

The Lion, the ‘Which?’ and the Wardrobe – Reading Lewis as a Closet One-boxer

Huw Price

July 17, 2007

Abstract

Newcomb problems turn on a tension between two principles of choice: roughly, a principle sensitive to the causal features of the relevant situation, and a principle sensitive only to evidential factors. Two-boxers give priority to causal beliefs, and one-boxers to evidential beliefs.

A similar issue can arise when the modality in question is chance, rather than causation. In this case, the conflict is between decision rules based on credences guided solely by chances, and rules based on credences guided by other sorts of probabilistic evidence. Far from excluding cases of the latter kind, Lewis’s Principal Principle explicitly allows for them, in the form of the caveat that credences should only follow beliefs about chances in the absence of “inadmissible evidence”.

In this paper I exhibit a tension in Lewis’s views on these two matters. I present a class of decision problems – actually, I argue, a species of Newcomb problem – in which Lewis’s view of the relevance of inadmissible evidence seems to recommend one-boxing, while his causal decision theory recommends two-boxing. I propose a diagnosis for this dilemma, and suggest a remedy, based on an extension of a proposal due to Ned Hall and others from the case of chance to that of causation. The remedy dissolves many apparent Newcomb problems, and makes one-boxing non-controversial in those that remain.

I Two decision rules

The original Newcomb problem goes something like this. God offers you the contents of an opaque box. Next to the opaque box is a transparent box containing \$1,000. God says, “Take that money, too, if you wish. But I should tell you that it was Satan who chose what to put in the opaque box. His rule is to put in

\$1,000,000 if he predicted that you wouldn't take the extra \$1,000, and nothing if he predicted that you would take it. He gets it right about 99% of the time.”

	Opaque box empty	Opaque box filled
Take one box	\$0 (0.01)	\$1,000,000 (0.99)
Take both boxes	\$1,000 (0.99)	\$1,001,000 (0.01)

Table 1: The standard Newcomb problem (with evidential probabilities)

Famously, this problem brings to a head a conflict between two decision rules. In the original presentation of the problem, these rules were **Dominance** and **Maximise Expected Utility**, but for many purposes it has turned out to be more interesting to represent the disagreement as a clash between two different ways of calculating expected utility (and hence two different versions of the rule **Maximise Expected Utility**).

Maximise Evidentially-grounded Utility ('V-utility'):

$$EV(Act_i) = V(O_1)P_{evidential}(O_1|Act_i) + V(O_2)P_{evidential}(O_2|Act_i)$$

Maximise Causally-grounded Utility ('U-utility'):

$$EU(Act_i) = V(O_1)P_{causal}(O_1|Act_i) + V(O_2)P_{causal}(O_2|Act_i)$$

Here $P_{evidential}(O_j|Act_i)$ is the “purely epistemic” conditional probability of the outcome O_j given the act Act_i ; while $P_{causal}(O_j|Act_i)$ is what we may call the *causal* conditional probability.

It is a simple matter to show that in the decision problem like that described above, these two rules give different recommendations. On the one hand,

$$\begin{aligned} EV(One-box) &= \$0 \times 0.01 + \$1,000,000 \times 0.99 \\ &= \$990,000 \end{aligned}$$

while

$$\begin{aligned} EV(Two-box) &= \$1,000 \times 0.99 + \$1,001,000 \times 0.01 \\ &= \$11,010 \end{aligned}$$

so that the rule **Maximise V-utility** recommends taking only the opaque box. On the other hand,

$$\begin{aligned} EU(Two-box) &= \$1,000 \times \alpha + \$1,001,000 \times (1 - \alpha) \\ &= \$1,000 + EU(One-box) \end{aligned}$$

(where $\alpha = P_{causal}(the\ opaque\ box\ is\ empty)$), so that – by Dominance reasoning, in effect – the rule **Maximise U-utility** recommends taking both boxes.

Philosophers disagree about which of these two decision rules provides the rational strategy. Among the famous “two-boxers” is the lion of twentieth century metaphysics (and my title), David Lewis. Lewis describes the issue as follows:

Some think that in (a suitable version of) Newcomb’s problem, it is rational to take only one box. These one-boxers think of the situation as a choice between a million and a thousand. They are convinced by indicative conditionals: if I take one box I will be a millionaire, but if I take both boxes I will not. Their conception of rationality may be called *V-rationality*; they deem it rational to maximize *V*, that being a kind of expected utility defined in entirely non-causal terms. Their decision theory is that of Jeffrey [(1965)].

Others, and I for one, think it rational to take both boxes. We two-boxers think that whether the million already awaits us or not, we have no choice between taking it and leaving it. We are convinced by counterfactual conditionals: If I took only one box, I would be poorer by a thousand than I will be after taking both. . . . Our conception of rationality is *U-rationality*; we favor maximizing *U*, a kind of expected utility defined in terms of causal dependence as well as credence and value. Our decision theory is that of Gibbard and Harper [(1978)] or something similar. (Lewis 1981b, 377)

Elsewhere, Lewis affirms his commitment to two-boxing like this:

[S]ome—I, for one—who discuss Newcomb’s Problem think it is rational to take the thousand no matter how reliable the predictive process may be. Our reason is that one thereby gets a thousand more than he would if he declined, since he would get his million or not regardless of whether he took his thousand. (Lewis 1979, 240)

My aim in this paper is to call attention to an apparent tension between this aspect of Lewis’s views, on the one hand, and his professed position concerning chance, evidence and rational credence, on the other. In his discussion of the Principal Principle, Lewis allows that chance does not provide an exceptionless constraint on rational credence: on the contrary, he holds, an agent who has access to “inadmissible information” may be rational to allow her credences to be guided by that information, rather than by her knowledge of the relevant objective chances. I want to argue that this amounts to endorsing one-boxing, in a particular

class of Newcomb problems; and that there seems to be no principled way of distinguishing these Newcomb problems from Newcomb problems in general. If I am right, then Lewis's commitments about the two matters are in tension with one another.

In the final section of the paper I consider a resolution of this tension which would extend a suggestion due to Ned Hall concerning the Principal Principle. Hall argues that Lewis's qualification of the Principal Principle to deal with "inadmissible" information is unnecessary and undesirable. Better, Hall argues, to say that there is no such thing as inadmissible information: properly understood, chance tracks expert credence in such a way that such cases simply don't arise. If Hall is right, the effect is that in my chance-based Newcomb problems, chance-grounded and evidence-grounded reasoning must coincide – both recommend one-boxing.

I want to point out that Hall's move is analogous to an option that a number of people – me, for one – have found attractive in standard Newcomb cases, viz., that of arguing that where evidential reasoning really does recommend one-boxing, so too does causal reasoning, properly understood. I think that this approach can be seen as arguing that causation is best understood as a codification of an "expert function" for a deliberative agent, in the way that Hall treats chance as a codification of an expert function for a betting agent – an *evidential* agent, in both cases.

In neither case (for chance nor causation) is Hall's proposal compulsory, in my view. In either case, we might have grounds – e.g., perhaps, from physics – to postulate a modal notion which might in principle float free of rational agency, in unusual cases. But in this eventuality, once we recognise it for what it is, it will be immediate that the rational policy is to one-box, in the corresponding Newcomb problems. The apparent force of the Newcomb puzzles derives from the fact that we have allowed our modal and evidential notions to drift apart in this way, without being aware of the diagnosis. Once we understand these facts, we can either eliminate these cases altogether, via Hall's prescription and its causal analogue, or we can choose to live with them. But in the latter case the right option is the one that Lewis grasped in the case of chance: rationality and modal metaphysics part company, and the rational choice is to one-box.

2 A chancy Newcomb problem?

On the face of it, Newcomb problems turn on a conflict between *causal beliefs* and *evidential beliefs*. Hence it is natural to ask whether the same kind of conflict can arise for other kinds of objective modality. In particular, can it arise for chance?

It is easy to see that it can. Suppose God offers you the following options on the toss of what He assures you is a fair coin. Obviously, it is rational to choose to bet Heads.

	Heads	Tails
Bet Heads	\$100	\$0
Bet Tails	\$0	\$50

Table 2: A free lunch?

But now suppose that Satan informs you that although God told you the truth when He told you that it is a fair coin, he didn't tell you the whole truth. The whole truth, Satan reveals, is that on occasions on which you yourself bet on the coin, it usually (say, about 99% of the time) comes up Tails. What strategy is rational at this point? Should you assess your expected return in the light of the objective chances? Or should you avail yourself of Satan's further information? Call this the chancy Newcomb problem, or Chewcomb problem, for short.

	Heads	Tails
Bet Heads	\$100 (0.01)	\$0 (0.99)
Bet Tails	\$0 (0.01)	\$50 (0.99)

Table 3: The Chewcomb problem (with Satanic evidential probabilities)

What is the rational policy in this case? Presumably we should use our rational credences to calculate the expected values of the available actions, but there are two views as to what the rational credences are. According to one view, the rational credences are given to us by our knowledge of the objective chances. In this case, Satan's contribution makes no difference to the rational expected utility, and we should bet Heads, as before. According to the other view, our rational credence should take Satan's additional information into account, in which case (as it easy to calculate), our rational expected return is \$1 if we choose Heads and \$49.50 if we choose Tails.

Which policy should we choose? If we turn for guidance to the masters, we find that Lewis's discussion of the constraint that a theory of chance properly places on rational credence – the discussion in which he formulates the Principal Principle – seems to recommend the second policy in such a case. What it explicitly recommends – the point of Lewis's exclusion to the Principal Principle for the case in which one take oneself to have “inadmissible evidence” – is that in such

a case one's rational credences follow one's beliefs about the new evidence, rather than remaining constrained by one's theory of chance. Ned Hall's exposition of Lewis's view makes this clear:

Example: we have, on thousands of occasions before this one, consulted the Oracle about what the chancy future would bring—and every time, her predictions have been vindicated, in minute detail. This time, she tells us that the coin will land heads. All of our information is purely historical, concerning only the record of her past successes, plus her most recent prediction. What should we believe about the outcome of the toss, on the supposition that it has a 50% chance of landing heads? Answer: we should be certain, or nearly so, that it will land heads, favouring the reliable words of the Oracle over the guidance of objective chance. We should treat the Oracle as a crystal ball—even though she might merely be lucky, and even though the evidence guiding our opinions contains no information directly about the future. (Hall 1994, 508–509)¹

Or from Lewis himself:

The fatal move that led from Humeanism to contradiction is no better than the *obvious blunder*:

$C(\text{the coin will fall heads/it is fair and will fall heads in 99 of the next 100 tosses}) = 1/2$

or even

$C(\text{the coin will fall heads/it is fair and it will fall heads}) = 1/2$.
(Lewis 1994, 485, my emphasis)

Thus Lewis takes it for granted that someone who takes themselves to have inadmissible evidence should base their credences – and hence presumably their betting behaviour – on that evidence, rather than on their beliefs about the relevant chances. In the present case, then, this suggests that we should assess our options in the Chewcomb problem simply by replacing credences based on chances with credences based on the 'Satanic' evidential probabilities.

However, it turns out that the Chewcomb problem is just a Newcomb problem, in (light) disguise; and this policy – the second policy, above – amounts to one-boxing, when the disguise is removed. So Lewis's qualification to the Principal Principle thus seems to be in tension with his allegiance to two-boxing, at least in certain kinds of Newcomb problem.

¹We'll return to Hall's own view below.

In §3 below I'll introduce a new Chewcomb problem, which makes more explicit the fact that Chewcomb problems are a species of Newcomb problem. But we can already highlight the tension in Lewis's position in the present case, by considering this decision problem in the light of Lewis's own (1981a) formulation of causal decision theory. Lewis's version of CDT depends on a partition $K = \{K_0, K_1, \dots\}$ of "dependency hypotheses", each of which specifies how what an agent cares about depends on what she does. The expected U -utility of an act A , is then calculated as a sum of the values of each option allowed by this partition, weighted by the corresponding unconditional probabilities:

$$U(A) = \sum_i P(K_i)V(A \& K_i).$$

Thus in a standard Newcomb problem, where it is specified that the agent has no causal influence over the contents of the opaque box, the dependency hypotheses may be taken to be:

- K_0 : The opaque box is empty.
- K_1 : The opaque box contains \$1,000,000

We then calculate the causal utilities for taking both both boxes and taking one box as follows:

$$\begin{aligned} U(Two-box) &= P(K_0)V(Two-box \& K_0) + P(K_1)V(Two-box \& K_1) \\ U(One-box) &= P(K_0)V(One-box \& K_0) + P(K_1)V(One-box \& K_1). \end{aligned}$$

By dominance reasoning, the result is that $U(Two-box) > U(One-box)$.

If we wish to apply this framework to the Chewcomb problem, the first issue is what we should take the the dependency hypotheses to be. If we take it that because the outcome (H or T) is the result of a toss of fair coin – whose behaviour isn't influenced by the way we choose to bet, presumably – then the relevant dependency hypotheses are:

- K_H : The coin lands Heads
- K_T : The coin lands Tails.

The next issue concerns the probabilities $P(K_H)$ and $P(K_T)$. Lewis himself stresses that if CDT is to remain distinct from EDT, we need to use unconditional probabilities at this point, not probabilities conditional on action:

It is essential to define utility as we did using the unconditional credences $C(K)$ of dependency hypotheses, not their conditional credence $C(K|A)$. If the two differ, any difference expresses exactly that news-bearing aspect of the options that we meant to suppress. Had we used the conditional credences, we would have arrived at nothing different from V . (1981a, 12)

In the Chewcomb problem, however, the information we learn from Satan doesn't tell us that $P(K_H) = .01$, for example, but only that $P(K_H|Bet\ T) = .01$. If this isn't already clear, we could easily modify the example to make it explicit. As I set things up above, Satan's information concerns the class of cases in which the agent bets at all (either on H or T) – and it might be argued this yields an unconditional probability for an agent who already knows herself to be taking part in the game. However, if we specify that the agent has a third choice – viz., not to bet at all – then the situation is unambiguously as I claimed. In this case, Satan's information certainly concerns a “news-bearing aspect” (as Lewis puts it) of the act of choosing to bet rather than not to bet. Accordingly, Lewis's CDT then seems to require that we use $P(K_H) = P(K_T) = 0.5$ for calculating $U(Bet\ H)$, $U(Bet\ T)$ and $U(No\ bet)$, for there are no other *unconditional* probabilities available. The upshot is that CDT recommends the first of the two policies we distinguished above: it recommends betting on H , on the grounds (i) that H pays a higher return, and (ii) that K_H and K_T are taken to be equally likely, *in the only sense this decision theory allows to be relevant*.

Thus we have an apparent conflict between Lewis's CDT, on the one hand, and (and at least a naive application of) his view of the relevance of inadmissible information, on the other. To prevent an outright inconsistency, cases of this kind thus need “further work” at one end or other. Either we need to make CDT give a different result, or we need to explain why the injunction that rational credence should take into account inadmissible evidence doesn't yield a recommendation that conflicts with CDT.

I offer the following diagnosis of the difficulty. Newcomb problems are decision problems in evidential policies seem to give different recommendations from causal policies, and CDT is the decision theory that cleaves to the causal side of the tracks. Cases of inadmissible evidence are cases in which chance-based credences lead to different recommendations from (total-)evidence-based credences, and Lewis takes it for granted that the rational policy is to cleave to the evidential side of the tracks. Chewcomb problems are decision problems in which both these things happen at once. It follows that the two kinds of cleaving are liable to yield different recommendations in these cases. At least, they are liable to do so as long as our causal judgements cleave to our judgements about objective chance . . . but to give that up – to allow, instead, that causal judgements might properly follow the “merely evidential” path – would be to abolish the very distinction on which Newcomb problems rely (or at least to move in that direction).

Lewis himself seems to have been aware that cases like the Chewcomb problem lead to special difficulties. In the paper in which he presents his own version of CDT, he compares it to several earlier proposals by other writers. One of these

earlier proposals had been presented in unpublished work by Sobel,² and Lewis's discussion of Sobel's theory closes with the following remarks:

But [Sobel's] reservations, which would carry over to our version, entirely concern *the extraordinary case of an agent who thinks he may somehow have foreknowledge of the outcomes of chance processes*. Sobel gives no reason, and I know of none, to doubt either version of the thesis except in extraordinary cases of that sort. Then if we assume the thesis, it seems that we are only setting aside some very special cases – cases about which I, at least, have no firm views. (I think them much more problematic for decision theory than the Newcomb problems.) So far as the remaining cases are concerned, it is satisfactory to introduce defined dependency hypotheses into Sobel's theory and thereby render it equivalent to mine. (1981a, 18, my emphasis)

However, I don't know whether Lewis saw the dilemma that these cases pose for his own views – a dilemma that turns on a tension between his attitude to the relation between causal judgements and evidential judgements, on the one hand, and chance judgements and evidential judgements, on the other. Nor do I know whether he saw the fact that serves to highlight this dilemma, viz., that these cases are themselves a species of Newcomb problem.

3 Making the analogy closer

I referred to the Chewcomb problem above as a Newcomb problem in light disguise. Let's remove the disguise. Suppose that God offers you the contents of an opaque box, to be collected tomorrow. He informs you that the box will then contain \$0 if a fair coin to be tossed at midnight lands Heads, and \$1,000,000 if it lands Tails. Next to it is a transparent box, containing \$1,000. God says, "You can have that money, too, if you like." At this point Satan whispers in your ear, saying, "Pssst! It's definitely a fair coin, but my crystal ball tells me that in 99% of cases in which people choose to one-box in this game, the coin actually lands Tails. When people choose to two-box, it is only 1% Tails. Greed is good, my friend . . . though try telling that to 'His Excellency' over there!"

	Heads	Tails
Take one box	\$0 (0.01)	\$1,000,000 (0.99)
Take two boxes	\$1,000 (0.99)	\$1,001,000 (0.01)

Table 3: A better free lunch?

²To which at present I have not had access.

Assuming that you're convinced that both God and Satan are telling the truth, what's the rational decision policy in this case? Here the evidential and causal recommendations seem to be *exactly* as in the original Newcomb problem, as I presented it above. Your action won't have any *causal* influence on whether there is money in the opaque box. How could it do so, when that is determined by the result of a toss of a fair coin? To say that the chances are 50/50 is surely to say that *nothing* prior to the toss can have a causal effect on the outcome.³ Yet you have (or, what's relevant here, you *believe* yourself to have) evidence of a strong *evidential* correlation between your action and the result of the coin toss, such that you are much more likely to get rich if you one-box.

We've seen that in formulating the Principal Principle, Lewis explicitly allows an exception to the requirement that rational credences should align with beliefs about objective chances. The exception is to deal with cases of "inadmissible evidence" – the kind of evidence provided by crystal balls and the like. (See the Hall quote above.) Thus he appears to allow that in a case of the sort we have just described, it is rational to allow one's credences to be guided by the information provided by Satan. Yet we now see that this is just to recommend one-boxing, in what is formally just another Newcomb problem.⁴

4 One-box via the Hall way?

Ned Hall (1994, 2004) recommends that we replace Lewis's Principal Principle with a new principle, requiring that rational credences track *conditional* chances: chances *given our evidence*. On the face of it, this may seem to eliminate the problem cases. What matters isn't simply the chance of the coin coming up Tails, but the chance of it doing so given the extra information that Satan has whispered in our ear.⁵ On the face of it, then, this seems to be irenic resolution of the dilemma posed by the Chewcomb problems: they are pseudo-problems, artifacts of a mistaken rule for aligning credence with one's beliefs about chance: in one sense a victory for evidentialism, but a face-saving victory for the evidentialists' opponents, too, in that it maintains that they never had any good reason to disagree.

³This seems to block the move of trying to distinguish between this case and the original Newcomb problem by arguing that in this case, it is less clear that the outcome is not counterfactually dependent on one's choice. To admit that there might be such counterfactual dependence would be to admit that there might be causation, too.

⁴We could make the analogy with the original Newcomb problem even tighter, by specifying that the coin toss (which determines whether the opaque box contains \$1,000,000) has *already taken place*, when the agent makes her choice. (She, of course, does not yet have access to the result.)

⁵Or the information *that* he has done so, perhaps.

Like many irenic proposals, however, this one is a little too good to be true. To see this, we only have to imagine a proponent of a view of chance according to which it makes no difference what Satan whispers in one's ear: the real meta-physical chance of a fair coin's landing Tails is independent of such supernatural vocalisations (our objector insists), and so the shift to conditional chances makes no difference.⁶ In such a case, it remains an issue whether rational (conditional) credence should be guided by chance alone, or by other kinds of information.

I think that the real relevance of Hall's treatment of the Principal Principle to our present concerns lies in a different feature. Drawing (as he notes) on earlier proposals by Gaifman (1988) and van Fraassen (1989, 197–201), Hall suggests that "chance plays the role of an expert":

Why should chance guide credence? Because—as far as its *epistemic* role is concerned—chance is like an expert in whose opinions about the world we have complete confidence. (1994, 511)

In his (2004) paper Hall elaborates on this idea by distinguishing two kinds of expert—roughly, the kind of expert (a "database-expert", as Hall puts it) who simply knows a lot, and

the kind of expert who earns that status not because she is so well-informed, but rather because she is extremely good at *evaluating the relevance* (to claims drawn from the given subject matter) *of different possible bits of evidence*. (2004, 100)

"Let us call the second kind an analyst-expert," Hall continues. "[S]he earns her epistemic status because she is particularly good at evaluating the relevance of one proposition to another." (2004, 100) Hall takes chance to be the second kind of expert: "I claim that *chance is an analyst-expert*." (2004, 101)

Thus for Hall it simply becomes a matter of definition that chance and reasonable credence cannot come apart, once we have conditionalised on all the available evidence (including, in particular, what Lewis treats as "inadmissible" evidence). And it is this stipulation, rather than the conditionalisation move itself, that finally ensures that there cannot be a genuine Chewcomb problem – a genuine case in which chance and evidential reasoning come into conflict. Failure to conditionalise certainly gives rise to one class of (apparent) Chewcomb problems, and Hall's diagnosis correctly eliminates those. As we observed a moment ago, however, the conditionalisation move does not deal with a second class of potential Chewcomb

⁶I think that the possibility of this objection is obscured in Hall's (1994) discussion of "crystal balls" by his failure to treat the case in which the ball's prediction is itself probabilistic in nature, as in my Satanic example.

problems, viz., those in which a metaphysical view of chance simply “disconnects” from expert credence (conditional on all the evidence, in each case).

I’ve stressed this point because it is the latter aspect of Hall’s view that seems to me analogous to (what I find) an attractive resolution of the original Newcomb case. In Hall’s terminology, it is the view that causal judgments should be regarded as the judgments of experts about effective strategies, where these are a matter of maximising conditional V-utility (properly assessed from the agent’s point of view). Once again, the effect is to support one-boxing, but to see this as what maximising U-utility recommends, too, when causal dependence is seen for what it really is.

We can develop the analogy very directly using our Chewcomb example. The original argument for the causal independence of the outcome (Heads or Tails) on our choice of one or two boxes was that in either case, the chance of Heads and Tails remains the same. (How could we exert a causal influence, we reasoned, if we couldn’t influence the chances of the outcomes concerned?) According to Hall’s prescription, however, the conditional chance of Tails given one-boxing *is* higher than conditional chance of Tails given two-boxing (and higher than the conditional chance of Heads given one-boxing). And since we can choose which antecedent to “actualise” in these various conditional chances, we can also influence the resulting unconditional chance, in the obvious sense. Thus the intuitive connection between chance and causation now works in the opposite direction. It suggests that we do have influence and causation – in particular, causal dependence of States on Acts – in the sense of those terms that now seems appropriate, given that chance is to be understood as an expert function.

In neither case is Hall’s proposal or its causal analogue compulsory, in my view. In either case (for chance or for causation) we might have grounds – e.g., perhaps, from physics – to postulate a modal notion which could drift apart from expert credence and strategy, in unusual cases. But in this eventuality, once we recognise it for what it is, it will be immediate that the rational policy is to one-box, in the corresponding Newcomb problems.

The apparent force of the Newcomb puzzles seems to derive from the fact that we have allowed our modal and evidential notions to drift apart in this way, *without being aware of the diagnosis*. Once we understand these facts, we can either eliminate these cases altogether, via Hall’s prescription and its causal analogue, or we can choose to live with them. But in the latter case there’s no real dilemma. The right option, trivially, is the one that Lewis grasped in the case of chance: rationality and modal metaphysics part company (and rationality follows rationality – what else? – rather than metaphysics). There may be other puzzles and surprises in the vicinity: puzzlement that rationality and some kind of metaphysics do not keep step, in some sense, in the way that we have come to expect; or puzzlement

about how there can be an expert function at all, of the chance or the causation variety, in a world of a certain kind.⁷ But these are not the original puzzle of the Newcomb problem, which seems to evaporate from this perspective, along with the problem of inadmissible evidence.⁸

Bibliography

- Gaifman, H. 1988: 'A theory of higher order probabilities', in B. Skyrms, et. al. (eds.), *Causality, Chance, and Choice* (Dordrecht: Reidel).
- Gibbard, Allan and Harper, William 1978: 'Counterfactuals and two kinds of expected utility', in C. A. Hooker, J. J. Leach, and E. F. McClennen, eds., *Foundations and Applications of Decision Theory*, Vol. 1 (Dordrecht: Reidel), 125–162.
- Hall, Ned 1994: 'Correcting the guide to objective chance', *Mind* 103, 505–517.
- Hall, Ned 2004: 'Two mistakes about credence and chance', *Australasian Journal of Philosophy* 82, 93–111.
- Jeffrey, Richard C. 1965: *The Logic of Decision* (New York: McGraw-Hill).
- Lewis, David 1979: 'Prisoners' dilemma is a Newcomb problem', *Philosophy and Public Affairs* 8, 235–240.
- Lewis, David 1981a: 'Causal decision theory', *Australasian Journal of Philosophy* 59, 5–30.
- Lewis, David 1981b: "'Why ain'cha rich?'"', *Noûs* 15, 377–380.
- Lewis, David 1994: 'Humean supervenience debugged', *Mind* 103, 473–490.
- Price, H. 1986: 'Against causal decision theory', *Synthese* 67, 195–212.
- Price, H. 1991: 'Agency and probabilistic causality', *British Journal for the Philosophy of Science* 42, 15–76.
- van Fraassen, Bas 1989: *Laws and Symmetry* (Oxford: Oxford University Press).

⁷Another very interesting kind of puzzle that certainly survives these conclusions concerns the nature of the rational expert's recommendation, in difficult cases. Some people – I, for one (Price 1986, 1991) – have argued that while we do make a mistake if we one-box in the so-called medical Newcomb problems, the mistake is a failure of evidential rationality – a failure to assess the evidential probabilities correctly, from the agent's own point of view – rather than a failure to use the correct decision rule. This diagnosis, and the intriguing issue about rationality it tries to address, both seem untouched by the present conclusions,

⁸Many thanks to Jossi Berkovitz, whose current work on Newcomb problems prompted me to ask the question at the beginning of §2; to Rachael Briggs, for discussion of these ideas, and especially for suggesting that I look at Hall's response to Lewis on chance and inadmissible information; and to Adam Elga, for encouraging me to get to grips with Lewis's CDT in §2. Thanks also to Peter Menzies, Nick Smith, Mark Colyvan and an audience at the ANU–Sydney–Kyoto Probability Workshop in June, 2007, for many helpful comments.