Freedom, Fiction and Evidential Decision Theory

Phyllis Kirstin McKay

Erkenntnis (2007) 66:393--407

Abstract

This paper argues against evidential decision-theory, by showing that the newest responses to its biggest current problem – the medical Newcomb problems – don't work. The latest approach is described, and the arguments of two main proponents of it – Huw Price and CR Hitchcock – examined. It is argued that since neither new defence is successful, causation remains essential to understanding means-end agency.

1 Coco and the Evidential-Causal Debate in Decision Theory

Suppose Coco is trying to decide whether to accept a Mars Bar offered to him. Coco knows that he often suffers from migraine soon after eating chocolate. But Coco is also aware that he often suffers from a pre-migrainous state, ***PMS***, that causes him *both* to crave chocolate, *and* suffer from migraine. Should Coco accept chocolate?

A causal decision theorist says 'yes'. Causalists hold that agents should act only when they have reason to believe that their action is likely to *cause* the desired outcome. In the Coco case migraine is merely correlated with chocolate eating. Chocolate eating will not cause migraine, so Coco should go ahead and eat chocolate if he wants it.

It seems an evidential decision theorist must say 'no'. The evidentialist advocates choosing possible actions according to whether they are *evidence* or not for a particular desired result. Its claim is that an agent should choose those actions that render the result they desire most probable, or at the very least more probable if they perform the action than if they don't. But it is more likely that Coco will get migraine

if he eats chocolate than if he doesn't, because his eating chocolate is evidence for him *already being* in PMS. So Coco should refuse chocolate as a means to avoiding migraine.

It is almost universally agreed that Coco should have his chocolate, so this case is presented by causal decision theorists as a damaging counterexample to evidential decision theory. In general, this problem for evidentialists appears in every case where the desired result and the contemplated action share a common cause outwith the control of the deliberating agent. These are the medical Newcomb problems.

Causalists claim that the medical Newcomb problems show that causation is fundamental to agency, and hence that an evidential decision theory could never be adequate. This debate between causal and evidential decision theory is interesting because it explores important implications for the nature of agency, so long as decision theory is seen as a genuine attempt to model agency, to understand what agents are doing and ought to do in order to be successful in action.[1]

Evidential responses to the medical Newcomb problems have been attempts to show that evidential decision theory is not committed to prescribing that Coco reject chocolate. The favourite past defence has been the tickle defence. Very briefly,

---

[1] The causal-evidential debate is fundamental to this understanding. James Joyce argues that all decision theory is either causal or evidential, whether or not it is made explicit. All mathematical formalisations in decision theory come with tacit principles about what sort of decision problem you can legitimately apply the mathematical formalisation to. For example, certain formalisations are inappropriate for situations in which the probability of the background state of the world varies according to the action being chosen. Joyce says that causal decision theory is concerned with investigating these tacit principles, and that all decision theory needs to pay attention to them. Once a decision theory does that, Joyce thinks it will expose itself as either evidential or causal. (Joyce, especially section 1) I agree. The issue of whether you can describe agency accurately without reference to causes is vital to understanding agency. This is particularly interesting to me because I want to know whether you can have a genuinely non-circular *agency* theory of causation, but the issue is significant for other reasons too. For example, Nancy Cartwright thinks that decision-problems like this are actually evidence for the existence of deeper causal relations. (Cartwright Ch 1) She takes the need to refer to causation in correctly characterising agency as a key reason for holding a realism about causation.

tickle defenders claim that prior causes of an agent's actions always act via the beliefs and desires of the agent. Further, a rational agent should know his own beliefs and desires. But once the desires of the agent are set, these screen off the contemplated action from prior causes. For Coco this means chocolate is no longer correlated with migraine and he can go ahead and have his chocolate. But this defence has been thoroughly discussed in the literature.[2] In this paper I want to examine two new defences, produced by Huw Price and CR Hitchcock, who are dissatisfied with the tickle defence. Their defences have not yet been appropriately answered.[3]

This new idea is that rational agents must take up a certain perspective on their own deliberation. It seems to derive from the famous Kantian idea that an agent has to regard him or herself as *free* to see him or herself as genuinely an *agent*.[4] Price talks about taking up the 'agent's perspective' and evaluating probabilities from it. These "agent probabilities" are then the ones properly used in deliberation. Hitchcock thinks in deliberation we must "simulate the independent of our actions from factors beyond our control." (Hitchcock, p. 522.) There seems to be something right in the idea that an agent must view him or herself as in some sense free. Should you happen to be locked in a cage and miserably aware of this fact, considering whether to go shopping, or to see a film, might be as good a way to pass the time as any other, but it is not genuine deliberation. Here I am going to examine the detailed strategies used by Price and Hitchcock, show that their views are in fact quite different, and argue that they don't succeed in spite of their intriguing origin. Considering causes is essential to successful means-end agency.

2 Price and Wide-Ranging Generalisations

Price writes that the trouble with the tickle defence is "there is no guarantee that the effect of a physiological state on choice behaviour which gives rise to a medical

2 See for example Eells (1982, 1985).

3 Papineau discusses Hitchcock briefly, but there is more to say.

4 See for example the Groundwork of the Metaphysic of Morals part III where Kant seems to derive morality from freedom. The relationship seems to reverse in the Critique of Practical Reason, but that is a matter for Kant scholarship.

Newcomb problem will be mediated by any identifiable craving or tickle." (Price (1986, p. 203.) And I take it that it is this realisation which leads Price to attempt a different sort of defence of evidential decision-theory in the face of the medical Newcomb problems in two papers (1986 and 1991).[5]

Price claims in both papers that the crux of the problem is the question of when to apply statistical generalisations to individual cases, and this is the basis of his approach.[6] His idea is to save evidential decision theory by arguing that, in the crucial cases, the general correlation between eating chocolate and being in ***PMS*** or having migraines is not applicable to Coco, and Coco will always be in a position to realise this crucial fact.

It is certainly true that we sometimes have reasons not to apply generalisations to particular cases, and it is also perfectly possible that Coco have such a reason because of something he knows about his own deliberation. For example, if you are Coco, and you make a decision to eat chocolate with your dinner only and always when a tossed coin lands 'Heads , you can be fairly sure that next Tuesday, when dinnertime comes along and you eat your chocolate, this doesn' t provide extra evidence that you 're in ***PMS*** and are likely to suffer from migraine later on that evening. While there is a correlation between eating chocolate and migraine in the general population, there certainly isn t one between coin tossing and migraine. Evidentialists don' t have to say that Coco must act on an evidential correlation even when he has good reason to believe it doesn 't apply to him, or, even more narrowly, usually applies to him, but doesn 't for this particular decision under consideration. This is just a case of the

---

[5] In the 1991 paper, p. 64, he says that his argument undermines the need for a conceptual distinction between causal and evidential views in decision-theory. My arguments will suggest why an evidentialist might think this. But to a causalist there is a distinction, and Price falls in the evidentialist camp. I have chosen to use the Coco example because it is Price s formulation of the medical Newcomb problem for evidential decision theory. He makes it clear that he thinks evidential decision theory needs to allow Coco his chocolate: ''Whether Coco is in PMS or not ... accepting the Mars Bar will not make any difference. Its evidential bearing ought to be counted irrelevant to his decision'' (Price 1991, p. 162).

[6] He writes: ''as objections to evidential decision theory, medical Newcomb problems depend on the claim that spurious correlations translate into spurious evidential dependencies between contemplated actions and other events'' (Price 1991, p. 161). Later he writes: ''At the heart of the problem is the issue of the applicability of statistical generalisations to individual cases'' (Price 1991, p. 164).

old principle of using the most information available, or the narrowest possible reference class.[7]

Price thinks Coco would be mistaken in believing that eating this Mars Bar now makes it more likely that he is in PMS, and that a parallel mistake would apply to any medical Newcomb problem. To make this mistake, Price says Coco:

> has to overlook the fact that his choice to decline is fully explained by the very judgement in question (namely the judgement that if he were to decline, that would constitute evidence that he is not in PMS). This judgement has provided him with a reason to decline which is quite independent of whether he is in PMS. Given this judgement, anyone with the same background beliefs and desires would make the same choice, regardless of whether he or she was in PMS. Given that Coco has made this judgement, in other words, he now has no grounds for taking it to be true. Remember that it rested on his belief about the correlation between PMS and chocolate consumption; and its effect is to destroy that correlation. In similarly motivated agents who make this judgement, PMS is simply irrelevant to their resulting decision to choose chocolate. (Price 1991, p. 163.)[8]

So Price 's idea is that the motivation to decline out of fear of migraine is itself enough to tell Coco the problematic evidential correlation no longer applies to him. He writes: ‘‘He [Coco] has no reason to think that in cases in which he has such a motivation [i.e. fear of migraine] for eating or declining chocolate, there is any correlation between Mars Bars and PMS.’’ (Price 1991, p. 166.) Coco' s contemplated decision is based on a judgement about accepting chocolate being evidence for ***PMS***, but Price 's thought is that if a decision to decline chocolate is made on the basis of this correlation that usually applies, in *that instance* the correlation will *not* apply.

It seems to me that this will work in some special cases, but it is not generally true. If Coco s deliberations and motivations are very clear to him, there might be some occasions when he knows that his only reason for worrying about whether or not to eat chocolate is fear of the correlation between eating chocolate and migraine. In such a case, Coco would be entitled to reason that his decision about chocolate eating was

---

[7] Price doesn' t put the point this way in his 1991, but he does make the connection fairly briefly in his 1986.

[8] Price gives what seems to be the same argument again, but phrased in terms of the causal history of the action under consideration, at p. 166.

not being affected, in this instance, by PMS, it being ‘‘fully explained’’ by fear of the correlation. And in this instance, the correlation will no longer hold of Coco.

But this is going to be a very unusual situation; Coco is seldom going to know all this. Decision-making can be very complex. Coco has at least one reason for not accepting chocolate, fear of migraine, but he also has a reason for accepting chocolate, especially if in ***PMS*** -- he wants it! Given this desire, people make and stick to the decision to accept chocolate significantly more often when in ***PMS***, and make and stick to the decision to decline chocolate significantly more often when not in PMS, whatever they actually take their reasons to be. This is why the correlation exists. But these effects on their reasoning are not always immediately transparent to the agents concerned. On the contrary, human action is subject to many causes, some of which act via deliberation in a clear way, while some of them are comparatively opaque. For example, it is a well-known fact that when it comes to the intricacies of human relationships, especially close family relationships, many people' s behaviour can be affected by emotions that they may not be aware of, emotions like fear and jealousy. Sometimes people are aware of such emotions, of course, but sometimes they are not. We also find opaque deliberation in cases of self-deception, which can be simple and common. Suppose I interrupt my work to go to the local shop to buy milk. I succeed in buying milk but I am grumpy all the way home because they have sold out of Mars Bars. It doesn' t become clear to me until I am home that my need for milk wasn' t pressing enough to take time out to go to the shop. My real reason for walking to the shop was a suppressed desire for chocolate, affecting my behaviour in a way not immediately transparent to me.

So in cases like these, desires affect an agent s behaviour in a way the agent may find impossible to identify accurately. I think it would be a very unusual agent who was self-aware enough to follow all these complexities, and be in a position to know precisely what has ‘‘fully explained’’ her action in every case. So it will not generally be true that Coco is in a position to know that fear of migraine fully explains his choice to decline chocolate.

Price anticipates this response. He realises the causalist will claim that the correlation continues to apply to Coco in some cases. Price says this commits the causalist to

Coco believing ‘‘Whatever my reasons for consuming or not consuming chocolate, there is a positive correlation between my doing so and pre-existing PMS.’’ (Price 1991, p. 167.)

But the causalist doesn' t have to claim anything so strong, or so implausible. If Coco does have good reason to believe an evidential correlation does not apply to him in a particular case, he should ignore it. I have already described the coin-tossing case above, which alone would violate such a belief. The causalist only needs a weaker claim for Price s arguments to fall; he only has to insist that there is a correlation applying to Coco across cases sufficiently like the case where he decides ‘‘normally’’. So long as it is true that Coco chooses to go ahead with chocolate significantly more often when he is in ***PMS***, then his decision and action remains evidentially relevant to the existence of the prior state. The possible existence of cases using artificial random decision procedures like coin tossing, or extremely unusual situations like gun-to-head eat-or-I-shoot threats where the correlation fails to apply does not help. The usual case for agents deciding in medical Newcomb situations are not enough like the coin tossing case for the correlation to cease to apply, at least not for all agents in all possible medical Newcomb cases. The evidentialist is committed to giving the wrong prescription for action in at least some cases.
It is Price who must claim something implausible for his arguments to work. He needs an implausible picture of the reasoning of rational agents, just as the proponent of the tickle defence does. For Price, it would have to be true that the particular line of reasoning that Price describes was the only cause of agents decisions and actions, and agents also know this. It would have to be false that agents could be affected by other lines of reasoning that may or may not be clear to them. More opaque effects on agents deliberations, such as I have described, would have to be impossible.

And so at root the problem with Price 's defence is the same as the problem with the tickle defence. The tickle defence demands that, if you are a rational agent, your motivations are always clear to you, putting you in a reference class where the problematic evidential dependency no longer applies, but many philosophers reject that. Price knows this; it is why he too rejects the tickle defence. But Price 's defence makes analogously implausible claims, about the transparency of deliberation. I have argued that agents deliberation does not have to be transparent for it to be counted

rational – particularly when this is confined to being means-end rational. Rational agents can be affected by motivations whose effect on their decision-making is not fully transparent to them. They can be recognised as rational so long as they are engaging in successful means-end behaviour. Consider the simple case of self-deception I described above. Now, whether or not I realise why I am going to the shop, it is clearly a sensible means to get chocolate, to satisfy the suppressed desire. This behaviour is means-end rational, and so is the behaviour of the agents Price describes, even when they cannot be sure that their decision is the only cause of their action. Whether such behaviour is rational in any wider sense is not relevant to a decision theory, which only attempts to characterise behaviour in pursuit of given ends.

Price does not explicitly discuss rationality in this way, but it looks as if he is attracted to some sort of picture requiring strong claims about the transparency of deliberation, and that he thinks that when a particular line of deliberation is a cause of an action, there is no other cause of that action, not even other lines of deliberation. This picture would lead him to a strong claim he makes: ‘‘From the agent 's point of view probabilistic relevance and causal relevance cannot diverge. To introduce the agent is in effect to assume an independent causal history to the event A’’. (Price 1991, p. 169.) The event A is the action of the agent. But this cannot be right. The event A could only be assumed to have an independent causal history if we have reason to believe the decision of the agent is entirely independent of relevant causal factors such as PMS. We might believe this in some cases, where the agent is sure that the effects of prior factors on her deliberations are transparent to her, but we couldn' t possibly believe it true for every case of a deliberate human action in a medical Newcomb case.

So Price cannot plausibly claim that human deliberation and decision is fully transparent and that one clearly identified reason for action excludes other causes of action in the way he requires for his arguments in defence of evidential decision theory to be sound. There will still be cases of rational agents deciding in medical Newcomb cases where Price s defence of evidential decision theory cannot work, because the problematic spurious evidential dependency continues to apply. Evidential decision theory will still sometimes prescribe counterintuitive choices.

3 Hitchcock and Fictional Distributions

Hitchcock (1996) is engaged in a wider project to explore parallels between problems in probabilistic theories of causation, and problems in decision theory. He focuses on the fact that both use conditional probabilities evaluated within cells of what he calls the ‘‘c-partition’’. The ‘‘c-partition’’ is constructed by causalists when they want to describe their decision theory probabilistically. It is constructed to distinguish correlations that indicate a causal link between action and desired result -- which are good to act on – from correlations that are spurious since they indicate no such causal link and should not be acted on. Hitchcock notes that a parallel partition is important to probabilistic theories of causation, also to distinguish between correlations that indicate causal relations, and merely spurious correlations.[9] But in the section of the paper I am concerned with Hitchcock is looking for a *non-causal* rationale for using the c-partition, and what he gives amounts to a defence of evidential decision theory.

Hitchcock clearly likes Price’s work and in a certain sense his work is an extension of Price 's. He writes:

> The Ramsey-Price proposal strikes the right balance: it incorporates the insight that there is something funny about deliberating about actions that one knows to be caused by external factors, without withholding decision theory from agents that occasionally have their actions so caused. The idea is that rational deliberation demands of an agent that she entertain the *fiction* that her actions are *sui generis*; deliberation is the determination of what an agent would do if she were completely free to act in accordance with her interests. (Hitchcock p. 520. Emphasis in original.)

So Hitchcock shares Price’s thought that there is something special about the agent’s own perspective. Nevertheless Hitchcock then diverges radically from both Price’s

[9] The general idea is numbingly familiar to decision theorists and those working on probabilistic theories of causation. Hitchcock explains: ‘‘In Eells’’ theory [of causation], the causal relevance of C for E depends upon conditional probabilities within cells of a partition {B1, B2, ...}, which we will call the c-partition for C and E. Each cell Bi corresponds to a causally homogeneous background context. C is a (positive) cause of E if and only if P(E|C^Bi)>P(E|~C^Bi) for each cell Bi; C is a negative cause of E if and only if P(E|C^Bi)<P(E|~C^Bi) for each cell Bi; and C is a mixed cause of E if one of the inequalities P(E|C^Bi)>P(E|~C^Bi) or P(E|C^Bi) <P(E|~C^Bi) holds for some cell Bi, but not for all cells□“ (Hitchcock, p. 510).

work and the tickle defence. Hitchcock is not interested in information that puts you in some reference class or other, because he is not interested in reference classes *at all*. Unlike either Price or the tickle defence, Hitchcock is not concerned with agents acting on a *real* probability distribution. What Hitchcock does is construct a *fictional* probability distribution based on explicit assumptions that he derives from the Ramsey-Price proposal, and claim that agents should act on an evidential decision rule, using this fictional probability distribution.[10]

Formally, this seems to work. Hitchcock describes how to construct his fictional probability distribution, Pf, and, in the appendix to his paper, proves that when reasoning evidentially using it an agent will arrive at the right prescriptions for action in the medical Newcomb cases. That is, reasoning evidentially using Hitchcock 's fictional probability distribution is *formally equivalent* to reasoning using the standard causal criterion, and the real probability distribution. I don' t want to dispute this.

Nevertheless, this isn' t enough to defend evidential decision theory. What is needed is some reason for thinking that the reasoning Hitchcock describes is a *good* way of thinking about rational mean-send reasoning. After all, there could be indefinitely many systems for constructing a fictional probability distribution plus wrong criterion

---

[10] Hitchcock is quite explicit about this distribution being fictional. He writes that the assumption he advocates, the assumption the fictional distribution models, is an ‘‘assumption of a freedom of action which may, as a matter of fact, not exist’’ (Hitchcock, p. 523). In the light of this, it is worth noting that I have generally assumed some external constraints on probabilities used in decision theory. I have not concerned myself with the debate over objective versus subjective probabilities in decision theory because it is a vexing debate tangential to the issues here. In a nutshell this is because both Price and Hitchcock accept that the existence of a genuine correlation in the general population between chocolate eating and migraine, and Coco 's beliefs about that correlation, is what generates the problem in the first place. In more detail, Price is inclined to favour subjective probabilities in decision-making,

but he isn t aiming to convince only those who agree with him on that. He also presumably thinks subjective probabilities in some sense track more objective facts, since the entire focus of his argument is on how Coco comes to know that the correlation really does not apply to him. Hitchcock makes it quite clear he thinks rational degrees of belief will be guided by beliefs about objective probabilities. He writes ‘‘In other words, our beliefs about the objective probabilities that constitute the various relations of causal relevance will guide our deliberations’’ (p. 518). Hitchcock 's fictional distribution is different, of course. I do dislike the very idea. It would take a very strong argument to persuade me that it could be a requirement of rationality that you act on probabilities even when you know they are false, and it may be quite impossible to believe a falsehood, rather than merely pretend that a falsehood is true. Nevertheless, these views are incidental to my criticism of Hitchcock, which focuses on a possible rationale for the assumption of freedom he recommends.

for reasoning that nevertheless yielded correct prescriptions for action, because *together* they were formally equivalent to the causal criterion plus the correct probability distribution. For Hitchcock 's work to be a convincing defence of evidential decision theory, there needs to be an argument that using the fictional probability distribution plus the evidential criterion is the right way to think of means-end reasoning. There must be a rationale for using the fictional distribution, and one that is more convincing than that for using real probabilities and a causal criterion.

Hitchcock seems to think there is a further story. He says that construction of this fictitious distribution is intended to formalise Price 's work.[11] He also summarises his position: ‘‘we evaluate our actions in terms of these probabilities because they allow us to simulate the independence of our actions from factors beyond our control. Rational deliberation requires that we view our actions in this way, even if only in fiction’’ (Hitchcock, pp. 521--522). I am going to disentangle two ways in turn of constructing the necessary rationale for using the fictional distribution from Hitchcock' s work, and argue that neither is convincing. The first fails because it is implausible that we should take rational means-end action so, while the second cannot establish the necessary formal equivalence with the causal criterion plus the correct probability distribution.

So what kind of ‘‘independence’’ or ‘‘freedom’’ is Hitchcock 's fictional distribution meant to simulate. Hitchcock characterises his view: To evaluate her own actions, the agent must ‘‘entertain the fiction that her actions are *sui generis*, which, following Price, we will take to mean probabilistically independent of factors that are beyond her control.’’ (Hitchcock, p. 521.) This and remarks above and elsewhere in Hitchcock s paper strongly suggest the first kind of freedom I will discuss: that

[11] Hitchcock says he intends to take the Ramsey-Price proposal and make it mathematically precise. He writes: ‘‘It would be nice to have a clearer mathematical picture of what Price 's agent probabilities are’’, and: ‘‘There is something deeply right about the idea that deliberation requires that contemplated actions be viewed as *sui generis*.’’ (Hitchcock, p. 519) This is what he is taking from Price 's work. It is worth noting that I am going to criticise Hitchcock' s attempt to defend evidential decision theory by constructing a fictional probability distribution, but there is a great deal more interesting work in that paper on the relation between decision theory and probabilistic theories of causation which I will not touch on.

agents must treat their own actions as if uncaused by external factors and probabilistically independent of them.

The formal criteria Hitchcock gives for construction of the fictional distribution $\boldsymbol{P_f}$ seem to accord with this. First, Hitchcock says, the probability of an agent choosing action $\boldsymbol{A}$ can be arbitrarily assigned (by that agent), for the fictional distribution $\boldsymbol{P_f}$. Secondly, and much more importantly, in $\boldsymbol{P_f}$ the actions should be independent of the members of the c-partition, so that $\boldsymbol{P_f}$ (A_1&factor x) always equals $\boldsymbol{P_f}$ (A_1) times $\boldsymbol{P_f}$ (factor x).[12] This just states that the probability of any action 1 and any factor x, is the same as the probability of action 1 holding multiplied by the probability of factor x holding. This states that action 1 and factor x are independent. The agent must assume in her own fictional distribution that her actions are independent of any prior factors. And this *must* be what models the freedom of action – the fiction – Hitchcock says the agent must assume. This is because apart from these two points, the fictional distribution "$\mathbf{P_f}$ should otherwise be as similar to $\mathbf{P_d}$ as possible. … suffices that $\mathbf{P_f}$ agree with $\mathbf{P_d}$ about the relative likelihood of various factors beyond the agent's control, and about how her acts, in conjunction with these factors, *influence* the occurrence of the various possible outcomes." (Hitchcock, p. 521. Emphasis added.)

Hitchcock' s story can be examined and contrasted with Price's using Coco. Hitchcock is arguing that evidential and causal decision theory give the same prescriptions for action for Coco when Coco makes the necessary assumptions about his freedom to act. Hitchcock thinks Coco should allocate $\boldsymbol{Pf}$ (eating chocolate) = some arbitrary
number between 0 and 1. But this is not the crucial claim. That is that Coco should, in $\boldsymbol{Pf}$, assume that the probability of any action is independent of prior factors. Hitchcock is saying that Coco should, to be *rational*, assume that his eating chocolate is not correlated with any pre-existing ***PMS*** -- indeed, Coco should assume his action is independent of *any* possible pre-existing state.

[12] What Hitchcock actually writes is: "1) $P_f(A_j) = p_j$, with $\Sigma_j\, p_j = 1$" and then, "P_f should satisfy: 2) $P_f(A_jB_k) = P_f(A_j)P_f(B_k)$ for all j and k." (p521) I have merely standardised notation and expressed the same thought in the terminology I have been using all along.

On Price' s view Coco *ends up* in a very similar position, but Price tells a different story. Price identifies the causal history of Coco 's contemplated action and provides an argument to support the view that in the problematic cases Coco will know that the troublesome spurious correlations don' t apply. Coco acts on *real correlations*. For Hitchcock, Coco uses a purely *fictional* distribution.

So the freedom Hitchcock thinks agents must assume is modelled by (i) arbitrary allocation of a probability to an action and (ii) assumption that an action under consideration is independent of any prior factors. It is not clear whether these two assumptions are separate parts of the assumption of freedom, or meant to be linked. For Coco, the important correlation is of course between being in ***PMS*** and eating chocolate, and perhaps the idea is that if Coco is free to allocate the probability of himself eating chocolate, then Coco must assume that his eating of chocolate is uncorrelated with other things like prior causes that would affect the absolute probability of Coco eating chocolate. But this does not follow. It could be true that Coco can allocate an arbitrary probability to his action (especially a fictional one) while this action is still correlated with prior states. So the assumption that there are no spurious correlations has to stand alone -- it has to follow from the assumption of freedom.

But if this is the picture, then Hitchcock is committed to claims that must be false. That Coco has to assume that his action is uncaused, or uncorrelated with events we normally take to be in its causal history, in order to deliberate rationally cannot be true. In fact, it sounds a very odd thing to try to do. Why should Coco, knowing that his chocolate-eating is often caused by ***PMS***, and that there is a corresponding correlation between chocolate-eating and ***PMS***, have to try to *ignore* that knowledge and *believe the opposite*, i.e. believe that there is no such correlation, in order to act rationally? This is a very strong claim that is now quite different from its Kantian antecedents, and the intuitive plausibility of ‘‘deliberation’’ about whether to go shopping while locked in a cage not counting as genuine deliberation. Indeed, information about prior causes can be highly relevant to rational choices. Take the case of family relationships involving complex emotions I mentioned earlier. Your ability to recognise that your impulse to be mean to your sister arises from foolish jealousy caused by childhood events might be crucial to your ability to ignore

impulses and forge a healthy adult relationship with your sister. Hitchcock must say more to justify such a move, particularly when the causalist has a sensible story about why rational agents should act on causes and real probabilities.[13]

As I suggested earlier, there is a second possible, and more plausible, assumption of freedom Hitchcock might have in mind. Perhaps he wasn 't thinking that an agent must take her actions to be uncaused by prior factors, just that those prior factors must cause the actions of a rational agent via her own motivations -- the favourite assumption of the tickle defenders. Hitchcock only means that an agent, in treating herself as a free agent, should not worry about prior causes of a contemplated action, and instead just take current motivations and decide on the basis of them. This is weaker, since Hitchcock could accept that Coco' s decision about chocolate involves many causal influences, some of them correlated with the desired result, and can even allow that Coco knows this. Hitchcock is only requiring that Coco assumes that these causal influences act via his own beliefs and desires. Coco doesn' t have to worry about them, because they are taken into account along with his beliefs and desires.

If this is Hitchcock 's story then it is a good deal more plausible. And it won' t fall to my objections to Price since Hitchcock doesn 't require that the agent can *tell* what is causing her action. On the new reading the story applying to Coco goes: Coco wants to eat chocolate. Since he takes himself to be free to act he cannot worry about ***PMS***

---

13 There may be a lingering thought that the system Hitchcock describes is better because it is simpler. I think this is misleading. Hitchcock is *not* saying evidentialists can get away with assuming their actions are causally and probabilistically independent of prior states, and just acting on raw correlations. This would, after all, lead to Coco avoiding chocolate. On Hitchcock' s picture Coco will still need to do a

tricky bit of figuring out. I quote again Hitchcock describing how the fictional distribution does not differ from the real one: ‘‘***Pf*** should otherwise be as similar to ***Pd*** as possible. ... suffices that ***Pf*** agree with ***Pd*** about the relative likelihood of various factors beyond the agent s control, and about how her acts, in conjunction with these factors, influence the occurrence of the various possible outcomes’’

(Hitchcock, p. 521. Emphasis added). This last part is going to be difficult, but it is the only way Hitchcock' s criterion for action could end up formally equivalent to the causal criterion. Agents need this causal information to act effectively in such cases, and it cannot be got around. Hitchcock is not giving us a simpler system, just an odd story in favour of the same criterion the causalist advocates acting on.

causing him to act. He should only consider his current motivations, so he should eat chocolate. It is more reasonable to assume that your actions are entirely determined only by your motivations and reasoning, than that they are uncaused by prior factors at all.

Unfortunately for Hitchcock, the second more plausible sort of assumption is not enough to generate his fictional distribution. Since an agent 's motivations will still be correlated with desired results, via common causes, his actions will also still be correlated with desired results. Desired results will share a common cause with some actions, in the form of causes of certain motivations of the agent. Whether or not Coco is worrying about past causes of his current motivations, Coco eating chocolate will still be correlated with future migraine, and if Coco is reasoning evidentially he should refuse chocolate. So the weaker and more plausible assumption could not generate the criterion formally identical to the causal criterion, as Hitchcock says the assumption will. If Hitchcock 's fictional distribution is to generate a criterion the same as the causal criterion, Hitchcock needs the agent to make the assumption of radical freedom. It is no accident that it is this claim that is formalised in the mathematically framed assumptions Hitchcock uses to generate the new criterion. His claim cannot be weakened; his formal description of this assumption must stand.

This suggests that this approach is never going to work. If the assumption of freedom made is strong enough so that reasoning evidentially using it gets the right prescriptions for action, then that assumption of freedom is going to be implausible. Getting the right answer on what actions to perform is not enough to establish that Hitchcock has described a reasonable way of deliberating about action. He has merely given us the causal criterion and a bizarre account of how he gets it using an unmotivated fiction that is forced to use implausible assumptions. The causalist gives the causal criterion, with a sensible account of why it is the right way to reason in those circumstances. Hitchcock 's account cannot stand up in comparison.

4 Conclusion

The general evidentialist attempt to use facts about the application of general correlations, and assumptions of freedom to evade medical Newcomb problems doesn' t work. It springs from an interesting idea that 's been intuitively plausible in outline to many. Nevertheless, I have described two different ways of using it and argued that once examined in details these attempts fail. Price provides a genuinely evidential account, but is forced to impose too-high standards for rationality, while Hitchcock is forced to make implausible claims about freedom. This means that the only plausible rationale for the criterion he gives that is formally equivalent to the causal criterion is the causal story, which would make him a causalist. Evidential decision theory must find another way to evade the medical Newcomb problems.

Acknowledgements

I am indebted to many colleagues at both Bristol and Stirling for helpful comments and suggestions on this work. In particular I would like to thank Helen Beebee, Alexander Bird, Dorothy Edgington, Hannes Leitgeb, Samir Okasha, David Papineau, and two anonymous referees for comments leading to substantial improvements to the paper.

References

Cartwright, Nancy: 1983, *How the Laws of Physics Lie*, Oxford University Press, Oxford.
Eells, E.: 1982, *Rational Decision and Causality*, Cambridge University Press, Cambridge.
Eells, E.: 1985, 'Causality, Decision and Newcomb s Paradox', in R. Campbell & L. Sowden (eds.), *Paradoxes of Rationality and Cooperation*, University of British Columbia Press, Vancouver 183--213.
Hitchcock, CR: 1996, 'Causal decision theory and decision-theoretic causation', *Nous* **30(4)**, 508-526.
Joyce, James M.: 2002, 'Levi on causal decision theory and the possibility of predicting one's own actions', *Philosophical Studies* **110**, 69-102.
Papineau, David: 2001, 'Evidentialism reconsidered', *Nous* **35(2)**, 239-259.
Price, Huw: 1991, 'Agency and probabilistic causality', *British journal for the philosophy of science* **91**, 157-176.
Price, Huw: 1986, 'Against causal decision theory', *Synthese* **67**, 195-212.