Causation, Chance and the Rational Significance of Supernatural Evidence

Huw Price

January 2, 2012

Abstract

In 'A Subjectivist's Guide to Objective Chance,' David Lewis says that he is "led to wonder whether anyone but a subjectivist is in a position to understand objective chance." The present paper aims to motivate this same Lewisean attitude, and a similar degree of modest subjectivism, with respect to objective causation.

The paper begins with Newcomb problems, which turn on an apparent 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. I note that 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."

I then exhibit a tension in Lewis's views on these two matters, by presenting a class of decision problems – some of them themselves Newcomb problems – in which Lewis's view of the relevance of inadmissible evidence seems in tension with his causal decision theory. I offer a diagnosis for this dilemma, and propose 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 suggests a new view of the relation between causal decision theory and evidential decision theory, viz., that they stand to each other much as chance stands to credence, being objective and subjective faces of the same practical coin. This has much the same metaphysical benefits as Lewis's own view of chance, and also throws interesting new light on Newcomb problems, providing an irenic resolution of the apparent disagreement between causal and evidential decision rules.

1 Introduction

Contemporary metaphysicians are familiar with the claim that an adequate account of chance, or *objective* probability in general, needs to attend to the *subjective* face of probability, too. That is, it needs to make sense of the link between chance, on the one hand, and *credence,* on the other. The classic expression of this viewpoint is to be found in David Lewis's 'A Subjectivist's Guide to Objective Chance.' Lewis's 'Principal Principle,' introduced in that paper, aims to formalize this connection between the objective and subjective faces of probability by characterizing the constraint that knowledge of chance properly imposes on rational credence.[1]

Lewis himself takes it to be definitive of chance that it plays this role in guiding credence, or subjective probability. "Indeed," as he puts it, he is "led to wonder whether anyone *but* a subjectivist is in a position to understand objective chance!" (1986 [1980], 84) Returning to this theme in later work, he criticizes rival approaches on the grounds that they pay insufficient attention to this connection between chance and credence: "Don't call any alleged feature of reality 'chance' unless you've already shown that you have something, knowledge of which could constrain rational credence," he says. (1994, 484)

In this paper, I want to propose an analogous view of causation. To paraphrase Lewis, I want to make it seem reasonable to wonder whether anyone *but* a subjectivist is in a position to understand objective causation. (I want to show, also, that there is a tension in Lewis's views, in the light of the fact that he endorses this degree of subjectivism for chance but not for causation.) By 'subjectivism' here, I mean what I take Lewis to mean in the case of chance: the view that an adequate account of the objective modal notion in question (in my case, causation) necessarily begins with, or at least pays very close attention to, a psychological correlate of this modal notion.

At first sight, it is far from clear that there is a conceptual space for a proposal of this kind, in the case of causation. In particular, it is unclear what could

The beginnings of this paper were much indebted to Joseph Berkovitz, whose work on Newcomb problems prompted me to ask the question at the beginning of §3; and to Rachael Briggs, who suggested the link with Hall's response to Lewis on chance and inadmissible information (and who also gave me helpful comments at later stages). I am also very grateful to Arif Ahmed, Helen Beebee, Steve Campbell, Mark Colyvan, Andy Egan, Adam Elga, Alan Hájek, Chris Hitchcock, Jenann Ismael, Jim Joyce, Peter Menzies, Daniel Nolan, Wlodek Rabinowicz, Brian Skyrms, Nicholas J. J. Smith, Howard Sobel and Hong Zhou, and to audiences at the University of Sydney, ANU, MIT, Michigan and Oxford, for much discussion and many helpful comments, at various points. I am also indebted to several anonymous referees for *Philosophical Review;* and I am grateful to the Australian Research Council and the University of Sydney, for research support.

[1] Lewis's original formulation of the Principal Principle, which will do for our purposes, runs as follows:

> Let C be any reasonable initial credence function. Let t be any time. Let x be any real number in the unit interval. Let X be the proposition that the chance, at time t, of A's holding equals x. Let E be any proposition compatible with X that is admissible at time t. Then $C(A|XE) = x$. (1986 [1980], 87)

The notion Lewis invokes here of the 'admissibility' of the evidential proposition E will play a crucial role in what follows, and I shall return to it shortly.

comprise the subjective face of objective causation – the required psychological correlate. In the case of chance, the notion of credence, and its relation to rational action, seem relatively unproblematic. They come as a package, as Lewis himself notes:

> There is no great puzzle about why credence should be a guide to life. Roughly speaking, what makes it be so that a certain credence function is your credence function is the very fact that you are disposed to act in more or less the ways that it rationalizes. . . . No wonder your credence function tends to guide your life. If its doing so did not accord to some considerable extent with your dispositions to act, then it would not be your credence function. You would have some other credence function, or none. (1986 [1980], 108–109)

Thus we might say that credence is defined in terms of its role in a subjective decision theory (SDT) of the kind we have from Savage (1954); and that because that part of the story is unproblematic, we are then free to characterize chance in terms of credence, as Lewis proposes.

In the case of causation, however, we seem to lack an uncontroversial analogue of SDT. On the contrary, subjectivist evidential decision theory (EDT), such as that of Jeffrey (1965) – though initially motivated by the desire to extend Savage's SDT to the case in which outcomes may depend on an agent's choices – is widely held to deliver irrational recommendations in certain cases, such as Newcomb problems. Many writers, including Lewis himself, believe that EDT therefore needs to be supplanted, or at least supplemented, by a *causal* decision theory (CDT) – which, these writers claim, does give the right prescription in the difficult cases. Other writers disagree, and argue that EDT does give the correct recommendations, in the cases in question.[2] But both sides accept that there is a substantial difference: that the relations of causal dependence invoked by CDT simply *differ,* in the crucial cases, from the relations of evidential dependence invoked EDT. From both sides of the dispute, then, there seems to be no scope for the possibility that causal dependence might align with subjective evidential dependence – that the latter might comprise the subjective face of the former, as we put it above – in the way that Lewis takes chance to align with credence.

My main goal in this paper is to show that both sides are mistaken. The subjectivist option exists in the case of causation, just as for chance, and has similar metaphysical advantages. In particular, it easily explains the *practical* relevance of the modal facts in question – facts about chance in Lewis's case, about causation in mine – where less subjectivist approaches have trouble. (It is in this sense that, echoing Lewis, I think it is reasonable to wonder whether anyone but a subjectivist is in a position to understand objective causation.)

The subjectivist option also throws interesting new light on the dispute between EDT and CDT, suggesting a resolution of a decades-old stalemate at the heart of that debate. I provide a brief introduction to this dispute, along with

[2]Readers unfamiliar with these issues will find a brief overview in §2 below.

other preliminaries, in the next section. (Readers unfamiliar with this literature may safely skip the remainder of this section.) But to flag my course for those who do know the territory, it will be to argue that both sides in the dispute between EDT and CDT have missed an irenic proposal, which brings the two decision theories into sufficient proximity to allow a subjectivist analogue of the case of chance and credence. What is the secret of this irenic solution? In a sense, simply the subjectivist proposal itself. *By* treating it as definitive of causation that causal beliefs rationally constrain the conditional credences needed by EDT, we bring EDT and CDT into sufficient alignment so that they can be seen as the subjective and objective faces of the same practical coin (just as credence and chance are, in Lewis's own account).

For the moment, this suggestion is likely to strike experts on both sides of the EDT/CDT dispute as little more than a verbal trick – a proposal simply to change what we mean by the term *causation,* in effect. Surely – they will object – the differences between CDT and EDT will still be as stark as ever, whatever we *call* the decision theories in question? Changing the labels isn't going to change irrational acts into rational acts, or vice versa! So I have a lot of work to do, to establish that there is an option here worth taking seriously.

Lewis's views on chance and credence will provide a crucial point of reference, throughout the paper. They also provide a point of entry, in the following sense. As I noted, I will be arguing that in Lewis's own case, there is a significant dissonance between his views of chance on the one hand and causation on the other, consisting in the fact that he is more subjectivist about the former than the latter. Interestingly, this dissonance shows up in a sharp form in his views about decision theory, if we compare his recommendations about Newcomb problems, on the one hand, and certain exclusions to the Principal Principle, on the other. The Principal Principle (see fn. 1) describes the way in which rational credence tracks an agent's beliefs about objective chance, in many circumstances. The exceptions – the cases in which rational credence is not constrained by beliefs about chance in this way – are those in which agents take themselves to have what Lewis calls 'inadmissible information': direct epistemic access to the outcomes of chance processes. As Lewis points out, some cases of inadmissible evidence are commonplace. If a coin toss has already taken place, we may have a much better guide to its outcome than the knowledge that it was a fair coin. But Lewis allows that there might also be *abnormal* circumstances in which agents take themselves to have inadmissible information – e.g., the kind of direct epistemic access to the outcomes of chance processes traditionally associated with seers, crystal balls, and the like. As in the normal cases, Lewis takes it for granted that if we believe that we have such information, it properly overrules the Principal Principle, in determining our rational credences. (Again, readers unfamiliar with this discussion will find further details in the next section.) As I shall show, however, this recommendation turns out to be in tension with Lewis's opposition to EDT in certain Newcomb-like problems.

This tension in Lewis's views about decision theory is of some interest in its own right. Here, however, it functions as a path into the broader concerns

of the paper, and a motivation for my main proposal. I suggest a remedy for the tension that rests on extending to the case of causation a suggestion that Ned Hall makes for chance, in his own discussion of Lewis's Principal Principle. Roughly, Hall's suggestion is that we make *all* evidence admissible by stipulation. Cases of crystal balls thus become cases in which the chances are weird, not cases in which rational credence ignores chance, in some weird way. My proposal is that we say the same about causation in *some* Newcomb problems – especially those that rely on similar kinds of 'supernatural' sources of information as Lewis's *abnormal* cases of inadmissible information. My proposal is that we say that these are cases in which causation is strange, not cases in which causal and evidential dependence come apart in a strange way. When properly fleshed-out, and defended against various objections, this proposal injects the required element of subjectivism into our understanding of causation, and hence leads to the view that I want to offer as an analogue of Lewis's own view of chance.

As I said, my main goal is to get this modestly subjectivist account of causation into view. Its comparative invisibility, at present, seems to me a significant failing in contemporary metaphysics. But there are other benefits, besides those of an improved understanding of the metaphysical landscape. Injecting an element of subjectivism into our understanding of causation turns out to throw interesting new light on the cases on Newcomb problems and their cousins. By taking a cue from Lewis's account of chance, in other words, we achieve a view of causation which not only has virtues analogous to those of his own view of chance – e.g., in not leaving it mysterious why such facts should matter to us, in the way that they do – but also offers a resolution of some of the deepest puzzles that have plagued decision theory over the past forty years.

2 Preliminaries

The paper covers a lot of ground, and I want to begin by making explicit two simplifying assumptions – two issues that I am simply going to set aside, to avoid further complexity in what is already an intricate discussion. First, it might be argued that my subjectivist proposal would be better couched in terms of counterfactual conditionals, rather than causation; the former being the more basic notion, perhaps. I shall take no stand on this issue. Except at one or two points where I engage with an argument already cast in terms of counterfactuals, I'll simply talk of 'causation,' for convenience, on the grounds that it is the term in common use in the decision theory literature. I think that much of the rest of the discussion could be readily rephrased in terms of counterfactuals, but I won't try to defend that claim.

Second, and I think relatedly, I shall ignore the question as to whether the proper subject-matter of the kind of account I want to put on the map is causation itself – a relation in the world, as it were – or something like our *concept* of causation (or our use of the *term* 'causation,' perhaps). This is a very

good question, in my view, and it is certainly not unconnected to the issue as to whether either kind of investigation – of the relation of causation itself, or of our concept of causation – would benefit from an injection of subjectivism. But for present purposes I shall set it aside (relying on the hope that however it is best resolved, it will not spoil the analogy between causation and chance, on which much of what follows depends).

In the remainder of this section I offer a brief overview of the Newcomb problem and the debate it gives rise to between EDT and CDT; and then a guide to the structure of the remainder of the paper.

2.1 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 full
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 Nozick's original presentation of the puzzle (1997 [1969]), these rules were taken to be *Dominance* and *Maximize Expected Utility.* Intuitively, Dominance says that if an action A is guaranteed to leave you *no worse off* than action B, and may leave you better off; and if the choice between A and B makes no difference to those factors in the world that determine your payoff; then A should be preferred to B. For example, if you are offered a choice between two ways of betting on a fair coin, one of which offers the same payout as the other in the case of Heads but a higher payout in the case of Tails, then that's the one you should choose. (That choice *dominates* the other choice, as decision theorists say.)

Similarly in the Newcomb problem, apparently. Choosing two boxes leaves you $1,000 richer than choosing just the opaque box, whatever the opaque box contains. And your choice seems to make no difference to whether the opaque box contains money, because that is already determined, before you choose. So Dominance seems to recommend that you take both boxes.

The second principle, Maximize Expected Utility (MEU), recommends that you make your choice by considering a weighted average of the possible payoffs of each of your possible actions, where the individual payoffs are each weighted

by the probability that *that's* the payoff you will receive, given that you perform the action concerned. This weighted average is called the *expected utility* of the action in question, and MEU simply instructs you to choose the action with the highest expected utility.

In the Newcomb case, using the probabilities shown in Table 1, the calculation thus goes like this. For choosing just the opaque box, the expected utility is:

$$\$0 \times \mathrm{P}(\$0 \textit{ in opaque box}|\textit{One-box}) + \$1\mathrm{m} \times \mathrm{P}(\$1\mathrm{m} \textit{ in opaque box}|\textit{One-box})$$

Substituting the probabilities from Table 1, this gives us:

$$\$0 \times 0.01 + \$1\mathrm{m} \times 0.99 = \$990{,}000.$$

In a similar way, we see that the expected utility for choosing two boxes is:

$$\$1{,}000 \times 0.99 + \$1{,}001{,}000 \times 0.01 = \$11{,}000.$$

So, in contrast to Dominance, MEU recommends taking only the opaque box. The puzzle of the Newcomb problem is that it presents us with a stark conflict between these two decision principles, each of which has considerable intuitive plausibility, in the case in question.

Later writers often present the Newcomb problem not as a clash between Dominance and MEU, but rather as a disagreement between two different ways of calculating expected utility (and hence between two different versions of MEU). Thought of in this way, the difference turns on the nature of the probabilities used as weights in the calculation. The first option thinks of probability *epistemically*. In the cases in question, it measures the degree to which the occurrence of a particular action A would provide *evidence* for the occurrence of a given outcome O – in other words, as it is often put, the 'news-bearing' significance of A with respect to O.

This notion of probability gives us what we can call *evidentially-grounded* expected utility ('V-utility,' as it is often called in the literature):

$$\mathrm{V}(A_i) = \Sigma_j \mathrm{V}(O_j)\mathrm{P}_{\textit{evidential}}(O_j|A_i).$$

Here $\{O_j\}$ and $\{A_i\}$ are the relevant sets of Outcomes and Acts, respectively.

The second notion of expected utility relies on what, following Joyce (1999), we may call the 'causal probability' for an outcome given an action. Joyce notes that while this notion "has been interpreted in a variety of ways in the literature, ... the common ground among causal decision theorists is that [it] should reflect a decision maker's judgements about her ability to causally influence events in the world by doing A." (1999, 161) Intuitively, the intent is that $\mathrm{P}_{\textit{causal}}(O_j|A_i) \neq \mathrm{P}_{\textit{causal}}(O_j)$ only if O_j is *causally dependent* on A_i (positively or negatively, as the case may be). This notion of probability gives us what we can call *Causally-grounded* expected utility (or 'U-utility,' as it is known):

$$\mathrm{U}(A_i) = \Sigma_j \mathrm{V}(O_j)\mathrm{P}_{\textit{causal}}(O_j|A_i)$$

These two notions of expected utility then give us two decision rules: *Maximize V-utility,* and *Maximize U-utility.* And it is a simple matter to show that in the decision problem described above, these two rules seem to give different recommendations. By the calculation we have already performed, $\mathrm{V}(\textit{One-box}) = \$990{,}000$, while $\mathrm{V}(\textit{Two-box}) = \$11{,}000$. So the rule Maximize V-utility recommends taking only the opaque box.

On the other hand, so long as we assume that the choice of one or both boxes does not causally influence the contents of the boxes, then

$$\mathrm{P}_{\textit{causal}}(\$0 \textit{ in opaque box}|\textit{One-box}) = \mathrm{P}_{\textit{causal}}(\$0 \textit{ in opaque box}|\textit{Two-box})$$

It follows immediately that

$$\begin{aligned}\mathrm{U}(\textit{Two-box}) &= \$1{,}000 \times \alpha + \$1{,}001{,}000 \times (1-\alpha) \\ &= \$1{,}000 \times \alpha + \$1{,}000 \times (1-\alpha) + \$1{,}000{,}000 \times (1-\alpha) \\ &= \$1{,}000 + (\$0 \times \alpha + \$1{,}000{,}000 \times (1-\alpha)) \\ &= \$1{,}000 + \mathrm{U}(\textit{One-box}),\end{aligned}$$

where $\alpha = \mathrm{P}_{\textit{causal}}(\$0 \textit{ in opaque box}|\textit{One-box}) = \mathrm{P}_{\textit{causal}}(\$0 \textit{ in opaque box}|\textit{Two-box})$. So the rule Maximize U-utility recommends taking both boxes: the expected U-utility of taking two boxes is \$1,000 greater than that of taking one box.

Philosophers disagree about which of these two decision rules provides the rational strategy in such a case. Among famous 'two-boxers,' or 'Causalists,' is Lewis himself, who 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. (1981b, 377)

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

> Some—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. (1979, 240)

In the terminology introduced in the previous section, Lewis thus declares himself to favour CDT (causal decision theory) rather than EDT (evidential decision theory). However, he also remarks that the debate is "hopelessly deadlocked." (1981a, 5) As he puts it in another paper:

> It's a standoff. We [two-boxers] may consistently go on thinking that it proves nothing that the one-boxers are richly pre-rewarded and we are not. But [one-boxers] may consistently go on thinking otherwise. (1981b, 378)

In my view, one of the attractions of the proposal offered in the present paper is that it promises to end this standoff, by showing us how to bring together the seemingly conflicting intuitions of Causalists and Evidentialists (i.e., two-boxers and one-boxers). Briefly, this compromise agrees with traditional Evidentialists in certain cases, including those that involve 'supernatural' sources of information, in the sense mentioned above. It recommends one-boxing in these cases, in other words, backing up the recommendation with a new reply to the traditional Causalists' main objection to one-boxing. (Here, especially, the subjectivist option plays a crucial role.) But it agrees with traditional Causalists in more realistic Newcomb-like cases, explaining why Evidentialists should follow suit. It also proposes and motivates a criterion for deciding on which side of the line a particular decision problem lies.

2.2 Preview

As I noted, my route to this irenic proposal will begin with an apparent conflict in Lewis's own views about decision theory – a tension between his advocacy of CDT, on the one hand, and his professed position concerning chance, evidence and rational credence, on the other. The tension stems from the fact that 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 that an agent who believes that she has what Lewis terms inadmissible information – i.e., direct epistemic access to the outcomes of chance processes – may be rational to allow her credences to be guided by that information, rather than by her knowledge of the relevant objective chances. I shall argue that this amounts to recommending EDT rather than CDT, in a particular class of decision problems. Some of these problems are themselves Newcomb problems, and in these cases, Lewis's view of the relevance of inadmissible information seems literally to support one-boxing, rather than two-boxing.[3]

I shall suggest a resolution of this tension, which rests on extending a proposal by Ned Hall concerning the Principal Principle. Hall argues that Lewis's qualification of the Principal Principle to deal with inadmissible information is

[3] As we shall see, Lewis himself was certainly aware of the class of decision problems in question. He qualifies his own version of CDT by stipulating that it is not intended to apply to them. But if I am right that these cases include a particular class of Newcomb problems – a class in which Lewis's own views on the relevance of inadmissible evidence recommend one-boxing – then excluding them by fiat from CDT is hardly a satisfactory solution, from a two-boxer's point of view. It amounts to withdrawing from the field, in some of the cases in which the conflict with EDT matters most.

unnecessary and undesirable. Better, he argues, to say that there is no such thing as inadmissible information: properly understood, chance relates to rational credence in such a way that such cases simply don't arise.

I propose an analogous move in the case of causation: viz., that where evidential reasoning really does recommend one-boxing, so too does causal reasoning, properly understood. Causation is thus interpreted in such a way that CDT and EDT make the same recommendations, even in Newcomb cases.[4] As I shall explain, this approach treats causation as an 'expert function' for a deliberating agent, in much the way that Hall treats chance as an expert function for a betting agent – an *evidential* agent, in both cases. In both cases, the significance of the 'expert' metaphor is to flag the fact that according to the views in question, it is definitive of the modal notions in question (chance or causation) that they 'can't be wrong' about the corresponding evidential matters.

In the case of chance, the cash value of the metaphor thus lies in the fact that according to Hall's view, there can no such thing as inadmissible evidence – the cases Lewis treats as such are simply cases in which the chances are not what they seem, according to Hall's proposal. To understand the analogue in the case of causation, we need to know what the expert function produces, in that case. For chance, on Hall's view, the expert function's outputs are prescriptions for credences. For causation, according to my proposal, the outputs are prescriptions for the *conditional* credences of Outcomes given Acts that are required by an agent who acts in accordance with EDT – that is, *agentive* conditional credences (or *agentive* conditional probabilities), as I shall call them. As we shall see, the label 'agentive' does triple duty. It marks first the fact that these are probabilities an *agent* needs, according to EDT; second, the fact that they are probabilities conditional on *acts;* and third, crucially, the fact that they are assessed from the *agent's* distinctive epistemic perspective.[5] My proposal is thus that causal dependence stands to agentive conditional credence just as chance stands to credence according to Hall's proposal.

Hall's view of how chance stands to credence is very close to Lewis's own, of course. They differ, essentially, only in their treatment of cases of inadmissible evidence – Hall simply disallows it. Similarly for causation, I think. Someone sympathetic to the analogy I wish to draw might nevertheless prefer an analogue of Lewisean chance to an analogue of Hallean chance, in the case of causation. That is, it is compatible with the view that these modal notions (chance and causation) are both experts, first and foremost, that we might have grounds (from physics, perhaps) to prefer a conception of the modal facts which

[4] I advocated this aspect of the view in earlier work (Price 1991, 1993), but without the support of the analogy I draw here with the case of chance and inadmissible information.

[5] To forestall a possible concern, let me note at this point that the significance of taking these probabilities to be assessed from the agent's distinctive epistemic perspective will not be to open the door to wholesale relativism, relativism grounded in the fact that different agents have different evidence. On the contrary, I take it that in this respect, agentive probabilities can be objectified in familiar ways. Rather it reflects the familiar claim that agents have a distinctive epistemic attitude *to their own actions* – the fact that 'deliberation crowds out prediction', as I will put it later (§8.1), borrowing a phrase from Rabinowicz (2002).

allows they might float free of rational agency, in unusual cases. Exceptional cases, by their very nature, force us to make a trade-off between accuracy and conceptual tidiness. Lewis's picture of chance is tidier than Hall's, but pays for it by having to admit exceptions to the Principal Principle, in some (very) unusual cases.

This trade-off needs to be negotiated for causation, too, according to my proposal. In this case, the unusual cases are Newcomb problems. More precisely, they are *some* of the various decision puzzles called Newcomb problems; including the classic Predictor case described above, under some of its possible disambiguations. I shall have more to say later (§§6, 8, 9) about other cases, such as the more realistic 'medical' Newcomb problems and other versions of the Predictor case. I shall propose a way to sort the various problems into two classes, and argue that provided that causation is understood as I recommend, CDT and EDT can agree on both sides of the line: for one class of cases they agree to one-box, for the other they agree to two-box.

At this stage, the Newcomb problems I have in view are those in the first category, and I take good exemplars of such cases to be those I have already termed *supernatural* cases: cases in which the Predictor has access to Satanic powers, crystal balls, and things of that ilk. In case this diet of unrealistic examples might seem to trivialize matters, I remind readers that my aim is to develop an analogy that goes by way of similar examples in the case of chance.[6]

For the moment I am not in a position to introduce the sorting principle just mentioned, and hence I cannot avoid being non-committal about what other kinds of examples, if any, might fall into this first category, beside these obviously supernatural cases. However, what will eventually be definitive of the cases in this category is that they *really do* involve conditional evidential dependence of Outcomes on Acts, from the agent's point of view, of the kind depicted in Table 1.[7] Since it is standardly assumed that a classic Newcomb problem does exhibit such conditional dependence, I shall help myself for the time being to the label 'classic Newcomb problem' to refer generically to those Predictor cases of which this is true.[8]

Returning now to the trade-off between tidiness and accuracy, note that in the case of chance our preference on these matters does not affect our judgements about rational credence and rational action. Hall and Lewis agree on what credences are rational in the presence of what Lewis calls inadmissible evidence. They agree on this matter even though they disagree about whether the chances are such that these credences follow from the Principal Principle (properly formulated) itself. Similarly for causation, I shall argue. A preference

[6]Another example of a supernatural Newcomb problem, by my lights, is Egan's (2007) case involving time travel. Again, this example has parallels in the literature concerning inadmissible evidence, as indeed in Lewis's own discussion (see Lewis 1986 [1980], 94).

[7]In the second category, the appearance that there is such dependence will be held to rest on a mistaken view of the probabilities, as they are properly assessed from the agent's standpoint.

[8]Eventually, however, I shall want to argue that some of the cases usually thought of as typical Newcomb problems do not count as classic Newcomb problems in this sense, because the view that they exhibit the required conditional evidential dependence rests on a mistake.

for accuracy yields a view of causation such that CDT recommends one-boxing in the classic Newcomb problem. A preference for tidiness yields the verdict that such Newcomb problems are strange cases in which causal beliefs and rational decision behaviour do not keep step. But the rational policy is to one-box, in either case.

In my view, much of the force of the classic Newcomb puzzle 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 himself grasped for chance: rationality and modal metaphysics part company, and the rational choice is to one-box.

I stress again that although my proposal is motivated by an apparent tension in Lewis's views, and disagrees with Lewis about the rational policy in the classic Newcomb problem, it is in other respects Lewisean in spirit. In particular, it aims to extend to causation the well-judged conception of the relation between the subjective and objective aspects of a modal notion that Lewis himself offers us in the case of chance. That extension is the main project of this paper.

3 A chancy Newcomb problem?

As we saw in §2.1, Newcomb problems turn on a conflict between a decision rule couched in terms of an agent's *causal beliefs*, on the one hand, and a decision rule couched in terms of her *evidential beliefs*, on the other. Judgements of causal dependence of Outcomes on Acts support one sort of calculation of expected utility (U-utility). Judgements of evidential dependence of Outcomes on Acts support another sort of calculation of expected utility (V-utility). And classic Newcomb problems are cases in which these two calculations are thought to give different values of expected utility; and hence to give different recommendations for action, if we attempt to *maximize* our expected utility.

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

Table 2. A free lunch?

It is natural to ask whether the same kind of conflict between modally-grounded and evidentially-grounded decision rules can arise for other kinds of objective modality. Can it arise for chance, for example? It is easy to see that it can, at least on some intuitive understandings of chance. Suppose God offers you the payoffs shown in Table 2 on a bet on the outcome of a toss of a fair coin. It is a good bet either way, obviously, but a better bet on Heads than on Tails.

Now suppose that Satan informs you that although God told you the truth, and nothing but the truth, about the coin – in particular, it does have a 50%

chance of landing either Heads or Tails – He didn't tell you the *whole* truth. So far, this revelation shouldn't impress you. You were well aware that – as in the case of any event governed by (non-extreme) chances – there is a further truth about the actual outcome of the coin toss, not entailed by knowledge of the chances. "Tell me something I didn't know," you think to yourself.

"Okay," responds Satan, rising to this silent bait, "I bet you didn't know this. On those actual future occasions on which *you yourself* bet on the coin, it comes up Tails about 99% of the time. (On other occasions, it is about 50% 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.[9]

Let me make explicit some assumptions about this example. The most important are first, that the coin is (or, what really matters here, is *believed to be*) genuinely chancy (and fair); and second, that Satan is a source of inadmissible evidence, in Lewis's sense – in other words, that Satan is, as Hall puts it, a "crystal ball." As Hall says:

> Lewis himself notes [1986 [1980], 94] that there are possibilities (involving such things as time travellers, seers, and circular spacetimes) in which the past carries news from the future which, if known, breaks the connection between credence and chance. When the past does carry such news, I will say that it contains "crystal balls" (whether or not they take the form of magical quartz). (1994, 508)

Once again, as Hall himself notes, it doesn't matter whether there really are any crystal balls, only that our agents be assumed to have reasonable grounds for thinking that there are:

> There needn't actually be crystal balls, there need only reasonably seem to be. That is, a proposition E will be inadmissible at time t if it provides reasonable warrant for a hypothesis that there are crystal balls (along with information about what the balls say)—even if that hypothesis is false. (1994, 508)

The main role of these assumptions is to ensure that the present case is the kind of case that writers such as Hall and Lewis himself had in mind, in discussing chance and inadmissible evidence. Our use of the example is going to be within the framework defined by those previous discussions, which means that we can set aside certain concerns (for example, about the unrealistic nature of the example) that would be shared by other cases within the framework.

Another concern, which I want to defer though not set aside completely, is the thought that in the presence of a crystal ball (in this case, Satan himself) we might not be entitled to assume that the coin is still fair.[10] We will come back to

[9] At this stage, the justification for calling this a Newcomb problem is just that it presents us with a case in which modal and evidential beliefs deliver conflicting recommendations for action, in an unusual way. As we shall see, variants of it become Newcomb-like in additional respects.

[10] The mere fact of the assumed correlation between the result of the coin tosses and our own behaviour does not undermine the claim that the coin is fair, of course. The world is full of unlikely conjunctions. What is odd about this one is simply that it is known in advance.

this thought, which is related to Hall's own prescription about such cases. For the moment, I simply assume, following Lewis, that chance and inadmissible evidence can come apart in this way, and that our game involves an example of that phenomenon.

For future reference, let us also introduce a variant of the game, which makes explicit that we have the option not to bet at all, and that any credences we derive from Satan's revelation are conditional in nature: they are conditional on our choosing to bet. This gives us the two versions of the Chewcomb game, Unconditional Chewcomb and Conditional Chewcomb, as shown in Table 3 and Table 4, respectively. (I hope it already seems plausible that adding the No Bet option makes no difference to the rational choice in the Chewcomb game. Why decline a free lunch, after all, which is what we do if we choose not to bet?) Finally, let us assume that we believe that we, or others relevantly similar to us, will play these games many times in the future, so that Satan's information concerns a large class of future cases.

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

Table 3. Unconditional Chewcomb (with evidential probabilities)

	Heads	Tails
Bet Heads	\$100 (0.01)	\$0 (0.99)
Bet Tails	\$0 (0.01)	\$50 (0.99)
No bet	\$0 (0.5)	\$0 (0.5)

Table 4. Conditional Chewcomb (with evidential *conditional* probabilities)

What is the rational policy, in either version of the Chewcomb game? 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 accordance with the Principal Principle. 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 is easy to calculate), our rational expected return is \$1 if we choose Heads and \$49.50 if we choose Tails (in both versions of the game).

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 initially 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. As Lewis puts it, it would be an "obvious blunder" to take the Principal Principle to dictate the following credence: "C(the coin will fall heads/it is fair and will fall heads in 99 of the next 100 tosses) = 1/2." (1994, 485) So Lewis takes it for granted that someone who has inadmissible evidence should base their credences 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 is easy to configure the Chewcomb problem so that this recommendation is in tension with that of CDT. Lewis's own (1981a) formulation of CDT is based 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:

$$\mathrm{U}(A) = \Sigma_i \mathrm{P}(K_i)\mathrm{V}(A\&K_i).$$

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 simply be taken to be:

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

We then calculate the U-utilities as follows:

$$\mathrm{U}(\textit{Two-box}) = \mathrm{P}(K_0)\mathrm{V}(\textit{Two-box} \,\&\, K_0) + \mathrm{P}(K_1)\mathrm{V}(\textit{Two-box} \,\&\, K_1)$$
$$\mathrm{U}(\textit{One-box}) = \mathrm{P}(K_0)\mathrm{V}(\textit{One-box} \,\&\, K_0) + \mathrm{P}(K_1)\mathrm{V}(\textit{One-box} \,\&\, K_1).$$

The result is that $\mathrm{U}(\textit{Two-box}) > \mathrm{U}(\textit{One-box})$, by the kind of reasoning we used to calculate U-utility in §2.1.

Concerning the probabilities $\mathrm{P}(K_0)$ and $\mathrm{P}(K_1)$, Lewis stresses that if CDT is to remain distinct from EDT, we need to use our *unconditional* subjective 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)

What should we take the dependency hypotheses to be, to apply this framework to the Chewcomb problem? If we assume that because the outcome is the result of a toss of fair coin, it is not causally influenced by the way we choose to bet, then again the dependency hypotheses seem to take a simple form:

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

As we shall see, Lewis's own formulation of dependency hypotheses in such a case is a little more complicated; but it produces the same results, for present purposes, so for the moment we may work with this simpler alternative.

The next issue concerns the probabilities $\mathrm{P}(K_H)$ and $\mathrm{P}(K_T)$. As just noted, Lewis stresses that we need to use unconditional probabilities at this point, if CDT is to remain distinct from EDT. This means that if we set up the example so that Satan's inadmissible evidence yields *unconditional* probabilities, as in Unconditional Chewcomb (Table 3), Lewis can consistently allow that CDT yields the recommendation to bet on Tails. But in Conditional Chewcomb (Table 4), we specified that the information that we learn from Satan doesn't tell us that $\mathrm{P}(K_H) = 0.01$, for example, but only that $\mathrm{P}(K_H|\mathit{We\ bet}) = 0.01$. In this case, Satan's information certainly concerns a 'news-bearing aspect' of the act of choosing to bet rather than not to bet. Accordingly, Lewis's CDT then seems to require that we use $\mathrm{P}(K_H) = \mathrm{P}(K_T) = 0.5$ for calculating $\mathrm{U}(\mathit{Bet\ H})$, $\mathrm{U}(\mathit{Bet\ T})$ and $\mathrm{U}(\mathit{No\ bet})$, 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.*

We thus have two versions of the Chewcomb game, Conditional Chewcomb and Unconditional Chewcomb, where the difference consists in the availability of the No Bet option. The problem for Lewis takes the form of a trilemma (see Table 5). If he recommends betting Heads in both cases, the Unconditional case appears to be in violation of his own policy on the relevance of inadmissible evidence. If he recommends Tails in both cases, the Conditional case appears to be in violation of his own version of CDT. While if he recommends different policies in each case, the difference itself seems implausible. After all, the case has been set up so that it seems obvious that a rational agent will choose to bet – it's a free lunch. And the mixed case seems to yield different recommendations, depending on whether the agent is allowed first to choose to bet and then to choose *which* bet, or has to make both choices at the same time.[11]

Unconditional	**Conditional**	**Problem for Lewis**
Heads	Heads	Conflict with policy on inadmissible evidence
Tails	Heads	Implausible difference in recommendations
Tails	*Tails*	Conflict with CDT

Table 5. Two Chewcomb games – policies and problems

[11] As Arif Ahmed pointed out to me, this amounts to a violation of Independence. In the mixed case, the agent prefers betting on Tails to betting on Heads if she does not have the option not to bet at all, but betting on Heads to betting on Tails if she does have the latter option.

3.1 Following Lewis more closely

I noted above that in applying Lewis's CDT to the Chewcomb game, we used a different choice of dependency hypotheses. When Lewis considers the formulation of CDT in indeterministic worlds, he takes the relevant dependency hypotheses to be counterfactual conditionals whose antecedents are the actions an agent is considering, and whose consequences are full specifications of chances for relevant outcomes. In the Chewcomb game, these counterfactuals take a simple form. In the Unconditional version of the game they are simply:

Bet Tails $\boxdot\!\!\rightarrow \mathrm{Ch}(H) = \mathrm{Ch}(T) = 0.5$
Bet Heads $\boxdot\!\!\rightarrow \mathrm{Ch}(H) = \mathrm{Ch}(T) = 0.5$.

In the Conditional version of the game, where the agent has the option not to bet, we need to add:

No Bet $\boxdot\!\!\rightarrow \mathrm{Ch}(H) = \mathrm{Ch}(T) = 0.5$.

Since the consequent is identical in all three cases, we may take the dependency hypothesis to be simply a specification of the chances – i.e., in this case, the proposition (FC) that the coin is fair. It follows that:

$$\begin{aligned} \mathrm{U}(\textit{Bet Tails}) &= \mathrm{V}(\textit{Bet Tails} \,\&\, \textit{FC}) \\ \mathrm{U}(\textit{Bet Heads}) &= \mathrm{V}(\textit{Bet Heads} \,\&\, \textit{FC}) \\ \mathrm{U}(\textit{No Bet}) &= \mathrm{V}(\textit{No Bet} \,\&\, \textit{FC}). \end{aligned}$$

How should we calculate the utilities on the right hand side of these expressions? The first two expressions require a calculation of expected V-utility, and here, once again, we face the issue of what probabilities we use in the calculation. If we simply use the chances, the option of betting Heads will maximize U-utility. If we use the Satanic evidential probabilities, the option of betting Tails will do so.

Once again, the problem is that both policies seem defensible, in Lewisean terms. Lewis's views on the relevance of inadmissible information seem to recommend betting Tails. But in the Conditional game, this again has the effect of making U-utility sensitive to "that news-bearing aspect of the options that we meant to suppress" (as we saw that Lewis himself put it, in the case of the weights on the dependency hypotheses).[12]

3.2 Discussion

So, as I say, there seems to be a tension here, from Lewis's point of view. I offer the following diagnosis of the difficulty. Newcomb problems are decision

[12]Joyce (2011) argues that Lewis mischaracterizes CDT when he takes its essential feature to be that it uses unconditional probabilities. Rather, Joyce proposes, it is acceptable to use your act as evidence so long as it is evidence for what your act will cause. (Thanks to a referee here.) In the present case, however, this move will only help Lewis if he is prepared to concede that the causal dependencies are non-standard in the kind of cases we are considering – which is the shift I am recommending.

problems in which evidential policies seem to give different recommendations from causal policies,[13] 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 introduce 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).

As I noted earlier, Lewis recognized 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 proposals had been presented in unpublished work by Sobel, 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 difficulty that these cases pose for his own views – a difficulty 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.[14]

In any case, the move of simply setting aside these cases can hardly be regarded as satisfactory, by Lewis's own lights. His own policy on inadmissible evidence seems to yield a clear recommendation in the Unconditional version of the Chewcomb game; and hence a clear recommendation in the Conditional case, too, given the implausibility of the mixed option. We thus have a class

[13] Again, see §2.1 for an account of what this means.

[14] Lewis also notes the difficulty posed by these cases in correspondence with Wlodek Rabinowicz in 1982, saying:

> It seems to me completely unclear what conduct would be rational for an agent in such a case. Maybe the very distinction between rational and irrational conduct presupposes something that fails in the abnormal case. (1982, 2)

(I am grateful to Howard Sobel for alerting me to the existence of this correspondence, and to Wlodek Rabinowicz, Stephanie Lewis and the Estate of David K. Lewis, for giving me access to it.)

of Newcomb-like problems in which Lewis's policy on inadmissible evidence concurs with EDT; and in which CDT escapes defeat only by withdrawing from the field.[15]

4 Making the analogy closer

So far, our Chewcomb problems have been Newcomb-like in two respects. Even the Unconditional version of the Chewcomb game is analogous to a Newcomb problem, in that it provides a case in which modal beliefs and evidential beliefs yield different recommendations.[16] (The difference between Unconditional Chewcomb and Newcomb is that the modality concerned is chance rather than causality.) But the introduction of the Conditional game produced a decision problem which is Newcomb-like in a more direct sense, namely, that (unlike Unconditional Chewcomb) it does involve an apparent conflict between CDT and an evidentially-based decision rule.[17] In other words, it shifts the tension between modally-grounded and evidentially-grounded decision principles from the case in which the modality is chance to the case in which it is causation (as it is in standard Newcomb problems).

On the face of it, we can go even further. We can produce a Chewcomb game whose decision table looks exactly like that of the classic Newcomb problem. 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, "It is definitely a fair coin, but my crystal ball tells me that in 99% of future cases in which people choose to one-box in this game, the coin actually lands Tails; and ditto for two-boxing and Heads."

As before, we assume that God and Satan are telling the truth. What is the rational decision policy in this case? Here the evidential and causal recommendations seem to be exactly as in the original Newcomb problem, as described in §2.1. Your action will not have any causal influence on whether there is money in the opaque box, apparently. How could it do so, when that is determined by the result of a toss of a fair coin?[18] But as Table 6 shows, you take there to be a strong *evidential* correlation between your action and the

[15]True, they are "extraordinary cases," as Lewis puts it. But so, too, is the classic Newcomb problem. Once CDT has become fickle in this way, what reason do we have to trust it in that case?

[16]Note that the conditionals employed by EDT are indicatives, not counterfactuals – see, e.g., our first quotation from Lewis in §2.1 – so modality is not 'snuck back in' on the evidential side. (Thanks to a referee at this point.)

[17]Provided, at least, that the latter is understood in the light of Lewis's policy on inadmissible evidence.

[18]In the next section I propose an understanding of causation that challenges this claim, but for the moment I am assuming that someone who says that the agent has no causal influence on the contents of the opaque box in the standard Newcomb problem will be inclined to say the same here.

result of the coin toss, such that you are much more likely to get rich if you one-box.

	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 6. Boxy Chewcomb (with evidential conditional probabilities)

4.1 Remembering the counterfactuals

In this case there is no unconditional version of the game to highlight the tension in Lewis's position in the way that we did above. (The parallel with the original Newcomb problem depends on the fact that the high evidential probability of money in the opaque box is conditional on the agent's only choosing that box.) Accordingly, two-boxers will feel their standard reply will suffice, in this case. They will argue that whatever payout the one-boxer receives in Boxy Chewcomb (Table 6), it will always be true that *had* she two-boxed, she would have received the same payout plus \$1,000. Joyce puts the general argument like this:

> Having gotten the \$1,000,000, [the one-boxer] must believe that she would have gotten it whatever she did, and thus that she would have done better had she taken the \$1,000. So, while she may feel superior to [the two-boxer] for having won the million, [she] must admit that her choice was not a wise one compared to *her own* alternatives. ... She made an irrational choice that cost her \$1,000. (1999, 153)

The availability of this argument in the present case turns on the fact that Boxy Chewcomb is crucially different from the original Conditional Chewcomb game (Table 4), where no such appeal to counterfactuals is possible, by the Causalist's lights. In that case, an agent who bets Tails and wins cannot be told that she would have won even more, had she bet Heads (or not bet at all): on the contrary, presumably, she would have won nothing, in either case.

Accordingly, a Causalist might try to hold the line here – defending two-boxing in Boxy Chewcomb – while conceding both the Conditional and Unconditional versions of the previous game to the Evidentialist. It would still need to be explained how CDT can be formulated so as to follow EDT in Conditional Chewcomb, without also endorsing one-boxing in Boxy Chewcomb, too. But the counterfactuals associated with this response to the one-boxer seem to mark a line at which a stand might be made.

However, I want to try to pre-empt this defensive strategy, by counterattacking on what seems to two-boxers the safe side of the line. I think that Evidentialists typically concede too much to their Causalist opponents, in granting them the counterfactuals on which the charge that one-boxing is irrational always depends. The analogy with the case of chance and inadmissible evidence, and a proposal made in that context by Ned Hall, together suggest a much more forceful response.

5 One-boxing via the Hall way?

Hall (1994, 2004) recommends that we replace Lewis's Principal Principle with a modified principle, requiring that rational credences track *conditional* chances: chances *given our evidence*.[19] At first sight, 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.[20] 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, this would be a victory for Evidentialism: it agrees with the Evidentialist that we should take the Satanic information into account. But it would be a face-saving outcome for the Evidentialists' opponents, too, in that it maintains that they never had any good reason to disagree (they were simply misled by relying on unconditional chances rather than conditional chances).

Things aren't so simple, however. 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 metaphysical chance of a fair coin's landing Tails is insensitive to such supernatural vocalizations (our objector insists), and so the shift to conditional chances makes no difference.[21] In such a case, it remains an issue whether rational credence (conditional or otherwise) 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 not in the requirement that rational credences track conditional chances, but in the fact that he proposes a view of chance which makes it automatic that conditional chances *are* sensitive to such evidence. Drawing 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

[19] Translating Hall's terminology to match that of Lewis, as in fn. 1 above, we may write Hall's version like this: Let C be any reasonable initial credence function. Let t be any time. Let x be any real number in the unit interval. For an arbitrary proposition E, let X_E be the proposition that the *conditional chance* at time t of A's holding *given* E equals x. Then $C(A|X_E E) = x$.

[20] Or the information *that* Satan has provided this information, perhaps.

[21] In symbols, the suggestion is thus that $\mathrm{Ch}(\textit{Tails}|\textit{Satanic whisper}) = \mathrm{Ch}(\textit{Tails}|\textit{No Satanic whisper})$. I 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.

> 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. "She 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*," he says. (2004, 101)

Thus for Hall it becomes a matter of definition that chance and reasonable credence *cannot* come apart, once we have conditionalized on all our evidence;[22] even if some of that evidence counts as 'inadmissible,' by Lewis's lights. And it is this stipulation, rather than the conditionalization move itself, that ensures that there cannot be a genuine Chewcomb problem – a genuine case in which chance and evidential reasoning come into conflict.

I've stressed this point because it is the latter aspect of Hall's view – the view that chance is an analyst-expert – that seems to me analogous to an attractive resolution of the original Newcomb case. In Hall's terminology, the resolution turns on the proposal that *causal* dependence should be regarded as an analyst expert about the conditional credences required by an *evidential* decision maker. A little more formally, the proposal goes something like this:

> (EC): B is causally dependent on A just in case an expert agent would take $\mathrm{P}(B|A) \neq \mathrm{P}(B)$, in a calculation of the V-utility of bringing it about that A (in circumstances in which the agent is not indifferent to whether B).

Since this suggestion takes it to be definitive of causal belief that its role is to guide a particular kind of evidential judgement, I shall call it the *EviCausalist* proposal (hence 'EC,' above). It may seem to fall victim very swiftly to some well-known counterexamples – more on those in a moment. First, in its defence, note that it is not merely an *analogue* of Hall's proposal in the case of chance. It is something close to a *consequence* of it, at least if we wish to retain the intuitive connection between *causing* something, on the one hand, and *raising the chances* of it, on the other. Consider our Boxy Chewcomb game (Table 6), for 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 'actualize' 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 suggests that we *do* have causal dependence of Outcomes on Acts. By choosing to one-box rather than two-box, we greatly *increase* the chance of Tails.[23]

[22] As I noted in §2.2, this is the meat of the 'expert' metaphor. Chance simply 'can't be wrong' about rational credence, as it were.

[23] The fact that Hall's proposal is formulated in terms of conditional chances might seem to give

In effect, the EviCausalist proposal is simply that we should take seriously *in general* the view of causal dependence that is thus forced on us in this particular case, if we wish to combine Hall's view of chance with an intuitive understanding of the relation between chance and causation. Comparing Table 1 and Table 6, it is easy to see that EviCausalism will treat the classic Newcomb problem in just the same way as the Boxy Chewcomb problem. In other words, it implies that the contents of the opaque box in the classic Newcomb problem *are* causally dependent on the agent's choice, in the sense of causal dependence now proposed. Accordingly, the EviCausalist will regard the classic Newcomb problem not as a case in which CDT and EDT come apart, but simply one in which the causes are not what we initially assume. We might take this to imply that it is not really a Newcomb problem at all, on the grounds that as Joyce (1999, 152) puts it, "[i]t is part of the definition of a Newcomb problem that the decision maker must believe that what she does will *not* affect what the psychologist has predicted." But this is a terminological matter, on a par with that as to whether, in the light of Hall's proposal, we want to continue to speak of 'inadmissible evidence.' The substantial point is that in the classic (so-called) Newcomb problem, EviCausalism proposes an understanding of the causal structure of the case such that CDT and EDT agree in recommending one-boxing.

It is not news, of course, that CDT recommends one-boxing if the agent's choice affects what the Predictor puts in the boxes. Retrocausal variants of the original Newcomb problem are familiar – they feature in Nozick's (1997 [1969]) paper. What EviCausalism adds to this background is a proposal about the nature of causal dependence itself, such that the Newcomb problem cannot *but* be retrocausal, if there is genuine evidential dependence of the Predictor's behaviour on the agent's choice, from the agent's point of view.[24]

As I remarked above, however, the EviCausalist proposal may seem an obvious non-starter, blocked by familiar and ordinary 'medical' cases, in which it is (widely thought to be) clear that causal dependence and evidential dependence do not align with one another, in the way that EviCausalism appears to suggest. I turn to this objection in a moment. But before that, I want to stress one more lesson to be drawn from the analogy with Hall's view of chance. In neither case, for chance or for causation, is Hall's view or its causal analogue the only game in town. In either case, we might have grounds to prefer a modal

rise to a difficulty for my proposed analogy. Chance is a kind of probability, and so the notion of conditional chance makes sense. But doesn't the analogue require a notion of conditional causation? If so what could that be? (Thanks to a referee at this point.) The solution is to note that the EviCausalist proposal takes the basic notion to be one of *causal dependence*, defined probabilistically in a standard way. The principle (EC) invokes expert assessments of two probabilities, one conditional and the other unconditional ($\mathrm{P}(B|A)$ and $\mathrm{P}(B)$, respectively), and requires that we compare one to the other. There would be no problem, in principle, in introducing a further conditionality into this principle, so that an evidential proposition E came to play the same role here as it does in Hall's account. But that would be an unnecessary complexity, for present purposes – as I noted, the crucial part of Hall's proposal, for our purposes, is the treatment of chance as an expert function. The role of conditionality is secondary.

[24] And if the agent really has a choice in the matter, of course.

notion that could drift apart from evidence, in unusual cases. I merely want to claim that in this eventuality, once we recognize it for what it is, it should seem clear that the rational choice goes with the evidence, not with the modal notion.

As we have seen, this already looks unremarkable to us in the case of chance. In that case, Lewis himself offers us a modal notion (chance) that can diverge from evidence, in strange cases. He regards it as obvious that rational credence follows the evidence, not the modal facts, in such exceptional cases. I am proposing (and will be arguing) that this should seem just as unremarkable in the case of causation.

6 A cigarette at bay?

Whatever the appeal of EviCausalism in Chewcomb cases, it may seem that there are familiar Newcomb problems in which causal dependence and evidential dependence are clearly distinct. Consider the famous case of the Smoking Gene, for example, in which an agent believes that there is a gene which predisposes both to smoking and cancer, ensuring that these two outcomes are positively correlated (see for example Jeffrey 1981, 476–78). In general, the fact that someone is a smoker thus indicates that she is more likely than otherwise to have the gene, and hence more likely than otherwise to develop cancer. EDT is therefore held to recommend that even if such an agent prefers smoking to not smoking, other things being equal, she should decide not to smoke, in order to minimize the evidential probability that she will develop cancer (and thereby maximize her expected V-utility). But it would add idiocy to irrationality, surely, to try to justify this recommendation by claiming that causation should be understood in such a way that this agent can *cause herself* to lack the gene.

Indeed it would, and I make no such claim. Instead, I propose that in these familiar cases, the agent is making a mistake – a mistaken probabilistic inference, not a mistaken decision – if she concludes that her choice as to whether to smoke is evidentially relevant to whether she carries the gene in question, *from her own point of view.*

In support of the claim that this proposal is at least not obviously absurd, I appeal first to the authority of some of my (traditional) Causalist opponents, who recognized long ago that Evidentialists could get fairly close to this claim. Here is Brian Skyrms (1980, 130), for example: "There is a defense for [the Evidentialist] which can be pushed very far, but not, I think, far enough." Skyrms is talking about what was then becoming known as the *Tickle Defence:* an argument that in a case such as the Smoking Gene, an agent should indeed regard her action as probabilistically independent of whether she carries the gene. The essence of the Tickle Defence is the thought that an agent's special epistemic access to her own beliefs and desires inevitably screens off, for her, the evidential relevance of any prior factors, such as the Smoking Gene, that might be correlated with her choice. Perhaps she feels a 'tickle' (an urge to smoke), or perhaps her evidence is more subtle; but in principle, in some way

or other, she has access to information that screens off the correlation between her choice and whether she gets cancer.[25]

Lewis himself goes even further than Skyrms, in assessing the prospects of this argument:

> I [say] that the Tickle Defence does establish that a Newcomb problem cannot arise for a fully rational agent, but that decision theory should not be limited to apply only to the fully rational agents. Not so, at least, if rationality is taken to include self-knowledge. May we not ask what choice would be rational for the partly rational agent, and whether or not his partly rational methods of decision will steer him correctly? (1981a, 10)

It seems to me that at least in hindsight, this assessment positively *invites* a response framed in terms of expert functions. More about this in a moment, but before that, a couple of preliminary points.

Obvious no longer

First, a remark on the relevance of the dialectic of these old discussions to the present case. As noted, the acknowledged successes of the Tickle Defence do much to meet the objection that there are cases in which it is *obvious* that my proposed analogue of Hall's suggestion will attribute causal dependency, where actually there is none. These successes force the Evidentialist's opponents to retreat in one of two directions: either to less familiar and less realistic examples, in which it is correspondingly less plausible to say that the causal structure is not a matter for debate; or, as noted, to less rational agents, about whom there is inevitably an issue about the nature of their irrationality. So long as we Evidentialists can find an alternative interpretation to *decision-theoretic* irrationality, these agents need not trouble us.

Both points are well made by Paul Horwich. Concerning the first, Horwich notes that there are analogues of medical Newcomb problems that might evade the Tickle Defence:

> It is not difficult to concoct highly artificial examples in which, *ex hypothesi*, there is no tickle, no screen, and therefore no argument for the convergence of the evidential and causal principles. For example, we could simply have stipulated that cancer be correlated with smoking . . . (1985, 435)

However, he continues

> such scenarios do not constitute clear counterexamples to the evidential principle because they are extremely unrealistic—in exactly the same way as Newcomb's problem itself—and cannot, therefore, provide the material for authoritative intuitions. (1985, 435)

Later, taking up the Lewis's objection that "decision theory should not be limited to apply only to the fully rational agents," Horwich makes the second point. He points out that Lewis's objection

[25] Skyrms says, "I have heard this defense independently from Frank Jackson, Richard Jeffrey, David Lewis, and Isaac Levi." (1980, 130) More recent versions of this argument include those of Horgan (1981), Eells (1981, 1982, 1984), Horwich (1985) and Price (1986, 1991).

> neglects a certain systematic equivocation in the evaluation of actions. They are always judged in relation to desires and beliefs which are themselves susceptible to evaluation. Therefore, an act may be criticized as irrational because it was based on irrational beliefs, even though it was correct relative to those beliefs. (1985, 438)

We'll return to this observation below, when we have some more terminology in place. At that stage, we will be able to amplify Horwich's point with reference to the analogy with chance and credence.

A causal shortcut to evidential virtue

Next, I want to call attention to an advantage of EviCausalism with respect to medical Newcomb problems which is not shared by more orthodox versions of Evidentialism (such as Horwich's). Suppose, as the EviCausalist claims, that information about causal dependencies *just is* expert information about the corresponding evidential dependencies, from an agent's point of view. In the simplest case, in other words, the information that events of type A are (positively) causally relevant to events of type B is the information that rationality requires that an agent contemplating an action of type A take it to be positively evidentially relevant to the occurrence of an outcome of type B.[26] Then, at least in familiar and uncontroversial cases, such as that of the Smoking Gene, an agent with a proper grasp of the causal concept and firm beliefs about the causal structure of a particular case can no more be confused about the evidential dependencies, than, according to Lewis, an agent with firm beliefs about chance and a good understanding of the concept can be confused about the associated credences. For causation as for chance, the EviCausalist insists, confusion in uncontroversial cases is simply an indication that the agent in question does not have a proper grasp of the concept.

Indeed, the EviCausalist can go further. Having interpreted causal information in this evidential manner, she can allow that it is a considerable advantage of CDT, in many cases, that it operates directly with this encoded form of evidential information. Like computers programmers more comfortable in C++ than in machine code, ordinary agents find it much easier to operate at the causal level of description – much easier, thereby, to avoid the perils of probabilistic inference, a task which most of us are prone to get wrong.[27]

But this convenience comes with a cost. In unfamiliar circumstances, it may

[26] Readers may baulk here at the subjectivism of this proposal. "Were there no causal dependencies before there were agents?," as the baulker might put it. Indeed there were, just as according to Lewis there were chances before there were any creatures with credences. Nevertheless, Lewis holds that we cannot properly characterize chance unless we do so in terms of credence – unless we say, in effect, that information about chance *is* information about rational credence. The EviCausalist says the same about causal dependence. (See also fn. 5 above.)

[27] This is not to say that we are not prone to make mistakes in causal reasoning, too, but simply that it can be an advantage to package information in a way that puts its practical implications front and centre, so that the recipients do not need to work them out on the fly. As a further analogy, compare the information that the thing in the bushes is a large hungry carnivore with sharp claws, with the information that the thing in the bushes is *very dangerous.* It is often helpful not to have to go back to first principles.

seem to us that the causal facts and evidential facts pull in opposite directions. In familiar cases, we rely on various associations between causal facts and other features of situations – in other words, we take various criteria to be grounds for ascribing or withholding causal claims (i.e., really, on this view, evidential claims). But in unusual circumstances, these criteria can be a poor guide to the evidential structure of the case in question. We are habituated to regarding them as good guides to causal structure, and so it seems that causal and evidential dependency are coming apart. But it is an illusion, generated by the mistaken assumption that we were dealing with two distinct kinds of information in the first place – by the fact that we have allowed the causal realm to take on a life of its own, distinct from our evidential point of view.

Is CDT to EDT what chance is to credence?
Once again, the analogy with chance is helpful at this point. According to my EviCausalist, *causal dependence* stands to the conditional subjective probabilities needed by EDT, much as *chance* stands to the subjective probabilities, or credences, required by decision makers whose rational behaviour is modelled by an unconditional decision theory of Savage's sort. Savage's (1954) theory is a *subjective* rational decision theory: it prescribes rational behaviour for agents with a given set of credences and preferences, but remains silent about the rationality of those credences and preferences themselves. The Principal Principle steps into the latter gap (in the case of credence), imposing a rationality constraint on credences, in the light of the agent's beliefs about chances (or in the light of the *facts* about chances, if we wish to interpret the Principal Principle as an objective constraint on rational credence).[28]

Note that in principle we could combine these two levels, formulating an analogue of Savage's theory directly in terms of beliefs, or even facts, about chances.[29] Why wouldn't that be preferable? Well, because it would formalize Horwich's "systematic equivocation in the evaluation of actions," for one thing;[30] and thereby, arguably (more on this in §7.1), obscure something very important about the 'subjective,' 'pragmatic' or 'practical' foundations of the concept of chance itself – the sense in which the concept has its roots in subjective decision.

Let SDT_{ch} be such an 'objective' version of Savage's decision theory, formalized in terms of chances, and SDT_{ev} the familiar subjective version. The EviCausalist regards the relation between CDT and EDT as closely analogous to

[28]Interpreted in the usual (former) way, the Principal Principle tells an agent that if she *believes* a coin to be fair, and has no inadmissible evidence about the matter, she should assign a credence 0.5 to the coin's landing Tails. Interpreted in the latter way, it tells her that if a coin *is* fair, and she has no inadmissible evidence about the matter, she should assign a credence 0.5 to the coin's landing Tails.

[29]In the first case, we use the Principal Principle as our guide to the beliefs about chances relevant for substitution for each of the subjective probabilities in SDT; in the second case, we simply substitute chances directly for subjective probabilities.

[30]In other words, it would obscure the distinction between an agent who acts irrationally given her credences, and an agent who has irrational credences given her beliefs (or the facts) about the relevant chances.

that between SDT_{ch} and SDT_{ev}. CDT is simply the 'objectified' version of EDT, and hence runs together two issues: the *subjective* issue of the rationality of a decision policy, given certain preferences and conditional credences, and the *objective* (or at least *less subjective*) issue of the rationality of certain conditional credences – credences of outcomes given actions – given the facts, or the agent's beliefs, about causation. (CDT then has the analogous disadvantage to SDT_{ch}, in that it obscures the practical, subjective roots of the concept of causation itself, and invites Horwich's "systematic equivocation.")

Two dimensions of expertise

With all this in hand, let us return to Lewis's remark that "decision theory should not be limited to apply only to the fully rational agents." Lewis is right, of course, that (subjective) decision theory should not simply fall silent, in the case of an agent whose beliefs and preferences are not fully rational. But this is compatible with the insight that decision theory itself is supposed to be an expert, and therefore *intolerant* of irrationality in decisions made *on the basis* of those beliefs and preferences. Agents themselves may be irrational at this step, but decision theory aims to codify the standard that rational agents are *trying* to meet. So there is a sense in which decision theory does "apply only to the fully rational agents." Taken *descriptively,* it does not apply – at least not strictly – to agents who are not fully rational in making decisions on the basis of their credences and preferences. Taken *prescriptively,* it does tolerate irrationality; but only in the *acquisition* of beliefs and preferences, not in its own domain.

Subjective decision theory is the expert we consult as we try to do the best with the credences and preferences we actually possess. But to what experts do we turn to avoid the kind of irrationality that decision theory itself tolerates? That is, for help with the credences themselves? Hall has already given us a large part of the answer, perhaps all of it. We need two experts: first, the database-expert, who knows all the evidence that we ourselves would have, under idealization; and second, the analyst-expert, who knows what credences to assign on the basis of that evidence.

However, in the cases for which we need Jeffrey's subjective decision theory rather than Savage's – cases with conditional dependence of States and hence of Outcomes on Acts – these two experts have a special job to do. They need to collaborate to consider the special epistemic situation of the deliberating agent – I will have more to say about this special epistemic situation in §8.1 – in order to determine the rational conditional credences of States given potential Acts, from her point of view. Because the task involves this collaboration, it will be helpful both for the two experts and for their clientele to create a single shopfront, through which requests for guidance may conveniently be channelled. The EviCausalist proposes that this expert shopfront – the 'Agency Guidance Agency,' perhaps – is causal dependence.

6.1 Resuscitating the medical objections?

This program for aligning CDT and EDT would be undermined by a genuine medical Newcomb problem – i.e., a realistic case in which it was clear that the relevant causal dependencies really differed from the evidential dependencies, from the agent's point of view. With such a case in hand, critics could fairly object that EviCausalism amounts, at best, simply to changing the meaning of 'causal dependence,' in a way that obscures the genuine difference between CDT and EDT.

Realistic cases seem to be hard to find, however, and this is certainly good news, from the EviCausalist's point of view. But shouldn't the EviCausalist expect even better news? After all, if the EviCausalist wants to claim that there is some sort of conceptual tie between causal dependence and agentive evidential dependence, shouldn't it be more than a contingent matter that there are no cases in which these notions clearly diverge, in the way that the proposal seeks to disallow?

In response to this challenge, I want first to emphasize, once again, that EviCausalism need not claim that it offers the *only* acceptable understanding of causal dependence. On the contrary, it should acknowledge that causation has other conceptual ties, and – in a good Quinean spirit – allow that the preservation of these ties might seem preferable, in some quarters, when the concept comes under pressure for revision in strange cases.[31] By the resulting lights, it will indeed seem that causal dependence can 'come apart' from conditional evidential dependence, even when the latter is assessed from the agent's point of view.

The significance of this loophole should not be exaggerated, however. For one thing, it would not help in the face of a genuine medical Newcomb problem, where it would be implausible to maintain that the notion of causal dependence was under any sort of conceptual pressure. For another thing, EviCausalism is committed to an 'in principle' claim of a weaker sort, even in strange cases. If causal dependence and conditional evidential dependence are allowed to part company in this way – if causation's other conceptual ties are judged to be more worth preserving, in strange cases – the EviCausalist wants to maintain that it should be nevertheless clear, at least when all the cards are on the table, that rationality goes with conditional evidential dependence, rather than with causal dependence. In other words, it should be clear that if causation is taken this way, such cases provide counterexamples to CDT.

So EviCausalism has some work to do, and I want to suggest a line of attack. It relies on a feature of the landscape where CDT and EDT already find common ground, in the thought that in certain cases, Evidentialists will

[31] Thus Michael Dummett (1954, 32ff.), though defending the conceptual possibility of circumstances that would support deliberation for past ends, argues that these would be cases of "quasi-causation," rather than genuine causation. (The principle he thereby preserves is that remote causes should *begin* a process that leads to their effects.) Dummett's terminological choice illustrates the present point very nicely. The fact that he takes quasi-causation to support means–end reasoning makes it clear that in his view, rational decision does not cleave strictly to *causal* dependence, in unusual cases of this kind.

actually do better than Causalists. This is the basis of the famous *Why Ain't You Rich?* argument against two-boxing. What will be important for my argument will be that the EviCausalist and her opponent will agree about when EDT leads to greater riches (i.e., higher expected V-utility) than CDT, under certain specified circumstances (namely, that it is a random matter which decision policy an agent follows).

Agreement on this matter means that we have a criterion acceptable to both sides for dividing Newcomb-like decision problems into two kinds of cases. In one kind of case, where randomly-assigned Evidentialism does lead to riches, EviCausalism will be able to appeal to a novel response to the Causalist's usual objection to the *Why Ain't You Rich?* argument, to argue that in these cases it is irrational not to follow the Evidential policy, even if one prefers for other reasons not to label the case in question as one of genuine 'causal' dependence. In the other kind of case, where randomly-assigned Evidentialism does not lead to riches, the EviCausalist will be able to argue that there is no agentive evidential dependence, in the relevant sense; and hence that her own version of EDT does not differ from CDT, with causation standardly understood.

If this argument works, it provides both the insurance the EviCausalist seeks about the non-existence of 'realistic' Newcomb problems, and support for her claim that rationality goes with EDT, even if causation is understood in such a way that EDT and CDT diverge, in 'unrealistic' cases. So first, then, to the EviCausalist's response to the Causalist's objection to *Why Ain't You Rich?*, in the classic Newcomb case.

7 We're all Causalists now

As we saw in §4.1, the standard Causalist response to the *Why Ain't You Rich?* argument goes something like this: "Sure, one-boxer, you're rich. But if you *had* two-boxed in those same games, you would have been even richer." To quote Joyce (1999, 153) once more, "[t]he 'If you're so smart why ain't you rich?' defense does nothing to let [the one-boxer] off the hook; she made an irrational choice that cost her $1,000."

However, unlike a traditional Evidentialist, who accepts the Causalist's conception of the modal landscape, my EviCausalist will simply *deny* that "she would have gotten [the million] whatever she did." On the contrary, as she understands the counterfactuals – regular *causal* counterfactuals, as she sees them, not backtrackers[32] – she would have received only $1,000, had she two-boxed. It is the two-boxer who is irrational in this counterfactual sense, by the EviCausalist's lights: *had* the two-boxer one-boxed instead, she would have had the million.

At this point, a lively discussion is likely to ensue about who has the 'proper' notions of causation and counterfactual dependence. But as we have already

[32]This is how EviCausalism differs from the view of Horgan (1981). As Horgan puts it, "I do recommend acting *as if* one's present choice could causally influence the being's prior prediction, but my argument does not presuppose backward causation." (1981, 340–41)

seen, the EviCausalist is prepared for this. "Keep your notions of causation and counterfactual dependence, if you wish," she says to her traditional Causalist opponents:

> "But recognize, with me, how we came to the present juncture, where we need to make a choice. In the case of chance, a 'supernatural' source of information about the future (i.e., a source not envisaged by our usual physical theories) would confront us with a choice about how to continue to use the notion of chance: we could hold fixed our notion of chance, and deal with the unusual cases by allowing an exception to the Principal Principle; or we could modify our notion of chance, and preserve the universality of the Principal Principle. But whichever we choose, it is clear that it makes no difference to rational betting behaviour. Either way, we should not ignore the new information – that's why the first choice requires an exception to the Principal Principle, after all. The only wrong option is the choice that muddles and mixes the two right options, by holding fixed the standard notion of chance, *and* insisting on the universality of the Principal Principle.
>
> Similarly for causation. We can imagine cases – the classic Newcomb problem is one – that confront us with a choice about how to continue to use the notions of causation and counterfactual dependence. Again, we have two choices. We can hold fixed the traditional notions of causation and counterfactual dependence, and allow exceptions to CDT (which is the analogue, here, of the Principal Principle); or we can preserve the universality of CDT, by allowing that the causal structure of these strange cases is not what initially we took it to be. Again, this choice makes no difference to the rational behaviour in such a case: either way, it is to one-box. The only wrong option is the choice that muddles and mixes the two right options, by holding fixed the standard notion of causation, *and* insisting on the universality of CDT."

The traditional Causalist will want to disagree rather vigorously at this point, of course. She will want to defend this 'mixed' option – and to deny that it involves any sort of 'muddle'! To do so, she needs to *explain* the relevance of causality, as *she* understands it, to rational strategic deliberation. To see what is at issue here, it is helpful, once again, to compare the analogous problem in the case of chance.

7.1 The limits of objectivism

Some philosophers feel that there is a problem about explaining the link between beliefs about objective probabilities and rational credence, of the kind encapsulated in the Principal Principle. David Papineau, for example, calls this connection the "Decision-Theoretical Link":

> We base rational choices on our knowledge of objective probabilities. In any chancy situation, a rational agent will consider the difference that alternative actions would make to the objective probabilities of desired results, and then opt for that action which maximizes objective expected utility. (1996, 238)

"Perhaps surprisingly," Papineau continues, "conventional thought provides no agreed further justification [for this principle]":

> Note in this connection that what agents want from their choices are desired *results*, rather than results which are objectively *probable* (a choice that makes the results objectively probable, but unluckily doesn't produce them, doesn't give you what you *want*). This means that there is room to ask: *why* are rational agents well advised to choose actions that make their desired results objectively probable? However, there is no good answer to this question .... Indeed many philosophers in this area now simply take it to be a primitive fact that you ought to weight future possibilities according to known objective probabilities in making rational decisions. ... It is not just that philosophers can't agree on the right justification; many have concluded that there simply isn't one. (1996, 238)

Not all views of probability will agree with Papineau that there is any such a problem, however. One tradition, variously known as *subjectivism, pragmatism,* or *Bayesianism,* regards it as a pseudo-problem, generated, in effect, by starting one's account of probability in the wrong place. Provided we *start* with the insight that probabilistic models are guides to decision-making under uncertainty in particular domains, there's no further mystery as to why they may be used for that purpose. There is no "primitive fact" needed, and no decision-theoretic missing link. There may be other interesting questions in the vicinity: for example, about how, and why, such probabilistic models are linked to other kinds of models, such as those provided by physics in various domains; and about whether these links uniquely constrain the associated probabilistic models. But these are not the practical puzzle about why probability properly guides action. That isn't a puzzle at all, from the subjectivist point of view.

Another interesting issue, famously explored in Lewis's own account of chance, is the extent to which this subjectivist insight can be combined with an objectivist, or metaphysically realist, theory of chance.[33] As we noted at the beginning, Lewis thought that it could be. In his account, the tie to subjectivism consists in the fact that it is *definitional* of objective chances that they support the Principal Principle. If something doesn't do that, it isn't properly called chance. "A feature of Reality deserves the name of chance to the extent that it occupies the definitive role of chance," as Lewis (1994, 489) puts it.

But consider a view of chance of the kind Papineau has in mind, prepared to take it to be, as Papineau puts it, "a primitive fact" that chance constrains rational credence in accordance with the Principal Principle. We can imagine that such a view – emboldened by its own courage in making a stand on this point – might also dig in its heels concerning the rationality of betting Heads, in the first version of our Chewcomb game, despite the availability of inadmissible evidence. "To hell with Satan," says this mad-dog objectivist, thumping the

[33] This is a distinct issue because the subjectivist point of view just mentioned is quite compatible, on the face of it, with antirealism about chance. Indeed, if we have a theory of subjective probability, what need or place is there for a theory of objective probability, too? The issue I have in mind here concerns responses to that challenge.

table. "By betting Tails, you *irrationally* forgo an *equal chance* of a *greater reward*." Or in the past tense: "No matter that you actually won; you were nevertheless irrational, because you sacrificed an equal chance of a greater reward." Or in the long run: "No matter that you have won many times, and are now rich; and that I, betting on Heads, am not rich. This is not a mark of irrationality on my part, but merely a sign that the rewards were reserved for the irrational."

I am not sure whether anyone actually thumps the table to this dialectical end, in the case of chance. But many people, including Lewis, staunchly defend what I take to be its analogue in the case of causation: that is, orthodox two-boxing. The full set of analogies is depicted in Table 7. On the right hand side are views that take modal beliefs to constrain practical rationality, even in cases of exceptional evidence. On the left hand side are views that construe practical rationality in evidential terms; typically combining this preference with some element of subjectivism about the associated modal judgements. In the middle are mixed positions, that allow that there may be exceptional cases in which evidence and objective modality part company, and in which practical rationality goes with the former. Lewis himself holds the mixed view in the case of chance, but the full modal priority view in the case of causation – and the combination creates internal difficulties, as we have already seen.[34]

	Evidential priority	**Modal priority with exceptions**	**Modal priority**
Chance	Hall	Lewis	The table-thumper
Causation	EviCausalists	Horgan, Horwich	Two-boxers

Table 7. Three views of practical rationality

Is my comparison of the orthodox two-boxer position to table-thumping objectivism about chance a fair one, or can causal objectivism do better than its probabilistic cousin? Can the causal objectivist *justify* (rather than simply assume as primitive) the claimed link between causal judgement and rational decision (and hence explain why the objective modality takes precedence, in case of conflict with exceptional evidence)?

What does the history of these debates tell us about the prospects for such an argument? It reveals a widespread acceptance, even on the part of two-boxers themselves, that there is no such argument to be found. As we noted earlier, Lewis himself says that the debate is "hopelessly deadlocked," (1981a, 5) and "a standoff." (1981b, 378)

Hopeless deadlock will be bad enough for present purposes, but I note in passing that Horgan (1981) argues persuasively for an even less promising conclusion, from the two-boxers' point of view. He notes an apparently ineliminable circularity in their attempt to *justify* two-boxing, turning on the

[34]The most important distinction here is the one marked by the double line. To the left of this line, evidence rules rationality, and modality plays a vice-regal role. (Evidence is the throne behind the powers, so to speak.) To the right of this line, modality rule rationality, and evidence defers.

fact that attempts at justification always return to the same kind of counterfactuals. One-boxers do better, Horgan argues, by confining their attention to deliberation about *actuality.*[35]

Why is even deadlock bad news, from a two-boxer's point of view? For a reason which Lewis himself puts his finger on, with respect to 'unHumean' theories of chance, whose proponents are more willing than he himself is to postulate metaphysical primitives:

> Be my guest—posit all the primitive unHumean whatnots you like. ... But play fair in naming your whatnots. Don't call any alleged feature of reality "chance" unless you've already shown that you have something, knowledge of which could constrain rational credence. I think I see, dimly but well enough, how knowledge of frequencies and symmetries and best systems could constrain rational credence. I don't begin to see, for instance, how knowledge that two universals stand in a certain special relation N* could constrain rational credence about the future coinstantiation of those universals. (1994, 484)

My EviCausalist makes the same demand of an account of causation:

> Be my guest – posit all the primitive whatnots you like. But play fair in naming your whatnots. Don't call any alleged feature of reality 'causation,' or 'counterfactual dependence,' unless you've already shown that you have something, knowledge of which could constrain rational deliberation.[36]

Had Lewis himself been in a position to meet this challenge to his own account of causation and counterfactuals, he would have had the key required to break the deadlock between one-boxers and two-boxers. The fact that he thought the deadlock hopeless therefore supports my contention that two-boxers occupy a position analogous to that of hardline primitivist objectivists about chance – a further manifestation, in my view, of a deep tension in Lewis's own view.[37]

---

[35]Cf. Price and Weslake (2009), who note that Lewis's "deadlock" reflects the difficulty that accounts of counterfactuals such as Lewis's have in explaining the connection between counterfactuals and deliberation. Like Horgan, Price and Weslake argue that we do better if we begin with non-counterfactual modes of deliberation – with "material deliberation," as they term it.

[36]Is this demand is compatible with the concessive policy I have recommended earlier, viz., that of allowing alternate choices about the use of these concepts in exceptional cases, provided it is conceded (as Lewis himself concedes for chance in cases of inadmissible evidence) that the usual ties with rational action are broken in these cases? Yes, provided we read the demand as calling only for an explanation of the concepts' connection with rational action in normal cases. (Lewis must take it this way in the case of chance, of course, if his own view is not to fail the test.)

[37]How might Lewis's version of CDT be modified to remove this tension? Horgan's account suggests a possibility. As we noted earlier, Horgan says that while he does "recommend acting *as if* one's present choice could causally influence the being's prior prediction, ... my argument does not presuppose backward causation." (1981, 340–41) Analogously, Lewis's CDT might be modified to say that in exceptional cases such as the classic Newcomb problem, we need not genuinely causal dependency hypotheses, but rather hypotheses sensitive to the dependencies revealed by inadmissible information and the like. As in Horgan's case, the modified account would "recommend acting *as if* one's present choice could causally influence the [Predictor's] prior prediction."

EviCausalism, by contrast, has precisely the advantages of Lewis's own subjectivism in the case of probability. By building its account of causality *on* deliberation, evidentially construed, it ensures that causality does not lose conceptual or practical touch with deliberation.

The Principal Principle can be regarded as a codification of the relation that something must bear to credence, to count as chance, or objective probability. As we might put it:

$\text{SDT}_{ch} = \text{SDT}_{ev} + \text{PP}$.

In other words, PP is the rationality condition one needs to add to SDT_{ev}, to produce the 'bundled,' two-experts-in-one theory represented by SDT_{ch}.

In the same spirit, the EviCausalist proposes that we should expect a codification of the relation that something must bear to conditional agentive credence, to count as causality – in other words, a rationality condition one needs to add to EDT to produce CDT (which is the two-experts-in-one version of conditional decision theory). What is this principle CP, such that

CDT = EDT + CP?

Essentially, it is the principle that in assessing one's agentive conditional credences for Outcomes given Acts, one should be guided by one's causal beliefs.

8 Random riches

Now to the task deferred above: using *Why Ain't you Rich?* as a point of agreement between EviCausalism and traditional Causalism, in order to argue that EviCausalism is as general as it needs to be – there are no nasty surprises, lurking around the corner.

In the standard Newcomb problem, two-boxers accept that one-boxers will get rich, and that they themselves will not. But the bare description of the case admits a variety of understandings of what actually underlies the prediction. In order to mark what will turn out to be an important distinction, I want to introduce the following variant. Let us suppose that it is proposed to allocate agents randomly to a one-boxer stream or a two-boxer stream (perhaps with an additional inducement, so that all parties agree that it is rational to play the game, agreeing thereby to follow the policy appropriate to the stream to which they are assigned); this random assignment to take place *after* the Predictor has allocated his cash to the boxes. The function of randomness is to guarantee that agents are epistemically neutral between the two options,[38] and we therefore stipulate that the device be the equivalent of what the players take to be a fair coin. (They believe that it makes assignment to each of the two streams equally likely, on other words.)

Will Causalists and Evidentialists still agree that the expected return (i.e., the expected V-utility) for the one-box stream is much greater than for the two-box

[38] As we shall see in a moment (§8.1), it thus simulates something that is true anyway, from the agent's perspective.

stream? In some cases, we may assume that they will. Again, our supernatural cases will serve as examples: imagine that the Predictor is believed to be a seer, who knows the agent's choice, even if it is the result of our random process. The details don't matter here, since for present purposes we simply need an example that falls on this side of the line, and it can be as unrealistic as it needs to be. But note that the more realistic we make a Predictor version of the Newcomb problem – by basing it on not-too-implausible extrapolation from the predictive powers of real predictors, perhaps – the less likely it is that the assumed correlation between one-boxing and wealth will still obtain in this random variant. Not-too-implausible predictors will fail dismally when the task is to predict random choices.

Similarly in the medical cases, presumably: there, Causalists will not expect that agents randomly assigned to the No Smoking stream will have a lower incidence of the cancer gene than those assigned to the Smoking stream. If they themselves are the players in this random game, then their own conditional credences of having the gene, conditional on being assigned to the No Smoking stream or to the Smoking stream, are identical. By the Causalist's lights, in other words, there is no *Why Ain't you Rich?* challenge to be answered in this case: those randomly assigned to decline a cigarette will be no 'richer' (i.e., healthier) on average than those who do not. On average, they will be worse off, once the denied pleasure of smoking is taken into account.

Somewhere between these two kinds of cases thus lies a boundary, by a regular Causalist's lights. On one side of the line, a randomly prescribed 'Evidentialist' choice leads to higher expected V-utilility. On the other side, it does not.[39] So long as EviCausalism can cleave to this same line, recommending *only* the choices that have higher expected V-utility in this random game, then when it differs from traditional Causalism, it will always be able to appeal to *Why Ain't you Rich?* (with the response outlined above in hand, to deal with the Causalist's objections).

The EviCausalist thus requires that her own agentive conditional probabilities of Outcomes given Acts (e.g., of Cancer given Smoking, in the Smoking Gene example) are the same as the corresponding conditional probabilities in the random case. So long as this equality holds, the EviCausalist will assign the same expected V-utilities to Acts in the two cases – that is to say, when she chooses the Act, and when the random device chooses the Act. Accordingly, her own decision policy will recommend a particular Act if and only if the Causalist agrees that that Act maximizes V-utility in the random game.

So the EviCausalist needs to show that she is entitled to ignore any apparent evidential dependency that wouldn't hold if her action were randomly chosen. How might this result be established? The most direct option would be to maintain that it is simply a primitive, *constitutive* fact about the free agent's point of view that she regards her actions as 'uncaused,' in such a way that she is automatically committed to the claim that any evidential dependency between

[39] As we observed a moment ago, semi-realistic versions of the original Newcomb problem will fall on the latter side of the divide.

her Actions and Outcomes would survive if her actions were randomly chosen (this being simply one way in which her choices may be uncaused).[40] In earlier work I proposed such a view of free action (though without the claim that it is primitive), and attributed it to Ramsey:[41]

> Ramsey [identifies] what he takes to be the crux of the agent's perspective, namely the fact that from the agent's point of view contemplated actions are always considered to be *sui generis,* uncaused by external factors. As he puts it, "my present action is an ultimate and the only ultimate contingency." [Ramsey 1978, 146] I think this amounts to the view that free actions are treated as probabilistically independent of everything except their effects. (1993, 261)

Similar views are defended by Hitchcock (1996) and Joyce (2007). Hitchcock's version perhaps comes closest to regarding this as simply a primitive feature of free action: he suggests that we might regard it as a kind of 'fiction,' central to our practice of regarding ourselves as free agents.

Can we do better than regarding this as a primitive feature of agency, fictional or not? The Tickle Defence and its descendants comprise a sustained attempt to do better; to show that an agent's evidential perspective is *guaranteed* to have this distinctive character, in virtue of differences between her own epistemic situation and that of external observers. As we noted in §6, the core of the Tickle Defence is the thought that an agent's information about her own beliefs and desires necessarily screens off, for her, any prior factors, such as the Smoking Gene, that might be correlated with her choice. If the argument works, it provides a kind of case-by-case defeater for evidential connections that would otherwise make her actions *unlike* randomly chosen actions (in being correlated with such prior states of affairs).

As we noted, Lewis himself offered an optimistic assessment of the prospects for this endeavour. In this context, where the task is to provide EviCausalism with a guarantee that there are no cases in which her own policy need differ from that of the random game, we need not be concerned about Lewis's remarks about the unsuitability of the Tickle Defence for imperfectly rational agents (where, as Lewis put it in the remark we quoted in §6.1, "rationality is taken to include self-knowledge.")[42] All the same, I think there is a sense in which the Tickle Defence puts the emphasis in the wrong place, and misses a more elegant and economical route to the *de facto* randomness that the EviCausalist requires.

[40] Readers familiar with the notion of an *intervention,* as employed by writers such as Pearl (2000) and Woodward (2003), may feel that it would be sufficient that we treat our actions not as 'uncaused,' but simply as having a causal history sufficiently independent of the events at issue. (Thanks to a referee here.) But this will not do for my EviCausalist, who wants to ground such causal judgements on evidential judgements. That said, however, it is certainly true that the notion the EviCausalist is after is very close in spirit to Pearl's and Woodward's notion of an intervention.

[41] I go on to suggest that this point be read "in reverse," so that we regard the effects of an action, as my EviCausalist now proposes, as those outcomes properly regarded as (positively) conditionally probabilistically dependent on the action in the context of deliberation.

[42] Though Lewis's caution about the coherence of free choice for a perfectly rational agent might lend support to fictionalism after all.

This route also turns on the special epistemic perspective of a deliberating agent, but in a less piecemeal manner.

8.1 The epistemics of deliberation

The new argument turns on the special epistemic authority of an agent's deliberations concerning her own actions. One recent writer who calls attention to this authority, and notes its potential to offer an alternative to the Tickle Defence, is Joyce (2007). Joyce himself takes the crucial point to be that "an agent's beliefs about her own free decisions and actions provide evidence for their own truth." (2007, 558) Such beliefs are "self-supporting," as he puts it.[43]

An alternative way to put this thought, preferable in my view, is to say that there is an important sense in which, as she deliberates, an agent simply *does not have* knowledge, beliefs or credences about the action in question. In this form, Joyce's thought corresponds to a familiar view, nicely characterized by Wlodek Rabinowicz in the following passage:[44]

> On this view, the relevant distinction is between the *first-person* perspective of a practical deliberator and the *third-person* perspective of an observer. While the observer can predict what I will do, I can't, insofar as I deliberate upon what is to be done. Deliberating in this way is incompatible with predicting the outcome of deliberation. To put it shortly, *deliberation crowds out prediction.* (2002, 91)

One route to this thesis, as to Joyce's version, turns on the special epistemic authority of the deliberating agent, concerning her own actions. This authority 'trumps' any merely predictive knowledge claim about the same matters, rendering it necessarily unjustified.

A familiar application and illustration of this point, in a superficially different guise, is Dummett's (1964) observation that an agent can coherently believe that she can *affect* some past state of affairs only if she takes herself to be unable to *know* whether the state of affairs in question obtains before she decides whether to perform the action she takes to be required to bring it about. Given retrocausality, any claimed knowledge about this matter could be 'bilked', as the familiar argument has it – that is, the agent could choose to act so as to defeat the knowledge claim in question.[45] Dummett's point applies equally to effects in any temporal relation to actions, of course. It is especially striking in

[43]Joyce makes these points in order to block an objection from Richard Jeffrey, to the effect that Newcomb problems are not really cases of free choice at all, because the agents involved know too much about their own actions. Joyce is thus defending Causalism against an Evidentialist objection. In my view, however, the point ultimately counts in favour of Evidentialism, by showing how the Evidentialist can justifiably ignore spurious evidential correlations: they fall into the category of evidence properly ignored by the deliberating agent.

[44]Rabinowicz himself opposes this view, which he attributes particularly to Spohn and Levi.

[45]The bilking argument is more familiar with the opposite orientation, pointing out that *given* such knowledge, the claim of retrocausal influence could be bilked. But it cuts equally well in either direction.

the retrocausal case only because we typically assume that we do have epistemic access, at least in principle, to states of affairs in the past. In the usual future-directed case, the presumption goes the other way. We assume that we do not have epistemic access, in advance, to the future effects of contemplated actions, or to the actions themselves. But again, if it were claimed that we did have such access, the claim could be bilked. The bilking argument simply reminds us that the agent herself holds the epistemic trump card. And Joyce's point – already implicit, I think, in Dummett's discussion – is that the character of deliberation mandates this assumption.[46]

What is the source of this special epistemic authority of the deliberating agent concerning her own actions? The most general proposal I know is that of Jenann Ismael, who argues that it is simply a special case of a familiar form of 'epistemic degeneracy,' as Ismael calls it, typical of self-representing representations:

> Alethic constraints . . . on representational activity are empty when applied reflexively, i.e., when what is being represented is the representational act itself. The most familiar examples of this degeneracy are self-representing linguistic performances: "I promise to X", "I declare that Y". Such performances are perfectly good representational acts. They have truth conditions that can fail to obtain; someone else can certainly falsely ascribe a promise to me, and I can misrepresent my own past promises and declarations. But because they provide their *own* truthmakers, they are unconstrained at the time that they are made. They are self-fulfilling. . . .
>
> [This] degeneracy is unavoidable for any system that includes its own activity in the field of representation. The desire to tell the truth in general will not guide my answer the question "Will I A?" and the ordinary epistemic procedures for getting information about whether [someone] A'd will not apply. Guidance has to come from elsewhere. . . .
>
> The emptiness, or degeneracy of alethic constraints . . . when applied to one's own actions opens up the space for deliberation. I believe that it captures the sense in which, from the point of view of the participant in a dynamical process, her own actions have the status of what Ramsey called "an ultimate contingency". (2007, §3)

As Ismael remarks, the epistemic authority of deliberation seems to explain the striking feature of action noted by Ramsey, viz., its apparent 'contingency,' from the agent's point of view. Once again, the point is easily made by adapting Dummett's condition for the coherence of a belief in retrocausality. An agent who took her own future actions to be *caused by* an earlier state of affairs of

[46] Is there a tension between the recognition that Dummett's point applies equally with respect to the future and taking seriously the possibility that we might have inadmissible information about the future, of the kind on which our Chewcomb problems depend? (Thanks to a referee for this question.) Answer: Not in general, but Dummett's observation does indeed imply that *some* kinds of inadmissible information about the future would be incompatible with the view that we had a choice about the matter in question. If I'm undecided about whether to go to a conference tomorrow, and think that I really have a choice in the matter, it is no use my checking my crystal ball to see if I'm there. In these circumstances, I must regard the crystal ball as unreliable. That's Dummett's point, applied to the future.

which she had knowledge, before she made up her mind what to do, would be in exactly the same incoherent epistemic position as Dummett's agent, who took herself to have knowledge of the past *effects* of a future action. (The difference between the two cases is simply the direction of the causal link, which makes no difference to what matters here, namely, the evidential significance of the link in question.) A little more generally, this argument shows that as she deliberates, a free agent cannot take her action to be correlated with *anything* of which she might in principle have knowledge, before she makes up her mind what to do. Again, any such correlation could be bilked.

8.2 Is this enough?

This appeal to the special epistemic situation of a deliberating agent provides much of what EviCausalism needs. It offers a strong argument, grounded on what is arguably an essential feature of deliberation, that many evidential correlations are properly ignored from a deliberating agent's point of view.

But does it go far enough? Couldn't there be some *non-causal* correlation between an agent's actions and some state of affairs of which she *could not have knowledge,* even in principle, as she deliberates? The epistemic inaccessibility of the state of affairs in question would then enable the correlation to survive under deliberation, from the agent's point of view, precisely as in Dummett's examples of coherent conceptions of retrocausality. But if it is not *really* a causal correlation, isn't the EviCausalist still in trouble?

The EviCausalist will agree that such cases are possible by the traditional Causalist's lights, but deny that they are possible by her own lights. On the contrary, she insists, such a correlation would automatically count as causal, by her standards. That's what causal dependency *is,* by her lights, after all: evidential dependency conditional on Acts, from the rational agent's point of view.

Moreover, since the correlation in question survives (by assumption) under deliberation, it survives in particular in the random choice case (which is simply a special case of deliberation, a choice to be guided by the outcome of a random event); which means that the traditional Causalist will have to admit that it leads to riches, when linked to suitable Outcomes. So it falls on the right side of the line, by the EviCausalist's lights.

If this argument works, it gives EviCausalism a guarantee that the conditional credences to which its version of EDT appeals yield the same assessments of expected V-utility as in the random game. From this point, the argument is straightforward. On one side, with the (strong version of the) classic Newcomb problem, EviCausalism recommends one-boxing, responding to the two-boxer's objection to *Why Ain't you Rich?* in the way described, and challenging traditional Causalists to offer a non-question-begging defence of the rationality of their own policy. On the other side, where we find the familiar medical problems, and less supernatural versions of Predictor cases, EviCausalism recommends two-boxing, or its equivalent – the same choice as traditional CDT.

In both cases, the EviCausalist stresses that by her lights – according to her understanding of causality, and her view of the probabilities required by EDT – CDT and EDT actually coincide. She recognizes that traditional Causalists understand the term 'causation' somewhat differently, and hence that by their lights, the cases in which EviCausalism recommends one-boxing are cases in which CDT differs from EDT. Her challenge to these opponents is to defend the claim that CDT remains rational, in these exceptional cases, if 'causation' is understood as they prefer. To meet this challenge, traditional Causalists need a response to the *Why Ain't you Rich?* objection – but now in the hands of an opponent who, unlike meeker traditional Evidentialists, is not prepared simply to concede the Causalist her counterfactuals.[47]

9 Conclusion

We have covered a lot of ground. I close with a summary of the main points, and some remarks about the limits of the present conclusions.

9.1 Summary

(i) The Chewcomb problems reveal a significant tension in a popular combination of views (a combination exemplified by Lewis himself, amongst others) concerning the rational practical significance of exceptional evidence, in the case of chance on the one hand, and causation on the other.

(iii) The tension can be resolved by adopting the same degree of subjectivism with respect to causation[48] that Lewis adopts with respect to chance – accepting that causation, too, has its roots in evidential decision making, at least in the sense that nothing deserves the name causation, unless we can explain its relevance to decision. For causation as for chance, the required degree of subjectivism is nicely captured by the proposal that the modal notion in a question is an expert, intended to represent ideal evidential practice.

(iv) As in the case of chance, there are two ways to develop this thought, which differ in their treatment of certain cases of exceptional evidence. For Lewis, chance is not an infallible expert, and is rationally set aside by someone who believes herself to have inadmissible evidence. For Hall, there can be no such cases: no evidence is inadmissible, from chance's point of view. The difference is largely a matter of taste, and the two views agree about the rational credences, in the exceptional cases. Similarly in the case of causation, in circumstances

[47] Not, at least, until the Causalist concedes that counterfactuals need not guide rationality, in exceptional cases.

[48] Or better, perhaps, the *concept* of causation. As noted in §2, I have been setting aside this distinction, in order to get on the table the main argument and the analogy with chance on which it relies. But the analogy holds here, too, in my view, and it is appropriate to ask whether Lewis's subjectivism is best thought of as an account of chance itself, or rather as an account of the *concept* of chance. I myself favour the latter reading, but shall not argue the point here.

such as the classic Newcomb problem. The Hall-like view treats these as cases in which the causal structure is abnormal (e.g., in involving retrocausality). The Lewis-like view treats them as cases in which rational choice does not follow CDT. But the two views agree on the rational policy: it is to one-box.

(v) If we go Hall's way in the case of causation – the EviCausalist proposal, as I have called it – then EDT and CDT now coincide everywhere. But this does not imply that CDT need endorse the equivalent of one-boxing (i.e., not smoking, in the Smoking Gene problem) in medical cases, or relatively realistic versions of the Predictor case. There, someone who believes that EDT recommends one-boxing is simply someone confused about the proper evidential bearing of her actions, as she deliberates. In normal cases, in which the causal structure is uncontroversial, our knowledge about it provides our best protection against such confusion; causation being precisely the expert we need to consult, in order to get these evidential judgements right.

(vi) This alignment between CDT and EDT entails that the usual Causalist response to *Why ain't you rich?* is powerless. Unlike conventional Evidentialists, the EviCausalist rejects the Causalist's claims about the relevant counterfactuals (insisting that had she two-boxed, she would have been nearly $1,000,000 poorer, not $1,000 richer, in the classic Newcomb problem).

(vii) At this point, the traditional Causalist will wish to defend her own reading of the counterfactuals, in the form of some objectivist rival to the EviCausalist's account of causal dependence. But such views are vulnerable to the charge that Lewis himself makes against analogous views of chance: what they offer us does not deserve the name causation, unless the Causalist can explain its relevance to rational deliberation. (And if the Causalist could do that, she would already have a response to the EviCausalist.)

(viii) Viewed by these lights, the classic Newcomb problem is pathology of rational deliberation, induced, to a significant degree, by excessive objectivism about causality. By obscuring the practical foundations of causal thought, this objectivism makes it hard to see that the Newcomb puzzle presents us with what amounts to conflicting information about the causal structure of a decision problem. On the one hand we are told, or simply infer by normal standards, that the agent's choice does not affect the contents of the opaque box. How could it do so, we are expected to think, if the money has been there (or not) since some time in the past? (Normal standards exclude retrocausality.) On the other hand we are given information about evidential dependence, which, combined with the assumption that we do genuinely have a choice, leads us by *different* normal standards to the conclusion that we can affect the contents of the opaque box – it is in our gift, as it were. Objectivism obscures the fact that we have a clash between two normal criteria for causality, not a clash between criteria for causality and something else entirely.

(ix) Once seen in the former terms, as a clash between two criteria for causality, the puzzle's intractability is easier to understand: our causal intuitions are

simply pulling us in two different directions, *because* the two criteria conflict. As in other cases of conceptual conflict, we then have a choice to make, about how to use the disputed notion in these strange cases. We can certainly choose to privilege the non-evidential factors, if we wish. (Again, see fn. 31 for an example of such a terminological choice.) But in that case we have no good grounds to insist that causality always constrains rational action. The circumstances driving the need for a terminological choice are sufficient to call that link into question; and, by choosing to be guided instead by the practical criterion, the EviCausalist automatically has the upper hand in the resulting dispute.

(x) A secondary contributing factor to the power and longevity of Newcomb's puzzle, if this diagnosis is correct, is the subtlety of the special epistemic status of the deliberating agent, as needed to ground an adequate subjectivist view of causation, immune from these illusions. The rough shape of the terrain is relatively familiar, the key feature being that fact that deliberation 'crowds out' prediction, as we put it earlier. But a detailed elucidation of this idea, in a well worked-out model of the epistemic dynamics of deliberation, remains a work in progress.

9.2 Limitations

I want to emphasize, first, that I am not proposing that EviCausalism offers a one-stop solution to all decision puzzles in the Newcomb tradition. The alignment EviCausalism allows between causal dependency and agentive evidential dependency does not imply, by any means, that it will always easy to determine what these dependencies are, in a particular decision problem. EviCausalism tells us that they lie in the same place, but not in *which* place – and that, of course, can still be hard to determine.

Among the hard cases we should expect various variants of the Predictor version of the Newcomb problem. I have been taking for granted that the classic version of the problem is one in which Evidentialism really does recommend one-boxing. But as I noted, it is easy to construct variants (based, for example, on not-too-improbable extrapolations from the capacities of real predictors) that lie on the other side of the line – in those cases, as in the medical cases, I take it that Evidentialism properly recommends two-boxing. In between, we may well find subtle cases that are hard to classify. Random experimentation remains the best guide, but it is not foolproof. We can never be certain that we have genuine randomness, for there might always be a lurking common cause, of which we are unaware. Nor can we be sure that the random mechanism itself does not have perturbing effects.[49] No matter – we inch our way forward, as in science in general, prepared always to retreat if necessary. The difference that EviCausalism makes is simply that as we do so, causal dependence and agentive evidential dependence keep step. Hypotheses about one are hypotheses about the other.

[49] As it does in Nozick's own presentation of the original case, in which the Predictor penalizes agents who choose by random means.

A related difficulty which survives EviCausalism is that it may be unclear whether a proposed decision problem is really a *decision* problem at all – that is, do its constraints really allow us to regard ourselves as making a choice? Here EviCausalism presumably has some bearing, for it reduces the potential parameters of a decision problem in a new way. But it seems unlikely to eliminate such puzzles altogether. As an extreme example, think of the variant of the classic Newcomb problem in which both boxes are transparent. Can an agent believe the usual story about the evidential significance of one-boxing, and yet believe that she has a choice in the matter? For a EviCausalist this is the same puzzle as a claimed case of causation in which the effect of a contemplated action is known in advance. But reducing two puzzles to one is not the same as eliminating them altogether.

This question connects with one which is both deeper and broader, and raises potential challenges to EviCausalism, that of the status of agency itself. As we have seen, the EviCausalist relies heavily on the idea that the epistemic viewpoint of an agent is distinctive in certain ways. Roughly, it requires that agents see their own actions as 'uncaused,' at least in the midst of deliberation about those same actions. This not only binds the fate of the EviCausalist, at least in some sense, to that of free will. It also means, potentially even more uncomfortably, that EviCausalism becomes a rope that binds *causation* to the fate of free will – no problem, perhaps, if these notions turn out to share the same fate, but a problem if they do not.

EviCausalism thus has a stake in some large issues about the metaphysics of causation, and related matters. And these issues in turn raise some general questions, of interest on all sides, which relate to our strategy of comparing chance and causation. Why has subjectivism seemed less attractive in the case of causation than in case of chance, for example? And is the difference well-grounded?

So various philosophical puzzles remain, in the vicinity of the Newcomb problem, even if we accept the EviCausalist proposal. Nevertheless, EviCausalism offers a solution to the central puzzle of the case: it mends, and ends, the strange divergence of causation and evidence that lies at the conflicted heart of Newcomb's famous problem. And it brings an attractive unity to chance and causation, extending the pragmatism of Lewis's treatment of the former to a treatment of the latter. It would be pleasing if it were true.

References

Dummett, M. A. E. 1954. "Can an Effect Precede its Cause?." *Proceedings of the Aristotelian Society Supplementary Volume* 38: 27–44.

———. 1964. "Bringing about the Past." *Philosophical Review* 73: 338–59.

Eells, Ellery 1981. "Causality, Utility, and Decision." *Synthese* 48: 295–329.

———. 1982. *Rational Decision and Causality.* Cambridge: Cambridge University Press.

———. 1984. "Newcomb's Many Solutions." *Theory and Decision* 16: 59–105.

Gaifman, H. 1988. "A Theory of Higher Order Probabilities." In *Causality, Chance, and Choice,* ed. Brian Skyrms and William Harper, 191–219. Dordrecht: D. Reidel.

Gibbard, Allan and Harper, William 1978. "Counterfactuals and Two Kinds of Expected Utility." In *Foundations and Applications of Decision Theory,* Vol. 1, ed. C. A. Hooker, J. J. Leach and E. F. McClennen, 125–62. Dordrecht: D. Reidel.

Hall, Ned 1994. "Correcting the Guide to Objective Chance." *Mind* 103: 505–17.

———. 2004. "Two Mistakes about Credence and Chance." *Australasian Journal of Philosophy* 82: 93–111.

Hitchcock, Christopher 1996. "Causal Decision Theory and Decision-Theoretic Causation." *Noûs* 30: 508–26.

Horgan, Terry 1981. "Counterfactuals and Newcomb's Problem." *Journal of Philosophy* 78: 331–56.

Horwich, Paul 1985. "Decision Theory in Light of Newcomb's Problem." *Philosophy of Science* 52: 431–50.

Ismael, Jenann 2007. "Freedom, Compulsion, and Causation." *Psyche* 13.

Jeffrey, Richard C. 1965. *The Logic of Decision.* New York: McGraw-Hill.

———. 1981. "The Logic of Decision Defended." *Synthese* 48: 473–92.

Joyce, James 2007. "Are Newcomb Problems Really Decisions?." *Synthese* 156: 537–62.

———. 2011. "Regret and Instability in Causal Decision Theory." *Synthese* Online first. dx.doi.org/10.1007/s11229-011-0022-6.

Lewis, David 1979. "Prisoners' Dilemma is a Newcomb Problem." *Philosophy and Public Affairs* 8: 235–40.

———. 1981a. "Causal Decision Theory." *Australian Journal of Philosophy* 59, 5–30.

———. 1981b. "'Why ain'cha rich?'" *Noûs* 15: 377–80.

———. 1982. Letter to Wlodek Rabinowicz, 11 March 1982. Accessible in an appendix to Huw Price, "The Lion, the 'Which?' and the Wardrobe – Reading Lewis as a Closet One-Boxer." philsci-archive.pitt.edu/4894/.

———. 1986 [1980]. "A Subjectivist's Guide to Objective Chance." In *Philosophical Papers,* Vol. II, 83–132. New York: Oxford University Press (originally published in Studies in Inductive Logic and Probability, Vol. II, ed. Richard C. Jeffrey, Berkeley: University of California Press).

———. 1994. "Humean Supervenience Debugged." *Mind* 103: 473–90.

Nozick, Robert 1997 [1969]. "Newcomb's Problem and Two Principles of Choice." In *Socratic Puzzles,* Cambridge, MA: Harvard University Press, 45–73 (originally published in *Essays in Honor of Carl G. Hempel*, ed. Nicholas Rescher, Dordrecht: D. Reidel, 107–33).

Papineau, David 1996. "Many Minds Are No Worse Than One." *British Journal for the Philosophy of Science* 47: 233–41.

Pearl, J. 2000. *Causality.* New York: Cambridge University Press.

Price, Huw 1986. "Against Causal Decision Theory." *Synthese* 67: 195–212.

———. 1991. "Agency and Probabilistic Causality." *British Journal for the Philosophy of Science* 42: 157–76.

———. 1993. "The Direction of Causation: Ramsey's Ultimate Contingency." In *PSA 1992, Volume 2,* ed. David Hull, Micky Forbes and Kathleen Okruhlik, 253–67. East Lansing, Michigan: Philosophy of Science Association.

Price, Huw and Weslake, Brad 2009. "The Time-Asymmetry of Causation." In *The Oxford Handbook of Causation,* ed. Helen Beebee, Christopher Hitchcock and Peter Menzies, 414–43. Oxford: Oxford University Press.

Rabinowicz, Wlodek 2002. "Does Practical Deliberation Crowd Out Self-prediction?" *Erkenntnis* 57: 91–122.

Ramsey, F. P. 1978. "General Propositions and Causality." In *Foundations: Essays in Philosophy, Logic, Mathematics and Economics,* ed. D. H. Mellor, 133–51. London: Routledge and Kegan Paul.

Savage, Leonard 1954. *The Foundations of Statistics.* New York: Wiley.

Skyrms, B. 1980. *Causal Necessity.* New Haven: Yale University Press.

Van Fraassen, Bas 1989. *Laws and Symmetry.* Oxford: Oxford University Press.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation.* Oxford: Oxford University Press.