition to choose the Petrograd game over the St. Petersburg game (since the relative expected utility of the Petrograd game over the St. Petersburg game is $\Sigma_{i=1}^{\infty} 1/2^i = 1$). It is also straightforward to show that so long as the nondominant utility in the Leningrad game is associated with a state with probability less than a half, we should choose the Leningrad game over the St. Petersburg game. It should also be intuitively clear that the above decision rules agree with dominance reasoning, whenever the latter applies, and agree with expected utility theory in at least cases with finite state spaces. That is, relative expectation theory is a conservative extension of both the rule of dominance and finite expected utility theory. (I will back up this last claim shortly.)

Relative expected utility theory clearly has a lot going for it. Indeed, it would seem to be a good candidate for the unification of dominance reasoning and expected utility theory which we seek. It re-

---

zero. See John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton: University Press, 1944). Like temperature scales (and unlike distance scales), the zero of utility theory is purely conventional. Thanks to Juliana Weingaertner for reminding me of this point.

[7] Thanks to an anonymous referee for this suggestion.

mains to demonstrate this though. I do this in the next section by proving a number of theorems about relative expected utility.

II. SOME ELEMENTARY THEOREMS

First a couple of theorems that shows that our decision rules are well defined.[8] That is, I show that the above rules do not give us contradictory advice.

Theorem 1 If $\mathrm{REU}(A_1,A_2) > 0$, $\mathrm{REU}(A_2,A_1) < 0$.

Proof: Without loss of generality, assume that the state space is infinite. $\mathrm{REU}(A_1,A_2) = \sum_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) > 0$, and since this series is defined it either converges or diverges to positive infinity.[9] In either case $\mathrm{REU}(A_2,A_1) = \sum_{i=1}^{\infty} p_i(u_{2i} - u_{1i}) = -\mathrm{REU}(A_1,A_2) < 0$.

A similar proof establishes a symmetry result for when the relative expected utility is zero.

Theorem 2 If $\mathrm{REU}(A_1,A_2) = 0$, $\mathrm{REU}(A_2,A_1) = 0$.

Having established that our new decision rules are well defined, I now prove a few results that establish the relationship between relative expectation theory and our earlier decision rules.

Theorem 3 If $\mathrm{EU}(A_1) > \mathrm{EU}(A_2)$, $\mathrm{REU}(A_1,A_2) > 0$.

Proof: Again without loss of generality assume that the state space is infinite. There are four cases to consider: (i) $\mathrm{EU}(A_1)$ and $\mathrm{EU}(A_2)$ both converge; (ii) $\mathrm{EU}(A_1)$ diverges to infinity and $\mathrm{EU}(A_2)$ converges; (iii) $\mathrm{EU}(A_1)$ converges and $\mathrm{EU}(A_2)$ diverges to negative infinity; and (iv) $\mathrm{EU}(A_1)$ diverges to positive infinity and $\mathrm{EU}(A_2)$ diverges to negative infinity. Consider case (i) first. Since $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i} > \Sigma_{i=1}^{\infty} p_i u_{2i} = \mathrm{EU}(A2)$, $\Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$. In case (ii), we have $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i}$ is infinite and $\Sigma_{i=1}^{\infty} p_i u_{2i} = \mathrm{EU}(A_2)$ is finite, so $\mathrm{EU}(A_1) - \mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{1i} - \Sigma_{i=1}^{\infty} p_i u_{2i} = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$. In case (iii), we have $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i}$ is finite and $\mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{2i}$ is negatively infinite, so $\mathrm{EU}(A_1) - \mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{1i} - \Sigma_{i=1}^{\infty} p_i u_{2i} = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$. In case (iv), we have $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i}$ is positive infinite and $\Sigma_{i=1}^{\infty} p_i u_{2i} = \mathrm{EU}(A_2)$ is negative infinite, so $\mathrm{EU}(A_1) - \mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{1i} - \Sigma_{i=1}^{\infty} p_i u_{2i} = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$.

And there is a similar proof that establishes the corresponding case for identity.

[8] For all these theorems I continue to assume that the states are independent of the acts. I will also make use of some results about well-behaved divergent series. See G.H. Hardy, *Divergent Series* (New York: Oxford, 1949) for more on this topic.

[9] In this and other proofs, I adopt the common convention of treating a series that diverges to infinity as being well defined and having infinity as its limit.

Theorem 4 If $\mathrm{EU}(A_1) = \mathrm{EU}(A_2)$ and is finite, then $\mathrm{REU}(A_1,A_2) = \mathrm{REU}(A_2,A_1) = 0$.

These two results tell us that relative expectation theory agrees with expected utility theory in finite cases (and whenever the states are independent of the acts). It is important to note that theorem 3 holds only in the direction stated. Indeed, it is the fact that relative expectation theory is more discriminating than expected utility theory that allows us to use relative expectation theory to justify choosing the Petrograd game over the St. Petersburg game. Inequalities in relative expected utilities do not imply inequalities in expected utilities and we do not require the expectation series to converge in order to make meaningful comparisons of actions.

Next a theorem relating relative expected utility to dominance.

Theorem 5 If A_1 dominates A_2, $\mathrm{REU}(A_1,A_2) > 0$.

Proof: Again without loss of generality assume that the state space is infinite. Since A_1 dominates A_2, each u_{1i} is greater than or equal to each u_{2i} and at least one u_{1i} is strictly greater than a corresponding u_{2i}. But this implies that $\mathrm{REU}(A_1,A_2) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) > 0$.

This result tells us that relative expectation theory agrees with dominance reasoning.

Next a couple of transitivity results.

Theorem 6 If $\mathrm{REU}(A_1,A_2) > 0$ and $\mathrm{REU}(A_2,A_3) > 0$, then $\mathrm{REU}(A_1,A_3) > 0$.

Proof: Once again, without loss of generality, assume that the state space is infinite. $\mathrm{REU}(A_1,A_2) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) > 0$ and $\mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) > 0$. It follows that $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) > 0$. But since the series in question either converge or diverge to positive infinity, $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) + \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{3i}) = \mathrm{REU}(A_1,A_3) > 0$.

Theorem 7 If $\mathrm{REU}(A_1,A_2) = 0$ and $\mathrm{REU}(A_2,A_3) = 0$, then $\mathrm{REU}(A_1,A_3) = 0$.

Proof: Once again, without loss of generality, assume that the state space is infinite. $\mathrm{REU}(A_1,A_2) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = 0$ and $\mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) = 0$. It follows that $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) = 0$. But $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) + \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{3i}) = \mathrm{REU}(A_1,A_3) = 0$.

Finally, the promised conservativeness result.

Theorem 8 Relative expected utility theory is a conservative extension of both finite expected utility theory and dominance reasoning.

Proof: This follows from theorems 3, 4, and 5.

III. DISCUSSION

Theorem 8 is the crucial one. This shows that relative expectation theory is the generalization we seek. Moreover, relative expectation theory provides a satisfying account of the near-dominance cases such as the Leningrad game versus the St. Petersburg game. And it does this in a unified way. There is no special pleading for problem cases.[10]

There are, however, some limitations to this approach that deserve commenting on. The first is the assumption that the states are independent of the acts. I relied on this assumption in the very definition of relative expected utility, when I required that the probabilities for each state do not depend on the acts. Dispensing with this assumption and thus generalizing the notion of relative expected utility would free things up, but it is not obvious how to do this. In any case, as things stand the definition of relative expected utility and thus the proofs of the theorems relied on the independence of states and acts, so we must live with this assumption for the time being at least.[11]

The second restriction is that the new theory can only deal with pair-wise comparisons of acts. Indeed, the very definition of relative expected utility only makes sense of pair-wise comparisons. In light of theorems 6 and 7, though, this is not such a serious limitation. For theorems 6 and 7 show us that we can make sense of decision problems with more than 2 acts—it is just that we need to compare the acts two-by-two. Theorems 6 and 7 tell us, in effect, that we can produce an ordering of the acts, and since any such ordering of finitely many acts will have a (perhaps nonunique) maximal act, we can solve the decision problem despite the limitation of using only pair-wise comparisons.[12]

[10] Although I will not pursue the details here, relative expectation theory also deals with a class of problems where expected utility theory is silent. I have in mind here the Pasadena game versus the Altadena game. See Harris Nover and Alan Hájek, "Vexing Expectations," *Mind*, CXIII (2004): 237–49; Hájek and Nover, "Perplexing Expectations," *Mind*, CXV (2006): 703–20; and Colyvan, "No Expectations." That this is so follows straightforwardly from theorem 5 above.

[11] This issue also touches on the treatment of Newcomb's paradox, see Robert Nozick, "Newcomb's Problem and Two Principles of Choice," in Nicholas Rescher, ed., *Essays in Honour of Carl G. Hempel* (Boston: Reidel, 1969), pp. 114–46. Since Newcomb's problem places dominance reasoning and expected utility theory in conflict with one another, it would suggest that dropping the independence condition will be no easy matter. For example, if we could prove analogues of theorems 3 and 5 without the independence assumption, Newcomb's problem would present itself as a serious problem for relative expectation theory. Perhaps what is required is a causal version of relative expectation theory; see James Joyce, *The Foundations of Causal Decision Theory* (New York: Cambridge, 1999).

[12] It is important to note that this is only the case when all the relative expected utilities are defined. There will be cases when they are not. Not surprisingly, relative

These limitations suggest that there is further work to be done here, but still relative expectation theory is genuinely useful in solving some persistent and troubling problems in infinite decision theory.

MARK COLYVAN

University of Sydney

expectation theory will not be able to solve these problems. (Consider, for example, the case of play or not to play the Pasadena game. Since the Pasadena game does not have an expectation, there is no relative expectation for playing the Pasadena game over not playing it. See Nover and Hájek, "Vexing Expectations," Hájek and Nover, "Perplexing Expectations," and Colyvan, "No Expectations.")