Erkenntnis (2006) 64:85–100
DOI 10.1007/s10670-005-7515-y

MICHAEL CLARK and NICHOLAS SHACKEL

THE DR. PSYCHO PARADOX AND NEWCOMB'S PROBLEM

ABSTRACT. Nicholas Rescher claims that rational decision theory "may leave us in the lurch", because there are two apparently acceptable ways of applying "the standard machinery of expected-value analysis" to his Dr. Psycho paradox which recommend contradictory actions. He detects a similar contradiction in Newcomb's problem. We consider his claims from the point of view of both Bayesian decision theory and causal decision theory. In Dr. Psycho and in Newcomb's Problem, Rescher has used premisses about probabilities which he assumes to be independent. From the former point of view, we show that the probability premisses are not independent but inconsistent, and their inconsistency is provable within probability theory alone. From the latter point of view, we show that their consistency can be saved, but then the contradictory recommendations evaporate. Consequently, whether one subscribes to evidential or causal decision theory, rational decision theory is not in any way vitiated by Rescher's arguments.

1. DR. PSYCHO

In his book *Paradoxes* Nicholas Rescher presents us with a paradoxical case in which he claims that rational decision theory breaks down (Rescher, 2001, pp. 269–272):

Dr. Psycho has given you an apple which you have eaten. He may have poisoned it with Z. He offers you a pill which contains X, fatally poisonous by itself but an antidote to Z, though with minor negative side effects. The doctor poisoned the apple iff he predicted you were going to take the antidote.

You have a life or death choice: do you take the pill or not?

Rescher produces two conflicting calculations of the expected values of taking and not taking the pill. He takes 1 to be the value of life, -1 the value of death, and 1^{-} to be the value of life with the negative side effects of the pill. (The superscript minus sign means "a smidgeon less".)

In his *Analysis 1* (pp. 270–271) he calculates the expected values relative to the correctness of Dr. Psycho's prediction, taking p to be the probability that Dr. Psycho predicts correctly. If the doctor predicts correctly, you survive; otherwise you die. The expected value of each possible action is the sum of the values of its outcomes weighted by their probabilities:

$$EV(\text{take}) = p(1^-) + (1-p)(-1) = 2p^- - 1$$

$$EV(\text{not take}) = p(1) + (1-p)(-1) = 2p - 1$$

So $EV(\text{take}) < EV(\text{not take})$

In his *Analysis 2* (p. 271) Rescher calculates the expected values relative to the state of the apple. If p is the probability that the apple was poisoned, then:

$$EV(\text{take}) = p(1^-) + (1-p)(-1) = 2p^- - 1$$

$$EV(\text{not take}) = p(-1) + (1-p)(1) = -2p + 1$$

Now, if $2p^- - 1 > -2p + 1$, as it will if $p > (1/2)^+$, $EV(\text{take}) > EV(\text{not take})$. So Rescher adds the information that it is (non-trivially) more likely than not that the doctor has poisoned the apple, and because of the discordance between the two analyses in this case concludes that "the standard machinery of expected-value analysis may leave us in the lurch because the probabilities may simply fail to be well-defined quantities" (Rescher, 2001, p. 272). For on his view there is nothing to choose between the two analyses.

2. LOOKING MORE CLOSELY

The "standard machinery of expected-value analysis" is constituted by standard probability theory, in particular, its definitions of the expectation of a random variable. On the whole, classical Bayesian decision theory (for example, that put forward by Jeffrey (1983)) adopts the standard definitions of expectation for the definitions of expected utility, while causal decision theory makes use of reformed definitions. For these reasons we think it is fair to analyse Rescher's paradox from a broadly Bayesian point of view, i.e. from the point of view of a standard evidential decision theory. After we have done so, we shall consider whether a charitable interpretation in terms of causal decision theory is of help to him.

If we determine expected utility in a Bayesian manner, the probabilities of which Rescher speaks are conditional probabilities, simply because what Rescher symbolizes by "EV(take)" is the expected utility on the supposition that you take the pill. When analysed on that basis we find that we get the same result whether we calculate relative to the correctness of the prediction or relative to the state of the apple.

Using "P" for *The apple is poisoned* and "T" for *You take the pill*, let $p_1 = \Pr(P\&T)$, $p_2 = \Pr(P\&{\sim}T)$, $p_3 = \Pr({\sim}P\&{\sim}T)$ and $p_4 = \Pr({\sim}P\&T)$. The prediction is correct and you survive iff $P\&T$ or ${\sim}P\&{\sim}T$, and the prediction is incorrect and you die iff $P\&{\sim}T$ or ${\sim}P\&T$.

The following Venn diagram makes the situation easy to envisage:

So, using C for *The prediction is correct*, we have as the first calculation, done by correctness:

$$\textit{Analysis } 1' \quad \begin{aligned} EV(T) &= \Pr(C|T)^{-} - \Pr({\sim}C|T) \\ &= (p_1/(p_1+p_4))^{-} - p_4/(p_1+p_4) \\ EV({\sim}T) &= \Pr(C|{\sim}T) - \Pr({\sim}C|{\sim}T) \\ &= p_3/(p_2+p_3) - p_2/(p_2+p_3) \end{aligned}$$

($\Pr C|T$) is the conditional probability of C on T, etc.)

If the calculation is done by whether the apple is poisoned, we get:

$$\textit{Analysis } 2' \quad \begin{aligned} EV(T) &= \Pr(P|T)^{-} - \Pr({\sim}P|T) \\ &= (p_1/(p_1+p_4))^{-} - p_4/(p_1+p_4) \\ EV({\sim}T) &= \Pr({\sim}P|{\sim}T) - \Pr(P|{\sim}T) \\ &= p_3/(p_2+p_3) - p_2/(p_2+p_3) \end{aligned}$$

Whichever way the calculation is done, the expected values are the same.

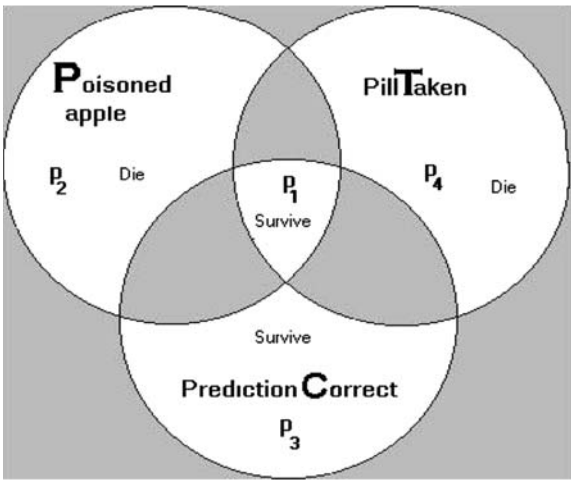

Figure 1.

3. DIAGNOSIS

So what has gone wrong with Rescher's argument?

Rescher's paradox uses three apparently independent premisses about probabilities. He proposes that

(1) The apple is more likely to be poisoned than not: $p_1 + p_2 > p_3 + p_4$.

In Bayesian terms, Rescher's use of the single symbol "p" for the probabilities amounts, in *Analysis 1*, to presuming that:

(2) Dr. Psycho's predictive competence is independent of whether you take the pill:

$$\Pr(C) = \Pr(C|T) = \Pr(C|\sim T), \text{ so } p_1/(p_1 + p_4) = p_3/(p_2 + p_3).$$

In *Analysis 2*, that use amounts to presuming that:

(3) The state of the apple is independent of whether you take the pill:

$$\Pr(P) = \Pr(P|T) = \Pr(P|\sim T), \text{ so } p_1/(p_1 + p_4) = p_2/(p_2 + p_3).$$

He then argues that applying rational decision theory gives us a contradiction. But the premisses are not independent. In fact, they are mutually inconsistent: you can have any two of them, but not all three.

To see that any two are consistent, consider the three cases in this table.

	Consistent pairs of premises			
	1 & 2	1 & 3	2 & 3	
p_1	9/16	9/16	1/4	
p_2	1/16	3/16	1/4	
p_3	3/16	1/16	1/4	
p_4	3/16	3/16	1/4	
$\Pr(C\|T)$	3/4	3/4	1/2	$\Pr(P\|T)$
$\Pr(\sim C\|T)$	1/4	1/4	1/2	$\Pr(\sim P\|T)$
$\Pr(C\|\sim T)$	3/4	1/4	1/2	$\Pr(\sim P\|\sim T)$
$\Pr(\sim C\|\sim T)$	1/4	3/4	1/2	$\Pr(P\|\sim T)$
$\Pr(C)$	3/4	5/8	1/2	
$\Pr(P)$	5/8	3/4	1/2	

For example, (1) is consistent with (2). With the values in the 1&2 column, (1) is true, since the probability that the apple is poisoned $= p_1 + p_2 = 5/8$, which is greater than the probability that it is not, $p_3 + p_4 = 3/8$. And with those values $\Pr(C) = \Pr(C|T) = \Pr(C|{\sim}T) = 3/4$, making (2) true. (1) is consistent with (3), since with the values in the 1&3 column the probability that the apple is poisoned $= p_1 + p_2 = 3/4 = p_1/(p_1 + p_4) = p_2/(p_2 + p_3)$, making both (1) and (3) true. And so on.

To see their mutual inconsistency, consider that

(4) $p_2 = p_3$, from (2) and (3).
(5) $p_1/(p_1 + p_4) = p_2/2p_2 = 1/2$, from (2) and (4).
(6) $p_1 = p_4$, from (5).
(7) $p_1 + p_2 = p_3 + p_4$ from (4) and (6), contradicting (1).

We can see the inconsistency embedded in the assumption that both analyses are compatible like this. In *Analysis 1*, Rescher assumes that $\Pr(C|T)$ and $\Pr(C|{\sim}T)$ are the same (he uses "p" for both). But, if your credence in the apple's being poisoned exceeds 1/2, as he postulates, then you should also think that if you don't take the pill you are more likely to die than to survive: $\Pr({\sim}S|{\sim}T) > \Pr(S|{\sim}T)$ (with S for *You survive*). Since, as is evident from Figure 1, $\Pr({\sim}S|{\sim}T) = \Pr(P|{\sim}T)$ and $\Pr(S|{\sim}T) = \Pr(C|{\sim}T) = \Pr(C|T) = \Pr(P|T)$, it follows that $\Pr(P|{\sim}T) > \Pr(P|T)$. But this contradicts Rescher's assumption in *Analysis 2* that the state of the apple is independent of whether you take the pill (i.e. (3) above).

Likewise, in *Analysis 2*, Rescher assumes that $\Pr(P|T) = \Pr(P|{\sim}T)$ (he uses "p" for both). If the apple is more likely to be poisoned than not, then you are more likely to need the antidote to survive, and so $\Pr(S|T) > \Pr(S|{\sim}T)$. But you survive iff the doctor predicts correctly (see Figure 1). So the chance of a correct prediction if you take the pill will exceed the chance of a correct prediction if you don't: $\Pr(C|T) > \Pr(C|{\sim}T)$. But this contradicts Rescher's assumption in *Analysis 1* that Dr. Psycho's competence is independent of whether you take the pill (i.e. (2) above).

Because Rescher has derived his paradox by assuming the probability premisses are independent, he maintains that the probabilities "fail to be well-defined quantities" – as we have noted. The basis for this claim is that "we have ... an aporetically inconsistent family of theses one of [which] must be jettisoned" (Rescher, 2001, p. 272). Because both of his analyses are equally cogent, it is "effectively impossible" to choose between them, so you might as well toss a coin (p. 272). But since that is unsatisfactory he concludes that the probabilities are not well defined. Now this last manoeuvre shows that Rescher is uncomfortable with where he has ended up. He is evading the strength of his own conclusion. Of course, we know a contradiction can't be true, but we also know that a sound argument leads to the truth. If there really is nothing to choose between the analyses they must rest on consistent assumptions. However, if they rest on consistent assumptions then we have derived a contradiction from those assumptions by use of standard probability theory. Therefore Rescher should locate the source of the incoherence somewhere between the premisses and the conclusion: in standard probability theory. He knows that can't be right, so tries to slip away from it by saying the probabilities aren't well defined.[1]

If the premisses of his argument are not consistent, then it may not be a matter of tossing a coin to choose between *Analyses 1* and *2*. Choosing one may amount to accepting substantial assumptions about the situation which are not consistent with choosing the other. Again, that would make his point entirely trivial. Decision theory is supposed to give different answers relative to different substantial assumptions. In particular, he must be claiming that his probability premisses are consistent and those used in *Analysis 1* do not represent substantial assumptions inconsistent with *Analysis 2*. For only if that is the case can he then make his *reductio* work – roughly:

(i) Suppose the probabilities are well-defined,
(ii) that the probability premisses are consistent
(iii) and the valuation of outcomes is correct.
(iv) Then take the pill and don't take the pill (classical Bayesian decision theory is true, applied to (i), (ii), (iii)).
(v) Absurd, therefore reject premiss (i).

As we now know, his probability premisses are not independent and their inconsistency is provable within probability theory alone. Therefore he has not shown that the supposition of well-defined probabilities is to blame for the inconsistency and cannot claim that "the standard machinery of expected-value analysis" has left us "in the lurch" (Rescher, 2001, p. 272).

Before you have the information that the apple is non-trivially more likely to be poisoned than not, and when you have no information about the doctor's predictive performance and no information relevant to the probability of the apple's being poisoned if you take it, it is not unreasonable to accept (2) and (3). In this case it is rational not to take the pill. But when you learn that it is non-trivially more likely than not that the apple is poisoned, then whether it is rational to take it depends on your subjective probabilities for *P*, *T* and *C*.

We now turn to considering whether interpreting Rescher in terms of causal decision theory can help him. The causal decision theorist rejects the use of conditional probabilities in calculating the expected utility of an act when that act is not causally related to the relevant outcome (cf. Lewis, 1981/1986, p. 314). By using conditional probabilities in such a circumstance, Bayesian decision theory gets it wrong by measuring the value of the act as news rather than what Gibbard and Harper (1978/1981, pp. 156–157) call the "genuine" expected utility. "The 'utility of an act'", they say, "should be its genuine expected efficacy in bringing about states of affairs the agent wants" (p. 168). For Gibbard and Harper, this is to be calculated from the probability of the relevant counterfactuals.[2] In our case, when calculating the expected utility of taking the pill, the relevant probabilities are the probabilities of "If I were to take the pill I would survive" and "If I were to take the pill I would not survive". The probability of the former is plausibly the same as the unconditional probability that the apple has been poisoned, Pr(P), and of the latter that it has not, 1−Pr(P). When calculating the utility of not taking the pill, the relevant probabilities are those of "If I were not to take the pill, I would not survive" and of "If I were not to take the pill, I would survive"; again, the former is Pr(P) and the latter 1−Pr(P). So this amounts to using *Analysis 2*. Gibbard and Harper's approach is consistent with other causal decision theorists, who might reason thus to the use of unconditional probabilities in *Analysis 2*: at the point of choice it is already settled whether the apple you have eaten is poisoned and this cannot be affected by whether you take the pill. Therefore the unconditional probability of the apple being poisoned should be used in calculating choiceworthiness of taking or not taking the pill. This gives us

$$EV(T) = \Pr(P)(1^{-}) + (1 - \Pr(P))(-1) = 2\Pr(P)^{-} - 1$$
$$EV({\sim}T) = \Pr(P)(-1) + (1 - \Pr(P))(1) = 1 - 2\Pr(P)$$

If the apple is more likely to be poisoned than not, then the rational choice is to take the pill.

Given that Rescher does not bring in conditional probabilities in his calculations, interpreting Rescher's *Analysis 2* in this way (i.e. as a causal decision theorist who intends p to be the unconditional probability Pr(P)) might be thought to be the more charitable way of understanding him. It may now appear that Rescher's paradox can be resuscitated. The causal decision theorist is committed to the outcome of *Analysis 2*, and his use of the unconditional probability of the apple's being poisoned means the earlier proof of the equivalence of the two analyses in terms of conditional probabilities (in Section 2) may not hold. So why isn't he also committed to the possibility of an outcome of *Analysis 1* contradicting *Analysis 2*, just as Rescher proposes?

The answer to this is quite simple, provided we remember Rescher's dialectic. The point is to formulate a *reductio* argument against the premiss that the probabilities are well defined by deriving a contradiction of identity: two numbers measuring the same aspect of the same thing are not equal to each other.

(i) $EV(T)_{\text{using probability of correctness}}$ and $EV(T)_{\text{using probability of apple poisoned}}$ both measure the expected utility of taking the pill
(ii) $EV(T)_{\text{on the basis of correctness of prediction}} = a$
(iii) $EV(T)_{\text{on the basis of the state of the apple}} = b$
(iv) but $a \neq b$.

Now we saw in Section 2 that standard probability theory blocks the contradiction by proving (iv) false. In causal decision theory it is no contradiction to discover that $a \neq b$. Because of their reforming definition of expected utility, $EV(T)_{\text{using probability of correctness}}$ measures one aspect of taking the pill and $EV(T)_{\text{using probability of apple poisoned}}$ measures another. In general, it is no fault in causal decision theory that different ways of evaluating the expected value of an act result in distinct values, because those different ways correspond to evaluating the same act with respect to distinct effects.

For example, consider what *Analysis 1* may amount to for Gibbard and Harper's variety of causal decision theory.[3] We take the probabilities to be the unconditional probabilities, $\Pr(C)$ (the probability that Dr. Psycho has predicted correctly) and $1 - \Pr(C)$. To determine the causal efficacy being measured in *Analysis 1* we must determine what counterfactuals correspond. Plausibly, $\Pr(C)$ corresponds to the probability of the counterfactual "if I were to take the pill, the apple would be poisoned", and $1 - \Pr(C)$ to the probability of "if I were to take the pill the apple would not be poisoned". So *Analysis 1* evaluates the expected efficacy of taking the pill with respect to the apple being poisoned. But whether the apple I have eaten is poisoned or not is unaffected by my taking the pill, so $EV(T)$ and $EV(\sim T)$ evaluated on this basis ought to come out the same. However, substituting $\Pr(C)$ in place of p in *Analysis 1* gives

$$EV(T) = 2\Pr(C)^{-} - 1$$

$$EV(\sim T) = 2\Pr(C)^{-} - 1$$

So they come out not the same, but $EV(T) < EV(\sim T)$ by a smidgeon. The smidgeon difference is not altogether satisfactory, but since it is only a smidgeon, we should perhaps allow Rescher this interpretation within causal decision theory.[4]

Consequently, in *Analysis 1*, $EV(T)_{\text{using probability of correctness}}$ measures the efficacy of taking the pill in bringing it about that the apple is poisoned, while in *Analysis 2*, $EV(T)_{\text{using probability of apple poisoned}}$ measures the efficacy of taking the pill in bringing about our survival. Since taking the pill has no effect on the state of the apple but may bring about our survival, and since Gibbard and Harper intend to measure expected efficacy towards a desired outcome, it is not a contradiction but a positive virtue that the act of taking the pill has distinct expected efficacies for these distinct aspects. Hence a causal decision theorist may accept that the outcome of *Analysis 1* contradicts that of *Analysis 2* without thereby being subject to Rescher's *reductio* of the assumption that probabilities are well defined.

So, whether we interpret Rescher's analyses in terms of standard Bayesian reasoning or in terms of causal decision theory, the paradox dissolves. The Dr. Psycho case is not one where decision theory breaks down because the probabilities are necessarily ill-defined. We may have to choose whether we are evidential or causal decision theorists, and, if the latter, make sure we measure the expected efficacy of our act relative to the outcome we are interested in bringing about. But both types of theory apply coherently to the case, and it is possible for both types of theorist to recommend the same action. When they disagree, we prefer the recommendation of causal decision theory.

4. THE NEWCOMB PROBLEM

Rescher offers a similar diagnosis of Newcomb's problem, and uses it as an argument for two-boxing (Rescher, 2001, pp. 264–66).

In the Newcomb problem a Predictor with a very good track record presents you with two boxes, one transparent, in which you can see \$1,000, the other opaque. You may choose either the opaque

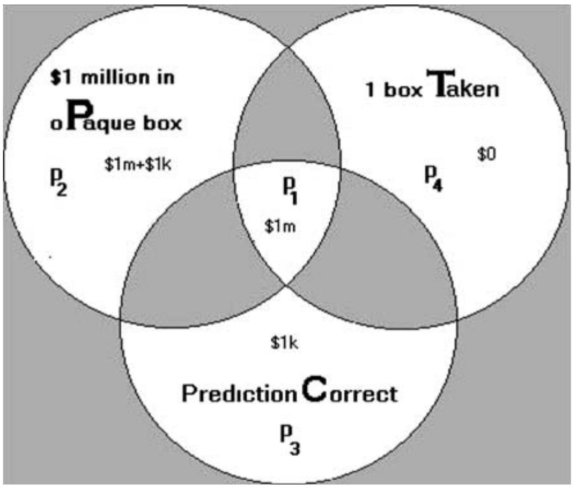

Figure 2.

box alone or both boxes. If the Predictor has predicted that you will take both boxes, he has left the opaque box empty, but if he has predicted that you will take the opaque box alone, he has put $1,000,000 in it. If you want as much money as possible, which choice should you make?

We may adapt the Venn diagram given above in Figure 1 to illustrate the Newcomb situation (Figure 2):

P: there is $1,000,000 in the opaque box.
T: the opaque box only is taken.
∼*T*: both boxes are taken.
C: the Predictor has forecast correctly.

Actually, Rescher introduces a needless complication by considering in addition the option of just taking the transparent box. But no rational chooser would choose the transparent box alone when he could possibly gain more without risking any loss by taking both boxes or follow the temptation to gain much more by taking the opaque box alone, so we shall adapt Rescher to eliminate this possibility. This makes for simplicity and conforms to the standard presentation of Newcomb. It also makes it easier to address his implication that Newcomb's problem and the Dr. Psycho paradox are of a kind, because in Newcomb's problem, too, "the probabilities ... are problematic; they may ... fail to be well-defined, in meaningful quantities" (p. 266). The argument for this claim is, once again, an implicit *reductio* argument.

He calculates the expected value of the alternative choices in two different ways.

First way (item 3, pp. 264–265). If p is the probability that the Predictor has forecast that you pick both boxes, and therefore has put nothing in the opaque box:

$$EV(T) = p(0) + (1-p)(1{,}000{,}000)$$
$$EV(\sim T) = 1{,}000 + EV(T)$$
$$\text{So } EV(\sim T) > EV(T)$$

The expectation from choosing one box is the probability that the Predictor has forecast that choice multiplied by a million; the expectation from choosing two boxes is a thousand more. Taking two boxes therefore has the higher expected value.

Second way (item 4, p. 265). If p is the probability that the Predictor has forecast correctly:

$$EV(T) = p(1{,}000{,}000) + (1-p)(0)$$

since, if the Predictor is wrong, there is nothing in the opaque box

$$EV(\sim T) = 1{,}000 + p(0) + (1-p)(1{,}000{,}000)$$

since, if the Predictor is right, there is nothing in the opaque box.

When p is greater than 0.5005, which it is bound to be given the Predictor's track record, $EV(T)$ will be greater than 500,500 and $EV(\sim T)$ will be less than 500,500. So taking one box has the higher expected value.

As with the Dr. Psycho paradox, Rescher's presentation of Newcomb's problem uses three apparently independent premisses about probabilities. He proposes that

(1′) The Predictor is (significantly) more likely to be right than wrong:

$$p_1 + p_3 > p_2 + p_4.$$

In Bayesian terms, Rescher's use of the single symbol "p" for the probabilities amounts, in *Second Way*, to presuming that:

(2′) The Predictor's competence is the same whether you pick the opaque box alone or both boxes: $\Pr(C) = \Pr(C|T) = \Pr(C|\sim T)$, so $p_1/(p_1 + p_4) = p_3/(p_2 + p_3)$.

In *First Way*, that use amounts to presuming that:

(3′) The probability that the Predictor has forecast that you pick both boxes (and so puts nothing in the opaque box) is the same whether you pick one box or two: $\Pr(\sim P) = \Pr(\sim P|\ T) = \Pr(\sim P|\sim T)$, so $p_4/(p_1+p_4)=p_3/(p_2+p_3)$. (Hence $\Pr(P)= \Pr(P|\ T)=\Pr(P|\sim T)$, so $p_1/(p_1+p_4)=p_2/(p_2+p_3)$, when the reader can see the exact parallel between 1, 2 and 3 above and 1′, 2′ and 3′ here.)

The table above shows once again that any pair of premisses is consistent, and a similar Bayesian analysis to that in Section 2 above shows that the expectations calculated by *First Way* and *Second Way* must be the same.[5] Once more, we obtain a contradiction from the assumption of all three:

(4′) $p_1 = p_4$ from (2′) and (3)′

(5′) $p_3/(p_2 + p_3) = p_4/2p_4 = 1/2$, from (3′) and(4)′

(6′) $p_2 = p_3$, from (5′)

(7′) $p_1 + p_3 = p_2 + p_4$, from (4′) and (6′), contradicting (1′).[6]

The standard analysis of Newcomb's Problem is that it is one in which the principle of dominance conflicts with classical Bayesian choiceworthiness. Rescher, however, thinks that "Newcomb's problem highlights the potential shortcomings that expected value calculations encounter in the presence of problematic probabilities" (p. 266). He argues that his two ways are equally plausible, and so cancel one another out, so that it is reasonable to apply the dominance principle and pick both boxes, since "you will then get whatever there is to be gotten" (p. 264). But the assumptions behind Rescher's two types of calculation are jointly incompatible with the Predictor's being right more often than he is wrong. Consequently, just as in Dr. Psycho, he cannot present the problem as if it were a case in which we had two contradictory ways of calculating the same expected value while keeping all relevant assumptions fixed, since the contradiction is not in the expected values but in his assumptions. He must choose which pair of the three probability premisses he will adopt.

Nor can causal decision theory help Rescher, any more than it did in Dr. Psycho. Obviously so, since causal decision theory was in part developed precisely so that it agrees with dominance on Newcomb's problem.[7] *Analysis 2* and *First Way* are analogous (being based on a state of affairs which will not be causally influenced by your action) and likewise *Analysis 1* and *Second Way* (being based on the correctness of the predictor). We find that *First Way* gets it right just as *Analysis 2* did. For example, for Gibbard and Harper, *First Way* corresponds to measuring the "expected genuine efficacy" of the act in maximizing the money got (Gibbard and Harper, 1978/1981, p. 168).

Since they are formally similar one might think that the Gibbard and Harper interpretation of *Analysis 1* from Dr. Psycho could apply equally to *Second Way*. In that case, Pr(C) would correspond to the probability of the counterfactual "if I were to take one box the opaque box would have $1m" and so on. Because the choice of boxes does not affect the contents of the opaque box, there is no difference between the efficacy of one-boxing and the efficacy of two-boxing with respect to whether there is $1 million in that box: but the two-boxer will win out, because he gets the $1,000 in the transparent box (cf. Gibbard and Harper, 1978/1981, p. 181). So $EV(T)$ and $EV(\sim T)$ evaluated on the basis of Pr(C) ought to come out the same or very close. In Dr. Psycho they do, up to a smidgeon and independently of the value of Pr(C). However, in Newcomb's Problem, if the probability Pr(C) is high then $EV(T)$ on basis of correctness is much higher than $EV(\sim T)$. Thus Rescher's *Second Way* cannot plausibly be interpreted in terms of causal decision theory. Rather, it corresponds to measuring the degree to which "news of the act ought to cheer the agent" (Gibbard and Harper, 1978/1981, p. 168).

So the two ways measure two distinct aspects of the same act – two different kinds of expected utility – and there is no contradiction in distinct aspects having distinct measures. Since maximizing cheering news "commends an irrational policy of managing the news so as to get good news about matters which you have no control over" (Lewis, 1981/1986, p. 305), while maximizing expected efficacy results in managing matters under your control, *Second Way* gets it wrong and *First Way* gets it right.

Just as with Dr. Psycho, with Newcomb's Problem Rescher thinks he has shown that "the probabilities at issue [in the *First* and *Second* Way] are problematic: they may ... fail to be well-defined, meaningful quantities" (p. 266). We have shown that on the one hand, in classical Bayesian theory, conflicting expected utilities of the same act would provide the contradiction his *reductio* required, but they do not arise. On the other hand, in causal decision theory, conflicting expected utilities of the same act measure distinct features of that act. In Dr. Psycho we saw that Rescher's conflicting expected utilities corresponded to distinct causal efficacies of the same act; in Newcomb's Problem they corresponded to the distinction between act as cheering news versus act as causally efficacious. We have to decide which features of the act are relevant to our decision and then use the related expected utility. In neither case do the conflicting expected utilities provide the contradiction his *reductio* required. So while it is possible for probabilities to be ill-defined, there is nothing about Dr. Psycho or Newcomb's Problem that requires that they should be. To avoid Rescher's mistakes we must first ensure that our subjective probabilities are mutually consistent. Secondly, we must distinguish the classical Bayesian expected value of an act, which always conditionalizes probabilities of outcomes on the act, from causal decision theory's expected utility, which restricts the use of conditional probabilities (and also from Bayesian decision theories which have been modified to get round Newcomb's problem). Although our own preference happens to be for causal decision theory, we have not sought to argue here for that approach, but only to show that Rescher's errors are independent of subscription to evidential or causal decision theory. Contrary to Rescher's claim, it is not true that Dr. Psycho and Newcomb's Problem are cases in which the application of the "standard machinery of expected-value analysis" is incoherent.

ACKNOWLEDGEMENTS

We are indebted to the journal's referees for very helpful comments on an earlier draft.

NOTES

[1] What does Rescher need "not well-defined" to mean for his argument? It is not enough that there are different incompatible probability functions on the relevant (sigma algebra of) events or propositions. Nor can it be a matter of ignorance or a case in which objective chances do not exist, since there seems no reason to rule out subjective probabilities. Rather he needs "not well-defined" to mean that no probability function *at all* can be defined on the relevant sigma algebra. But that is necessarily false. A probability space is an ordered triple, $<S,\Sigma,P>$, where S is the sample space of events, Σ is a sigma-algebra of S and P is a measure on Σ which obeys the probability axioms – $P(S)=1$ is sufficient. (For example, see Capinski and Kopp, 1999, p. 46.) A sigma-algebra is a set, Σ, of subsets of a set, S, (so $\Sigma \subseteq \mathbb{P}(S)$) that contains S and $\emptyset$, and is closed under complementation and countable union. The event set here is finite, so its power set is a sigma-algebra. Probabilities are always well definable from any sigma-algebra of events into [0,1]. Therefore there exist well-defined probability functions for Dr. Psycho.

[2] The following interpretations of *Analysis 2*, and (shortly thereafter) of *Analysis 1*, are in terms of Gibbard and Harper's causal decision theory, and they evaluate the

"*U*-utility...a measure of the expected efficacy of an act in bringing about states of affairs the agent desires" (Gibbard and Harper, 1978/1981, p. 168). We are not claiming that they are the uniquely correct *U*-utility interpretations, but only that they make reasonable sense and allow us to explore Rescher's conflicting analyses charitably in terms of a specific causal decision theory.

3 For Gibbard and Harper there is also an interpretation of *Analysis 1* which uses the conditional probabilities (such as $\Pr(C|T)$). On that interpretation, *Analysis 1* amounts to measuring what they call *V*-utility "the degree to which news of [taking the pill] ought to cheer the agent" (ibid.). But for causal decision theorists basing action on measuring cheering news is the wrong thing to do; hence such an interpretation is of no help to Rescher.

4 But see our discussion of the correlative case in Newcomb's Problem below.

5 The expected value calculations, when done properly, yield:

First way:

$$\begin{aligned} EV(T) &= \Pr(\sim P|T).0 + \Pr(P|T).1,000,000 \\ &= 0 + p_1/(p_1 + p_4).1,000,000 \\ EV(\sim T) &= \Pr(\sim P|\sim T).1,000 + \Pr(P|\sim T).1,001,000 \\ &= p_3/(p_2 + p_3).1,000 + p_2/(p_2 + p_3).1,001,000 \end{aligned}$$

Second way:

$$\begin{aligned} EV(T) &= \Pr(C|T).1,000,000 + \Pr(\sim C|T).0 \\ &= p_1/(p_1 + p_4).1,000,000 + 0 \\ EV(\sim T) &= \Pr(C|\sim T).1,000 + \Pr(\sim C|\sim T).1,001,000 \\ &= p_3/(p_2 + p_3).1,000 + p_2/(p_2 + p_3).1,001,000 \end{aligned}$$

Notice that in each case the formula for $EV(\sim T)$, the expected value when both boxes are taken, ensures that the probability of getting the thousand in the transparent box is 1, since $p_3/(p_2+p_3).1000+p_2/(p_2+p_3).1{,}001{,}000=(p_3+p_2)/(p_2+p_3).1000+p_2/(p_2+p_3).1{,}000{,}000$.

6 With (1′) and (2′) we do not have the sort of indeterminate case that Levi (1975, pp. 164–166) drew attention to. Levi pointed out that it didn't follow from (a) the probabilities of choosing one box conditional on a prediction of one box, and of choosing two boxes conditional on a two-box prediction, are both high, that (b) the probabilities of correct predictions conditional on one (two) boxes being chosen are both high: one of the latter could be high while the other was low, consistently with the overall track record of predictions being good. Then the probabilities would fail to be well-defined because the case was incompletely described by (a). However, (1′) and (2′) imply that the probability of a correct prediction is high whether one or two boxes are chosen, which gives us (b) directly. We are not illicitly inferring (b) from (a).

7 Of course, some two-boxers have maintained that Bayesian decision theory can also be applied in such a way that it recommends two-boxing. Cf. Price (1986).

REFERENCES

Capinski, M. and P. E. Kopp: 1999, *Measure, Integral, and Probability*, Springer, London, New York.

Gibbard, A. and W. L. Harper: 1978, 'Counterfactuals and Two Kinds of Expected Utility', in Hooker, Leach, and McLennen (eds.), *Foundations and Applications of Decision Theory*, Reidel, Dordrecht. Reprinted in Harper, Stalnaker, and Pearce (eds.), *Ifs*, (1981, Reidel, Dordrecht), from which page references are taken.

Jeffrey, R. C.: 1983, *The Logic of Decision* (2nd edn), University of Chicago Press, Chicago

Levi, I.: 1975, 'Newcomb's Many Problems', *Theory and Decision* **6**, 161–175.

Lewis, D.: 1981, 'Causal Decision Theory', *Australasian Journal of Philosophy* **59**, pp. 5–30, reprinted in his *Philosophical Papers* II, pp. 305–339 (1986, Oxford University Press: Oxford), from which page references are taken.

Price, H.: 1986, 'Against Causal Decision Theory', *Synthese* **67**, 195–212.

Rescher, N.: 2001, *Paradoxes*, Chicago and La Salle, Ill, Open Court.

Michael Clark
Department of Philosophy
University of Nottingham
University Park, NG7 2RD
Nottingham
UK
E-mail: michael.clark@nottingham.ac.uk

Nicholas Shackel
Oxford Centre for the Science of Mind
University of Oxford
Oxford
UK
E-mail: nicholas.shackel@philosophy.ox.ac.uk

Manuscript submitted 1 February 2004
Final version received 17 May 2005