JAMES M. JOYCE

LEVI ON CAUSAL DECISION THEORY AND THE POSSIBILITY OF PREDICTING ONE'S OWN ACTIONS

(Received 21 May 2002)

ABSTRACT. Isaac Levi has long criticized causal decision theory on the grounds that it requires deliberating agents to make predictions about their own actions. A rational agent cannot, he claims, see herself as free to choose an act while simultaneously making a prediction about her likelihood of performing it. Levi is wrong on both points. First, nothing in causal decision theory *forces* agents to make predictions about their own acts. Second, Levi's arguments for the "deliberation crowds out prediction thesis" rely on a flawed model of the measurement of belief. Moreover, the ability of agents to adopt beliefs about their own acts during deliberation is *essential* to any plausible account of human agency and freedom. Though these beliefs play no part in the *rationalization* of actions, they are required to account for the *causal genesis* of behavior. To explain the causes of actions we must recognize that (a) an agent cannot see herself as entirely free in the matter of A unless she believes her decision to perform A will *cause* A, and (b) she cannot come to a deliberate decision about A unless she adopts beliefs *about* her decisions. Following Elizabeth Anscombe and David Velleman, I argue that an agent's beliefs about her own decisions are *self-fulfilling*, and that this can be used to explain away the seeming paradoxical features of act probabilities.

Isaac Levi has long been among the most persistent and influential critics of causal decision theory. At the heart of nearly all his objections lie two claims: first, that the causal theory requires deliberating agents to make predictions about their own actions; second, that this is incoherent because "deliberation crowds out prediction".[1] Levi is wrong on both points. As the first two sections of this essay will make clear, nothing in causal decision theory *forces* an agent to make predictions about her own acts. While the specific *version* of the theory *I* defend does permit this, I am, as far as I know, the *only* causal decision theorist doctrinally committed to rejecting the "deliberation crowds out prediction" thesis. The essay's third section presents my reasons for opposing the thesis. We will see that none of the standard justifications for it, including Levi's, stand

Philosophical Studies **110:** 69–102, 2002.

up to scrutiny. Moreover, I shall argue, the ability of a decision maker to adopt beliefs about her own acts during deliberation is *essential* to any plausible account of human agency and freedom. While Levi suggests that a deliberating agent cannot see herself as free with respect to acts she tries to predict, precisely the reverse is true. Though they play no part in the *rationalization* of actions, such beliefs to are essential to the agent's understanding of the *causal genesis* of her behavior.

1. WHAT IS CAUSAL DECISION THEORY?

As Levi tells it, all arguments for causal decision theory are founded on a grand false dilemma: we causal decision theorists assume that Richard Jeffrey's "evidential" decision theory is the only viable alternative to our position,[2] argue that Jeffrey's view is wrong, and conclude that our view must be right. Allegedly, this ignores a bevy of expected utility theories, like the one found in Leonard Savage's *Foundations of Statistics*,[3] which are neither causal nor evidential. To reinforce the point, Levi notes that neither the causal or evidential theory "gets much press" outside philosophy.[4] The implication is that the theories used by "real" experts on rational choice – statisticians, economists, psychologists, and so on – simply ignore the causalist/evidentialist debate.

None of this is so. Contrary to what Levi suggests, *all* legitimate decision theories are, at least implicitly, causal or evidential. The causal and evidential approaches "get no press" outside philosophy not because they are irrelevant to the serious study of rational choice, but because the nonphilosophical experts have accepted the basic message from the start, and have been incorporating it into their work for years. I will elucidate this point using Savage's theory, but it holds for other versions of expected utility theory as well.

Savage portrays the rational decision-maker, hereafter *DM*, as appealing to beliefs about possible *states of the world* to choose *acts* that are likely to produce desirable *outcomes* (his "consequences"). States are the locus of all uncertainty in the model; *DM*'s opinions are captured in a *subjective probability function* **P** that is defined over states.[5] Outcomes, the objects of *DM*'s non-instrumental desires, are assigned utilities. Acts have only instrumental value, and are evaluated in terms of their *unconditional expected utilities*. These are computed using the formula:

$$\textbf{SAVAGE:}\quad \mathbf{Exp}(A) = \Sigma_S\, \mathbf{P}(S) \cdot \mathbf{u}(\mathbf{o}[A, S])$$

where *S* ranges over states of the world and **o**[*A*, *S*] is the outcome that *A* will produce if *S* obtains. *DM* can rationally choose to perform *A*, according to Savage, only if it maximizes her unconditional expected utility.

As the notation suggests, Savage assumed that state probabilities do *not* vary across acts; **P**(*S*) remains the same no matter which act's expected utility is being computed. This restriction is essential because employing SAVAGE when state probabilities vary with acts can lead to trouble. Consider the following decision:

	You will contract influenza this winter	**You will not contract influenza this winter**
Get a flu shot	Get the flu, and suffer the minor pain of a shot	Avoid the flu, but suffer the minor pain of a shot
Do not get a flu shot	Get the flu, but avoid the minor pain of a shot	Avoid the flu, and avoid the minor pain of a shot

When you apply SAVAGE using the same state probabilities for both acts it tells you to avoid the pain by forgoing the shot (because this is the *dominant* act), which is extremely bad advice given that your chances of getting the flu are markedly less with the shot that without it. Does this mean that SAVAGE should be rejected? Definitely not! If you present this problem to one of the experts from whom causal decision theory "gets no press" they will tell you that you have *misapplied* Savage's theory. You need to reformulate your decision problem using states whose probabilities do *not* vary with your choice of an act, like these:

- You will contract the flu whether or not you get the shot.
- You will not contract the flu whether or not you get the shot.
- You will contract the flu if you get the shot, but not otherwise.
- You will not contract the flu if you get the shot, but will otherwise.

Doing the numbers this way yields the right answer.

The moral is that "Savage's theory" is more than a mathematical formalism. It also involves an unwritten rule about the kinds of decision problems to which the formalism may legitimately be applied. Before we can use SAVAGE to identify *DM*'s optimal choices we need to describe her decision in a way that makes her probabilities for states independent of her choice of acts. There is nothing special about Savage's view here. Any utility theory that weights utilities of outcomes by *unconditional* state probabilities comes with a tacit warning: *do not apply unless probabilities of states are independent of acts*. The experts are content to leave this caveat implicit, and to rely on their good judgment to select decision problems for which their theories yield sensible answers. This is fine for those whose main interest lies in *applying* decision theory to solve practical problems, but we cannot obtain a complete *understanding* of what a decision theory says until we make its tacit principles explicit in the formalism. This is exactly what we causal decision theorists aim to do. Our efforts get little press outside philosophy not because they are irrelevant to what the experts are doing, but because the experts have *always* implicitly incorporated the basic message of into their practices of decision problem selection.[6] Indeed, Judea Pearl, who counts as an expert if anyone does, eschews the phrase "causal decision theory" in order "to suppress even the slightest hint that any alternative, noncausal theory can be used to guide decisions".[7]

In my view, the best way to make this message explicit is by generalizing SAVGE so that it allows for the calculation of expected utilities even when state probabilities depend on acts. To do this, each act A is evaluated *on the supposition that is it preformed*, and outcomes are weighted not by unconditional probabilities of states, but their probabilities *given A*. Savage's equation is thus replaced by

General Equation (GE): $\mathbf{Exp}(A) = \Sigma_S\, \mathbf{P}(S||A) \cdot \mathbf{u}(\mathbf{o}[A, S])$

where, for a each act A, $\mathbf{P}(\bullet||A)$ is a probability that represents *DM*'s degrees of confidence in various states of the world on the supposition that A is performed. Since GE reduces to SAVAGE whenever states are independent of acts it follows that SAVAGE applies exactly when $\mathbf{P}(S||A) = \mathbf{P}(S)$ for all acts A and states S.

The difference between the causal and evidential theories has to do with the interpretation of $\mathbf{P}(S||A)$. Evidentialists identify it with the subjective probability that *DM* should assign *S* upon *learning* *A*, i.e., with *DM*'s subjective probability for *S* conditioned on *A*, $\mathbf{P}(S/A) = \mathbf{P}(S \,\&\, A)/\mathbf{P}(A)$. Expected utilities are then computed using

$$\textbf{EDT:}\quad \mathbf{Exp}(A) = \Sigma_S\, \mathbf{P}(S/A) \cdot \mathbf{u}(\mathbf{o}[A, S])$$

Since this agrees with SAVAGE when states and acts are evidentially independent, evidentialists apply SAVAGE when $\mathbf{P}(S/A) = \mathbf{P}(S)$ for all *A* and *S*.

This model usually works quite well, but only because *DM*'s probabilities conditional on *A* so often capture her views about what *A* will *cause*. If *DM* thinks that doing *A* will causally promote *S* then, ordinarily, $\mathbf{P}(S/A)$ exceeds $\mathbf{P}(S/\neg A)$. There are cases, however, in which *DM*'s beliefs about what her acts might cause are not adequately reflected by her conditional probabilities, so that $\mathbf{P}(S/A)$ exceeds $\mathbf{P}(S/\neg A)$ even though she takes *S*'s truth to be *causally independent* of *A*'s . In such "Newcomb-type" problems acts can serve as *reliable indicators* of states without *causally promoting* them. EDT yields incorrect answers in such cases.

For a realistic Newcomb-type problem we can do no better than the *Twin's Dilemma*, a Prisoner's Dilemma with a twist. Two players, Row and Column, must decide whether or not to take some cooperative action. They make their choices simultaneously in separate locations so that there is no chance of either causally influencing the other. Their utilities are given by

Utilities (Row, Column)	***C* = Column Cooperates**	**¬*C* = Column Defects**
***R* = Row Cooperates**	(9, 9)	(0, 10)
¬*R* = Row Defects	(10, 0)	(1, 1)

The twist is that Row believes that she and Column are mildly like-minded, and thus she sees her act as evidence for what he will do. We can use the quantity $\mathbf{P}(C/R) - \mathbf{P}(C/\neg R)$ as a measure of the extent to which Row takes her cooperation to indicate Column's cooperation. If this difference is large enough, Row may be tempted by the following thought: "Since Column is more likely to cooperate if I cooperate than if I defect, and since I'm better off if Column cooperates no matter what I do, I should cooperate”. SAVAGE and EDT differ as to the legitimacy of this reasoning. SAVAGE always recommends choosing the *dominant* act $\neg R$, whereas EDT endorses R as long as $\mathbf{P}(C/R) - \mathbf{P}(C/\neg R) > 1/9$.

Causal decision theorists agree with SAVAGE. Since defecting puts Row a utile to the good *no matter what Column does* it follows that cooperating can *only* further Row’s ends by *influencing* Column’s act. Since Row does not believe she can do this, and since there is no cost in defecting, Row should defect to gain the extra utile. EDT goes wrong by weighting utilities of outcomes by the conditional probabilities of states given acts. Since the values of $P(C/R)$ and $P(C/\neg R)$ do *not* fully encode Row’s views about what her acts might *cause*, she ends up choosing means ineffective to her ends. To get the right result we must generalize SAVAGE as

$$\textbf{CDT:}\quad \mathbf{Exp}(A) = \Sigma_S\, \mathbf{P}(S\backslash A) \cdot \mathbf{u}(\mathbf{o}[A, S])$$

where $\mathbf{P}(\bullet\backslash A)$ is a probability function that captures *DM*’s views about what A is likely to cause. Different causal theorists interpret the “causal probability” $\mathbf{P}(\bullet\backslash A)$ differently,[8] but all agree that (a) it is not $\mathbf{P}(\bullet/A)$, and (b) it represents *DM*’s beliefs about what her acts will *causally promote*, so that $\mathbf{P}(S\backslash A)$ will exceed $\mathbf{P}(S\backslash\neg A)$ only if *DM* believes that A will causally promote S. Since CDT reduces to SAVAGE when states and acts are causally independent, another way to express the causal view is by saying that SAVAGE applies when $\mathbf{P}(S\backslash A) = \mathbf{P}(S)$ for all A and S.

When we look at things this way we see that both causal decision theory and evidential decision theory are *extensions* of Savage’s formalism. Far from *ignoring* Savage’s approach, both seek, in different ways, to *complete* it by allowing for the calculation of expected utilities even when state probabilities vary with acts. Since, for these purposes, CDT and EDT *are* the only live options, the case for causal decision theory does not rest on any false dilemma.

Levi only thinks otherwise because he mistakenly believes that the causal and evidential models are distinguished from other utility theories by the fact that they, and they alone, *force* agents to assign probabilities to their own actions. He asserts that CDT and EDT *only* come into conflict in Newcomb problems when deliberating agents try to predict their own behavior by assigning unconditional probabilities to their own acts. Levi sees this as *the* crucial divide in decision theory, and he regards any theory that does not traffic in act probabilities as "neither causal nor evidential". Having framed the issue in this way, he goes on to argue against the coherence of act probabilities in the hope of refuting the causal and evidential theories in a single swoop. We will consider his arguments in Section 3, but let's first understand why they would not refute causal decision theory even if sound. As we will see in the next section, one need *not* invoke act probabilities to distinguish EDT and CDT. Some *theorists* (e.g., Jeffrey and me) endorse act probabilities, but this is an *option*, not part of the "standard equipment" of either approach. Levi misunderstands this because he misconstrues the nature of the debate over Newcomb problems.

2. LEVI ON NEWCOMB PROBLEMS

According to Levi, the Twin's Dilemma is a "weak reed on which to rest a case for causal decision theory".[9] It is far too underspecified to distinguish CDT from EDT since there are versions of the problem in which EDT-maximizers defect. Indeed, Levi thinks that the Twin's Dilemma *only* yields a conflict between EDT and CDT when agents assign unconditional probabilities to their own acts, and "the unconditional probability of either prisoner confessing is approximately 0.5".[10] This is wrong. Levi is only able to arrive at this conclusion by expanding the class of Twin's Dilemmas to include problems that should not be there, and ignoring others that should. Let's consider cases.

Levi offers two examples of alleged Newcomb problems in which EDT recommends defecting. In the first,

> each prisoner judges that that it is highly probable that she and her twin will both confess or both not confess. This is part of what it is to judge that one is very much like one's twin. Causal decision theorists read this as erroneously implying that from [Row's] point of view, the conditional probability of [Column] confessing (not confessing) given that [Row] confesses (does not confess) is high. That is to say, given that $\mathbf{P}(R \ \& \ C) + \mathbf{P}(\neg R \ \& \ \neg C)$ is very close to 1 it is concluded that $\mathbf{P}(C/R)$ and $\mathbf{P}(\neg C/\neg R)$ are both near 1.[11]

Levi rightly notes that $\mathbf{P}(R \ \& \ C) + \mathbf{P}(\neg R \ \& \ \neg C)$ can be close to 1 when $\mathbf{P}(\neg C/\neg R)$ is near 0, and that in this event EDT recommends confessing. But, he errors when he goes on to conclude that "the Prisoner's Dilemma for like-minded twins does not specify whether the conditions relevant to discriminating between cases where [EDT] favors confessing and [EDT] favors not confessing are in force".[12] Levi is reading "like-minded" in a way that no causal decision theorist has or ever would. The *sine qua non* of Newcombhood in a Twin's Dilemma is that Row regards Column's actions as causally independent of her own, *but sees R as providing significantly better evidence than $\neg R$ does for C.* Given the utilities we are using, "significantly better" means something fairly weak, but quite precise: $\mathbf{P}(C/R)$ must exceed $\mathbf{P}(C/\neg R)$ by at least 1/9. It does *not* mean $\mathbf{P}(C \;\&\; R) + \mathbf{P}(\neg C \;\&\; \neg R) \approx 1$ as Levi suggests. Since Levi offers us with a case in which $\mathbf{P}(C/R)$ and $\mathbf{P}(C/\neg R)$ are *both* close to 1, the crucial inequality $\mathbf{P}(C/R) - \mathbf{P}(C/\neg R) > 1/9$ does *not* hold. So, Levi's first example is a red herring; it is *no Newcomb problem at all.*

Levi's second example is similarly flawed. Here he imagines that "the probabilities relevant for computing expectations are indeterminate", and states that, "everyone agrees that the dominating option is to be recommended" in such a situation.[13] I wholeheartedly agree that in any realistic case Row's credal state will be represented not by a single probability function, but by a set of them. Levi thinks this set must be convex, I don't, but no matter. What matters is that *every* function in the set be such that $\mathbf{P}(C/R) - \mathbf{P}(C/\neg R) > 1/9$. If this is not so, then we are *not* dealing with a Newcomb problem since Row does not unambiguously see R as providing significantly better evidence than $\neg R$ does for C. On the other hand, if $\mathbf{P}(C/R) - \mathbf{P}(C/\neg R) > 1/9$ for *every* $\mathbf{P}$ then an EDT-maximizer will cooperate even though $\mathbf{P}(C/R)$ and $\mathbf{P}(C/\neg R)$ are indeterminate because it *will* be determinate that EDT-utility of R exceeds that of $\neg R$. Here, as in the previous case, EDT and CDT *do* conflict as long as we are dealing with a *genuine* Newcomb problem.

Next let's assess Levi's claim that CDT and EDT only conflict when "the probability of each prisoner confessing is 0.5". I am not sure how Levi arrives at this result, but he would seem to need two auxiliary assumptions:

Symmetry: Row will recognize that Column's situation is identical to her own, and will therefore set $\mathbf{P}(C/R) = \mathbf{P}(R/C)$ and $\mathbf{P}(C/\neg R) = \mathbf{P}(R/\neg C)$.

Uncertainty: $\mathbf{P}(C) = \mathbf{P}(\neg C) = 1/2$.

We causal decision theorists are often guilty of presenting Newcomb problems in ways that make these assumptions seem compulsory, but both are optional. There are plenty of Twin's Dilemmas in which $P(R)$ and $P(C)$ are unequal and far from 1/2, and plenty in which Symmetry fails. There are even some in which agents do *not* regard their acts as strongly correlated. One example makes all these points. Consider any probability function of the following form, where $0 < x < 1$:

	$P(C) = 2/9 + 2/9 \cdot x$	$P(\neg C) = 7/9 - 2/9 \cdot x$
$P(R) = x$	$4/9 \cdot x$	$5/9 \cdot x$
$P(\neg R) = 1 - x$	$2/9 \cdot (1 - x)$	$7/9 \cdot (1 - x)$

Each of these is a Twin's Dilemma in which $P(C/R) = 4/9$ and $P(C/\neg R) = 2/9$. In every case except $x = 2/7$ Symmetry fails and $P(R)$ and $P(C)$ differ. Finally, there is no positive correlation between R and C since Row takes $\neg C$ to be *more* likely than C even when she cooperates. EDT recommends choosing R not because it makes Column's cooperation likely, but because it makes it slightly less unlikely. What we have, then, is a family of Twin's Dilemmas in which EDT and CDT conflict even though Symmetry and Uncertainty fail.

Given that none of Levi's examples succeeds, I remain confident that there is an unequivocal distinction between evidential and causal approaches to decision making in *genuine* Newcomb problems. To reiterate: in any Twin's Dilemma (with the given utilities), if Row judges Column's acts to be causally independent of her own, and if her credal state contains only probabilities such that $\mathbf{P}(C/R) - \mathbf{P}(C/\neg R) > 1/9$, then CDT recommends defection while EDT recommends cooperation. Any ambiguity Levi finds in this is something he is adding himself.

Still, Levi does not rest his whole case on this point about ambiguity. His more serious objection is that in all Newcomb problems the agent "is committed to assigning unconditional probabilities to the available options".[14] Since we causal decision theorists rely on these problems to distinguish CDT-maximization from EDT-maximization, Levi thinks that we are *forced* to sanction unconditional probabilities for acts. While *I* do sanction them, and will soon explain why, I do not want to see my brothers and sisters in causation tarred with Levi's overly broad brush. Contrary to what he claims, agents in Newcomb problems need *not* set determinate probabilities for their own acts. Let Row's credal state be the entire (convex) set of probabilities in the above table. Since $\mathbf{P}(C/R) - \mathbf{P}(C/\neg R) = 2/9$ holds for every **P** in her credal state Row faces a true Newcomb Problem. Yet, her subjective probability for R is *maximally* indeterminate; she has no views whatever about what she is likely to do. Thus, Newcomb problems *can* arise even for agents who assign no unconditional probabilities to their own acts. This is not surprising. Nothing in the formalism of either causal or evidential decision theory *requires* agents to assign unconditional probabilities to their own acts since in both theories *the evaluation of expected utilities proceeds without reference to the probabilities of the options being considered.* It is thus consistent with either view to institute a blanket prohibition against act probabilities as long as the requisite probabilities for states *conditional on acts* are determinate.

I do not favor instituting such a prohibition because I think causal decision theory, and decision theory generally, is best formulated in terms of a system of axiomatic constraints on preferences that were used by Jeffrey and Ethan Bolker to codify the evidential theory.[15] It is a feature of the Jeffrey-Bolker formalism that acts, states and outcomes can all be represented by propositions, and that any proposition can, in principle, be assigned a utility and a subjective probability. So, in *my particular formulation* of causal decision theory it is permissible to assign probabilities to acts. Even if Levi were able to prove that act probabilities are incoherent, the most he would have shown is that my particular way of formulating things is flawed. He only thinks he can show more because he believes that CDT and EDT can *only* diverge when agents assign unconditional probabilities to their own acts. As we have just seen, this is a mistake. Let us now see why it is also a mistake to think that probabilities cannot be coherently assigned to acts.

3. CAN RATIONAL DELIBERATORS PREDICT THEIR OWN ACTIONS?

Levi has long held that there is something deeply problematic about agents treating their own acts as objects of belief. He writes:

> Deliberation crowds out prediction, so that a decision-maker may not coherently assign unconditional probabilities to the propositions he regards as optional for him ... Although [he] may predict his future choices as well as the choices of others, [he] cannot coherently assign unconditional probabilities to his currently available options.[16]

Moreover, any decision theory that permits act probabilities will be “insensitive to the distinction between what is under the decision maker’s control and what is not”.[17] Precisely the reverse is true. By conflating issues about what a person can control with questions about probabilities of acts Levi is lead to embrace a wholly untenable view of human agency. His arguments against assigning probabilities to acts are flawed, and there are independent reasons to favor a model of the that allows agents to adopt opinions about what they will do.

To put Levi’s criticisms into perspective, let’s consider some general worries that one might have about letting agents assign probabilities to their own acts:

> *Worry-1*: Allowing act probabilities might make it permissible for agents to use the fact that they are likely (or unlikely) to perform an act as a *reason* for performing it.

> *Worry-2*: Allowing act probabilities might destroy the distinction between acts and states that is central to most decision theories.

> *Worry-3*: Allowing act probabilities “multiplies entities needlessly” by introducing quantities that play no role in decision making.

We will consider these concerns in order. Levi’s misgivings are best seen as a species of the second worry, so I will discuss his views at that point.

As to Worry-1, I entirely agree that it is absurd for an agent’s views about the advisability of performing any act to depend on how likely she takes that act to be. Reasoning of the form “I am likely (unlikely) to A, so I should A” is always fallacious. While one might be tempted to forestall it by banishing act probabilities altogether, this is unnecessary. We run no risk of sanctioning fallacious reasoning as long *A*'s probability does not figure into the calculation of its own expected utility, or that or any other act. No decision theory based on the General Equation will allow this. While GE requires that each act *A* be associated with a probability P(•||*A*), the values of this function do *not* depend on *A*'s unconditional probability (or those of other acts). Since act probabilities "wash out" in the calculation of expected utilities in both CDT and EDT, neither allows agents to use their beliefs about what they are likely to do as reasons for action.

The second worry has been clearly articulated by Itzhak Gilboa, whose views about act probabilities are similar to Levi's. Gilboa writes that any theory that "allows a decision maker to have beliefs about his or her own choices ... robs decision theory of one of its most cherished assets, namely, the theoretical dichotomy between states of the world (which cannot be controlled) and choices (regarding which there are no beliefs)".[18] I wholeheartedly agree that *if* allowing act probabilities robs decision theory of the act/state distinction, then they should be banished. But there is no reason to think this is so. Gilboa has really drawn *two* distinctions: one between what *DM* can and cannot control, and one between what is and is not a legitimate object of belief for her. Most decision theorists follow Savage in running these distinctions together. This makes sense when one is *only* concerned with assigning utilities to acts since, as we have just seen, act probabilities do not enter into such assignments. Still, it does not follow that the distinction between what *DM* can and cannot control is *the same as* the distinction between what can and cannot figure in her subjective probabilities. Even if act probabilities do not figure into the calculation of act utilities, they may have other roles to play in the process of rational decision making. Indeed, we shall soon see that they do. And, if this is so, then the common practice of using the act/state dichotomy to do double duty for the can/cannot-control distinction and the can/cannot-have-a-probability distinction looks to be a mere artifact of decision theory's focus on act utilities. Without further arguments Gilboa's worries should not bother us.

Levi does have further arguments to offer here.[19] He claims that any agent who assigns probabilities to her acts must cease to see herself as free to choose these acts. The only options she will regard as *available* for choice will be those that are *admissible* "in the sense that they are not ruled out [as irrational] by principles of choice given [her] beliefs and values".[20] Levi's argument for this rests on two premises:

> Premise-1: An agent who assigns probabilities to her present actions is required, on pain of irrationality, to assign a probability of zero to any inadmissible act.
>
> Premise-2: Once a deliberating agent assigns a subjective probability of zero to an action she no longer regards it as available for choice.
>
> ---
>
> Conclusion: An agent who assigns unconditional probabilities to her own acts cannot regard any inadmissible act as available for choice.

If sound, this would indeed sound the death-knell for act probabilities. For as Levi notes, it would render the decision-making process *vacuous* since an agent would not even see herself as *free* to choose irrationally. However, neither of the argument's premises is true, and there are independent reasons for doubting its conclusion.

3.1. *Why Premise-1 is false*

Levi's justification for Premise-1 depends on the assumption that subjective probabilities for acts are always revealed in betting behavior. As Frank Ramsey and Bruno de Finetti first noted,[21] *under ideal conditions* one can discover *DM*'s degree of confidence in a hypothesis H by seeing how she bets on its truth-value. Let *DM* be an expected utility maximizer for whom money is linear in utility. Suppose also that her utilities for H and $\neg H$ will not change if she accepts a wager $W = [\$x \text{ if } H; \$y \text{ if } \neg H]$ that pays her $\$x$ if H is true and $\$y \neq x$ if H is false. We can then ascertain *DM*'s probability for H by eliciting her *fair price* for W, that sum of money $\$\mathrm{p}_W$ at which she indifferent between having $\$\mathrm{p}_W$ or receiving W's schedule of payments. Since $\mathrm{p}_W = \mathrm{P}(H)\cdot x + \mathrm{P}(\neg H)\cdot y$ maximizes expected utility, her *betting quotient* for H, $\mathrm{b}_H = (\mathrm{p}_W - y)/(x - y)$, will be both (a) independent of the choice of x and y and (b) equal to $\mathrm{P}(H)$. Given (a), we can set $x = 1$ and $y = 0$, so that *DM*'s price for $W = [\$1 \text{ if } H; \$0 \text{ if } \neg H]$ reveals her probability for H directly.

There are a variety of ways to elicit fair prices. Levi likes to speak in terms of the conditions under which *DM* would accept or reject bets. For our purposes, it is better to exploit a trick, due to de Finetti, that makes it clear what *decision DM* faces when she fixes a fair price for *W*. In *Theory of Probability*,[22] de Finetti showed that stating a fair price for *W* is equivalent to making a straight choice among (advantageous) wagers of the following form, with $1 \geq \mathrm{p} \geq 0$:

$$W(p) = [\$(1 - (1 - p)^2) \text{ if } H; \$(1 - p^2) \text{ if } \neg H]$$

That is, $\$\mathrm{p}_W$ is *DM*'s fair price for *W* iff she prefers $W(\mathrm{p}_W)$ to $W(p)$ for all $p \neq \mathrm{p}_W$. *DM* has an incentive to fix p_W as close to *H*'s truth-value as she can since she loses $\$(1 - \mathrm{p}_W)^2$ when *H*'s truth-value is 1 and $\$\mathrm{p}_W{}^2$ when its truth-value is 0.

Given this identity of fair prices and degrees of belief, it is natural to think that *DM* only *has* a subjective probability for *H* if she assigns a definite fair price to *W* (or, equivalently, if she has a definite betting quotient for *H*). Applied to hypotheses that describe *DM*'s actions, the assumption comes to this:

Act Probabilities are Revealed in Fair Prices. *DM* has a definite subjective probability **P**(*A*) for an act *A* if and only if **P**(*A*) is her fair price for the wager W_A = [\$1 if *A*; \$0 if ¬*A*] or, equivalently, she prefers $W_A(\mathbf{P}(A))$ among all wagers of the form

$$W_A(p) = [\$(1 - (1 - p)^2) \text{ if } A; \$(1 - p^2) \text{ if } \neg A]$$

If this is right, then the measurement of act probabilities comes down to the measurement of fair prices for wagers like W_A.

Levi uses the thesis that act probabilities are revealed by fair prices to deduce that *DM* can only assign probabilities to her acts if she is certain she will choose optimally. He reasons as follows:[23] Suppose that (i) *DM* sees *A* and ¬*A* as her only options, (ii) she takes them to be fully under her control, and (iii) she strictly prefers *A* to ¬*A*. Now, imagine that we try to ascertain *DM*'s subjective probability for *A at a time before DM chooses* by having her choose among all wagers of the form $W_A(p)$. This *alters* her decision. Her options are no longer just *A* or ¬*A*; now they include all prospects of the form ±*A* & $W_A(p)$ where ±*A* may be *A* or ¬*A* and *p* is any real number such that $1 \geq p \geq 0$. Given this, it would clearly be irrational for *DM* to perform *A* and set W_A's price at anything less than \$1, or to perform $\neg A$ and set W_A's price at anything greater than \$0. Doing anything else would be to choose a strictly dominated option. So, once dominated options are eliminated, the issue boils down to a choice between A & $W_A(1)$ and $\neg A$ & $W_A(0)$. Since the payoff from the wager is \$1 in each case, *DM* has no reason to refrain from satisfying her preference for A. Thus, her only rational choice is A & $W_A(1)$, and the only fair price she can rationally assign to W_A is \$1. Since this price reveals her subjective probability for A it follows that $\mathbf{P}(A) = 1$.

This reasoning is fallacious. Betting quotients may not be used to measure probabilities of propositions whose truth-values the believer can control because *the measuring process alters the quantity measured*. When we try to ascertain *DM*'s betting quotient for A by eliciting a fair price for W_A we give her an *incentive* her to make up her mind about A *before* setting a price. This, as I will argue, is an incentive she would be irrational not to take. Accordingly, her betting quotient for A reveals the probability that she assigns to A *after* she has decided whether or not to perform it. This entirely undercuts the force of Levi's argument. He cannot prove anything about *DM*'s doxastic state *during* her deliberations by showing that she assigns extreme probabilities to her actions after her deliberations have ceased. Friends of act probabilities can gladly grant that, once deliberation ends, *DM* will be certain about both what she has decided and what act she will do as a result of her decision. But, since *DM*'s probability for A can (and usually will) *change* as a result of her deliberations, it is no news to be told that *DM*'s subjective probability for A must be 1 *after* she decides on A. The controversial claim, and the one Levi explicitly means to defend, is that friends of act probabilities are committed to thinking that *DM*'s probability for A must be 1 *during her deliberations*.

He is wrong about this. Since *DM* has the ability to do A or $\neg A$, the problem she faces is not merely that of *figuring out* W_A's worth, as it usually is when fair prices are being elicited, but of deciding what to *make* it worth. Given that *DM* has an incentive to set a price as close to A's truth-value as she can, it would be foolish of her to put a price on W_A until *after* she makes a decision about A. To see why, notice that in choosing among options of the form $\pm A$ & $W_A(p)$, *DM* must pursue one of the following deliberative strategies

Strategy-1: Choose first between A and $\neg A$ and then among the $W_A(p)$.

Strategy-2: Choose first among the $W_A(p)$ and then between A and $\neg A$.

Strategy-3: Choose simultaneously between A and $\neg A$ and among the $W_A(p)$.

Strategy-1 is clearly rational since it allows *DM* to secure her most preferred result A & $W_A(1)$. Moreover, even *before* she starts deliberating she will know that it will let her have $W_A(1)$ if she settles on A or $W_A(0)$ if she settles on $\neg A$. Either way she gets her preferred option plus a dollar. Thus, not only does Strategy-1 offer a risk-free guarantee of the best possible outcome, *DM* will know from the start that she cannot do better with any other strategy. It is also obvious that the choice of a fair price for W_A in Strategy-1 reveals *DM*'s *post*-deliberation probability for A. Thus, *DM* can rationally pursue Strategy-1, and doing so will always yield a fair price for W_A that reveals her level of confidence in A at the *end* of her deliberations.

Given this, it follows that *DM* can only rationally pursue Strategies 2 or 3 if she knows in advance that she can do as well with them as with Strategy-1. That is, she must be convinced that she can use them, risk-free, to secure both her preferred option between A and $\neg A$ and an extra dollar. Since she can only secure \$1 by choosing $W_A(1)$ or $W_A(0)$, the issue becomes whether it can be rational for her to set a price of \$1 or \$0 for W_A *before* she has decided between A and $\neg A$ (as in Strategy-2) or at the instant she decides (as in Strategy-3)?

Strategy-2 is clearly irrational if *DM* is at all uncertain about the outcome of her deliberations concerning A. Forgoing Strategy-1 and choosing $W_A(1)$ or $W_A(0)$ in such a case is like passing up a free chance to watch the end of the horserace before placing an all-or-nothing bet on its outcome. The only exception occurs when *DM* is *certain* about which action she will choose *before concluding her deliberations*. For in this case Strategy-2 will also seem like a riskless "sure thing" to her, and it can be rational for her to pursue it in lieu of Strategy-1. Thus, a necessary condition for the rationality of Strategy-2 is that *DM* must be certain about what she is going to decide before concluding her deliberations.[24]

Opponents of act probabilities might try to portray this as another way of making Levi's point. Since Strategy-2 can only be rationally pursued when *DM* is already certain of A, they will argue, *DM*'s probability for A before and during deliberation must be 1 for A to have any probability at all. This is inference is fallacious. All that has really been shown is that *DM*'s pre-decision probability for A must be 1 if she has any probability for A *and if she forgoes Strategy-1 to pursue Strategy-2*. It is no requirement of rationality, however, that she forgo Strategy-1. In fact, if *DM* assigns A a probability strictly between 0 and 1 and Strategy-1 is available, then she will *never* pursue Strategy-2 because Strategy-1 will offer her a higher expected utility. On the other hand, if Strategy-1 is *not* an option, say because we figure out some way to force *DM* to choose among the $W_A(p)$ before she begins deliberating about A, then she will *not* see the choice of $W_A(1)$ or $W_A(0)$ in Strategy-2 as *guaranteeing* an optimal outcome if she assigns A an intermediate probability. As an expected utility maximizer, she will choose $W_A(\mathbf{P}(A))$. Thus, if *DM* assigns A an intermediate probability she will either forgo Strategy-2 for Strategy-1 or, if Strategy-1 is not an option, she will *not* choose $W_A(1)$ or $W_A(0)$ but will fix on some intermediate price.

Of course, if *DM is* antecedently certain about the ultimate outcome of her deliberations, then she may be able rationally pursue Strategy-2 by choosing $W_A(1)$ or $W_A(0)$ (supposing that her attitude of certainty can be warranted). Still, this is no help to Levi, who is hoping to show that *DM must* choose $W_A(1)$ or $W_A(0)$ on pain of irrationality, and thereby to convince us that her subjective probability for A during deliberation *must* be either 1 or 0. The most we can say here is that *if DM*'s probability *A happens* to be 1 or 0 before or during deliberations, then she must choose $W_A(1)$ or $W_A(0)$. As yet, we have been given no reason whatever to think that the antecedent of this conditional *must* be satisfied.

This brings us to Strategy-3, in which *DM* simultaneously chooses a fair price for W_A and a truth-value for A. This seems to be the case Levi has in mind,[25] and we can understand where his argument goes awry by seeing why Strategy-3 will lead *DM* to choose a fair price that reveals her *post-decision* probability for A *even if she assigned it an intermediate probability during her deliberations*. Levi's reasoning proceeds in three stages.

1. *DM*'s decision among prospects of form $\pm A$ & $W_A(p)$ is reduced to a straight choice between A & $W_A(1)$ and $\neg A$ & $W_A(0)$.

2. On the basis of (1) and the premise that *DM prefers A* to ¬*A* it is concluded that *DM* will in fact *choose A* & $W_A(1)$ over ¬*A* & $W_A(0)$.

3. Given that *DM* will choose *A* & $W_A(1)$ it follows that **P**(*A*) = 1 during the course of her deliberations.

Step-1 goes through without a hitch; it merely reflects the fact that *DM* may only pursue Strategy-3 if it guarantees her at least what Strategy-1 does. There is no problem with (2) either; *DM* will indeed end up choosing *A* & $W_A(1)$. Step (3) is the dubious one. Even if *DM* sees *A* & $W_A(1)$ and ¬*A* & $W_A(0)$ as her only options, she will not settle on the former until she *recognizes* that *A* is preferable to ¬*A*, and this will happen only *after* her deliberations cease. Acts are non-basic prospects whose value depends on both an agent's beliefs and basic desires. Even someone who has full knowledge of her beliefs and basic desires will not know how to act without doing some thinking. Deliberating is a process by which *DM* uses data about her desires and beliefs, augmented by principles of rational choice, to figure out which acts will best serve her interests.[26] However this process transpires, *DM* will *not* generally know which act she prefers *during* her deliberations; the purpose of deliberating is to figure this out. Since *A* & $W_A(1)$ and ¬*A* & $W_A(0)$ each pay $1, it can only be rational for *DM* to choose the former if she prefers *A* to ¬*A and knows this*. Merely being committed to having the preference by her beliefs and basic desires is insufficient; until *DM realizes* that *A* is the better option she will lack any sound rationale for choosing *A* & $W_A(1)$. Since she will not come to this realization until her deliberations cease, it follows that her choice of *A* & $W_A(1)$ will reveal her level of confidence in *A after* she has decided to do it. Accordingly, the only valid inference that can be drawn from the conclusion of (2) and the identity of fair prices with subjective probabilities is that **P**(*A*) = 1 *after DM* completes her deliberations. Again, this is not the conclusion Levi seeks.

In sum, Levi's arguments do not challenge the rationality of intermediate act – probability assignments *during the period when the agent is deliberating*. It is quite true that the probabilities she assigns to acts during her deliberations cannot be elicited using wagers in the usual way, but this does not show that they incoherent, only that they are difficult to measure. Levi's argument for Premise-1 thus collapses.

3.2. *Why Premise-2 is false*

Let's play along though, and imagine that *DM* assigns *zero* probability to inadmissible acts. Does it follow, as Premise-2 has it, that *DM* cannot see herself as free to act irrationally? Levi thinks so. "If [*DM*] is convinced that [she] will choose rationally," he writes, "then [she] is convinced that every proposition describing a suboptimal course of action will be false. Suboptimal options will have been ruled out as possibilities, and hence, as available options for choice".[27] This calls for some interpretation. First, "possible" in the second sentence does not mean metaphysical or logical possibility. It denotes a kind of *epistemic* possibility that Levi calls *serious possibility*. Roughly, a possibility is serious for a person just in case she would be irrational if she failed to consider it in her deliberations about how to act.[28] Second, Levi's claim only makes sense if the antecedent of the first sentence is read as asserting the *de re* claim that *DM* is certain *of* each inadmissible act that she will not perform *it*. On a weaker, *de dicto* reading *DM* would merely be convinced that *whatever* act she ends up choosing will be rational. This is consistent with her *not* knowing, early in her deliberations, that certain irrational acts are irrational, and these acts would be epistemic possibilities for her at that time. Given these provisos, the issue boils down to this: does the fact that *DM* is certain that she will not perform a given act prohibit her from seeing that act as available for choice?

In a number of places[29] Levi tries to explain what it is for *DM* to see *A* as available for choice, and to regard it as under her control, during her deliberations. He states the analysis in various ways, but he is clearly committed to at least the following theses (whose names I have chosen):

Levi's Analysis of Availability. DM sees *A* as available for choice during her deliberations only if

Deliberation: DM is certain she is deliberating.

Ability: *DM* sees herself as having the *ability* to decide to do *A* on the basis of her deliberations. Minimally, this requires her to regard the proposition d*A* = "My deliberations will terminate in a decision to *A*" as a serious possibility.

> *Efficacy*: *DM* is convinced that her decision regarding *A* will be *efficacious* in the sense that she does not regards it as seriously possible that she will decide on *A* (or $\neg A$) but will not actually perform *A* (or $\neg A$).

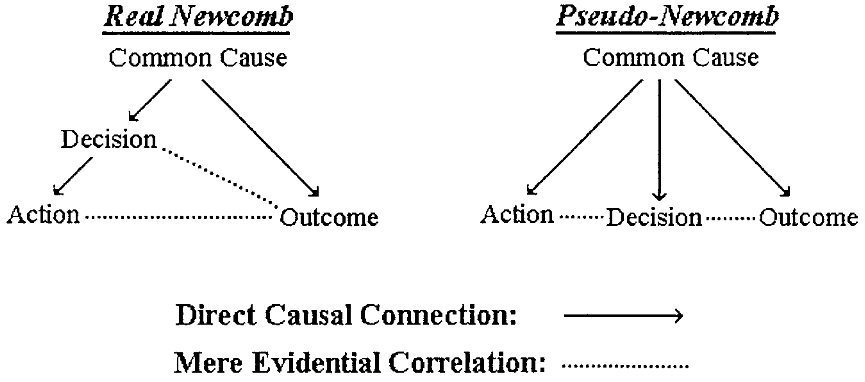
Even though the two key terms in this analysis – "ability" and "efficacious" – are causal in nature, Levi insists on explicating them in evidential terms. The lynchpin is the notion of serious possibility, which Levi cashes out in terms of subjective certainty (his "full belief"). A proposition is a serious possibility for *DM*, Levi maintains, just in case she not certain that it is false. With this understanding, the above conditions can be rewritten as follows (where d*A* says that *DM* will *decide on A*):

Levi's Analysis of Availability. DM sees *A* as available for choice during her deliberations only if[30] *DM*'s subjective probabilities are such that

> *Deliberation*. *DM* is certain she is deliberating.
>
> *Ability*$_E$. $\mathrm{P}(\neg \mathrm{d}A) < 1$ (*DM* is not certain she will not decide on *A*).
>
> *Efficacy*$_E$. $\mathrm{P}(A/\mathrm{d}A) = 1$ (*DM* is sure she will do *A* if she so decides).

It follows that *DM* cannot see *A* as an available option if (a) she is sure she will not perform it or (b) she is certain of a proposition *H* such that $\mathrm{P}(\neg dA/H) = 1$ or $\mathrm{P}(A/\mathrm{d}A \ \& \ H) < 1$.[31] So, *A* ceases to be even an option for *DM* when she becomes certain that she will do otherwise or when she becomes convinced of a proposition like *H* = "I will choose rationally and *A* is not a rational choice".

There are various flaws in this analysis. Let's start with Efficacy$_E$. Though likely true, it omits the most important part of the story. Even though $\mathrm{P}(A/\mathrm{d}A) = 1$ must hold for *DM* to see *A* as being fully under her control, this is a mere *symptom*. Genuine efficacious requires not only that *DM* be convinced that she will perform *A* if she so decides, but that she believe that her act will be a *causal consequence* of her decision. If *DM* does not see d*A* as the *direct (and total) cause* of *A*, then she will not see her decision to do *A* as (wholly) efficacious, and this is true no matter how high $\mathbf{P}(A/\mathrm{d}A)$ might be. To illustrate, consider a class of bogus decisions that I call a *pseudo*-Newcomb problems because they are so often confused with the real thing. In both real and *pseudo*-Newcomb problems *DM*'s action and a given desirable outcome are joint effects of a common cause that *DM cannot* control, but there is no direct causal link between the act and the outcome. In real Newcomb problems the act is a causal consequence of the agent's decision (which is itself an effect of the background state), while in pseudo-Newcomb problems the act and the decision are not casually connected.[32] Here is the picture:



In pseudo-Newcomb problems *DM* might be certain that she will do *A* if she so decides, but she will not see her decision as efficacious since she does not think it will cause her act.

To believe that a decision efficacious is, inescapably, to have a *causal* belief whose content outruns any purely evidential relationship that might hold between *A* and d*A*. Beliefs about the efficacy of one's decisions involve not only the evidential relationship $\mathbf{P}(A/\mathrm{d}A) = 1$, but the stronger causal one $\mathbf{P}(A\backslash \mathrm{d}A) = 1$. The causal connection is the one that counts as far as questions of agency are concerned. Far from being a "metaphysician's plaything",[33] causal probabilities are *essential* to understanding human agency. Unless we speak about *DM*'s causal beliefs we cannot even say what it *means* for her to see herself as having a choice about *A*. This alone suffices to undermine Levi's analysis, but there is more.

An even deeper flaw concerns Levi's standard of "serious possibility". One can distinguish *serious epistemic possibility* from *serious practical possibility*. A proposition *H* is a serious *epistemic* possibility for *DM* exactly if its truth is probabilistically consistent with all of her evidence. If we follow Levi in identifying *DM*'s evidence with her *corpus of certainties*, the set of propositions of whose truth she is subjectively certain, then we have the following definition:

H is a serious *epistemic* possibility for *DM* iff it is consistent with the laws of probability that she assign *H* a positive probability while assigning probability 1 to each proposition in her corpus of certainties.

In contrast, we may define practical possibility as follows:

H is a serious *practical* possibility for *DM* iff she is rationally required to factor the possibility of *H*'s truth into her decision making.

Practical *im*possibility appears in decision theory in the concept of a *null event*. An event *H* is said to be *null* if and only if altering the consequences that acts produce when *H* obtains makes no difference whatsoever to the agent's preferences. For example, if *DM* would rather attend a concert than a movie, and if the event that it snows in Muscat in May is null for her, then offering to pay her $1,000,000, or any other sum, if she goes to the movie and it snows in Muscat in May will not alter her preference: she will still hear the concert. When *H* is null *DM*'s unconditional preferences for acts correspond exactly with her *preferences conditional on* $\neg H$, which means that she can legitimately ignore the possibility of *H*'s truth when deciding what to do.

Though never put quite this way, it is a core tenet of expected utility theory that serious epistemic and practical possibility coincide for events expressible as disjunctions of *states of the world* (read *events over which DM has no control*). If *H* is such an event, then *H* is null for *DM* iff she is certain it is false. Levi proposes to extend this to cases in which *H* describes one of *DM*'s own acts. His thesis is that, insofar a she is rational, *DM* will only regard an action *A* as a serious practical possibility when it is also a serious epistemic possibility for her. Conversely, she will *not* regard *A* as a serious practical possibility, and will be able to legitimately ignore it in her decision making, whenever $\mathbf{P}(\mathrm{d}A)$ or $\mathbf{P}(A)$ is one or zero.

This sometimes makes sense. If *A* or *dA* is epistemically impossible because *DM* is certain of some *exogenous, uncontrollable* condition *H* that is incompatible with *A* or d*A*, then *A* really is a dead issue for her. To borrow an example from Levi, if *DM* suddenly realizes that she is about to suffer an asthma attack that will prevent her from playing the piano, then playing is practically impossible for her.[34] Levi thinks the same holds when a person becomes certain of facts *internal* to her deliberations. If, say, *DM* is sure she will maximize expected utility but discovers that A will not do this, then she no longer regards doing A even as an option. Or, if she is certain she will decide on some contrary act B for which $\mathbf{P}(B/dB) = 1$, then this too prevents her from seeing A even as an option.

Levi illustrates the latter case by imaging a manager who has been considering three applicants for a job, but who has already decided to hire one of the first two, and who regards his decision as efficacious. According to Levi,

> given [the manager's] decision to reject the third candidate and the efficaciousness of his choices, it is not epistemically or seriously possible that he choose the third candidate as far as he is concerned. Because hiring the third candidate is not a feasible option given [his] convictions, rationality does not require that he take that option into account in determining what to do.[35]

Levi's claim, then, is that *DM* cannot see A as available for choice if she is certain of *any* proposition H such that $\mathbf{P}(A/H) = 0$ or $\mathbf{P}(\mathrm{d}A/H) = 0$, and it does not matter whether H is some exogenous condition over which she has no control or a fact about her own decisions.

This overlooks an important difference between the asthma and manager cases. Piano playing really does cease to be a serious possibility for a person who learns that she will suffer an asthma attack, but not *merely* because she becomes certain that she will not play. Since she knows that there is nothing she can *do* to prevent an attack, becoming convinced that she will suffer it is, for her, becoming convinced of a proposition *whose truth-value she is certain she could not change even if she wanted to*. Thus, she both has evidence that convinces her that she will not perform the act, and she is certain that she can do nothing to alter or nullify this evidence. Things are quite different for the manager. While there are versions of the manager's problem in which *external* contingencies prevent him from changing his mind once he has decided against hiring the third candidate, this is *not* what Levi imagines.[36] Given that the manager's decision, and the certainty it engenders, is supposed to suffice *all by itself* to rule out hiring the third candidate as a serious practical possibility, we must suppose that his decision does not set into motion any chain of events that will prevent him from hiring the third candidate should he change his mind. In contrast to the asthma case, no *external obstacle* stands in the manager's way; the *only* thing preventing him from hiring the third candidate is his decision to do otherwise.

How can making a *revisable* decision turn what was a serious practical possibility into a practical impossibility? Levi seems to reason as follows: When the manager decides against hiring the third candidate (A) he becomes certain he has so decided and d$\neg A$ enters his corpus of certainties. Given Efficacy$_E$, this new evidence is probabilistically incompatible with A, so A loses its status as a serious epistemic possibility. Since there is no distinction between serious epistemic and practical possibility, the manager ceases to regard hiring the third candidate as an option. The manager is thus hemmed in by the beliefs his own free choices generate. He can no longer think A as an option because he has given himself conclusive evidence, namely d$\neg A$, which indicates that he will not do it.

This overlooks the fact that, unlike the asthma sufferer, *the manager controls what evidence he has concerning his own actions*. No external obstacle prevents him from changing his mind, and by doing so he can alter the constitution of his corpus of certainties. If he changes his mind and decides on A, then dA will replace d$\neg A$ in his corpus, and this both destroys his evidence for $\neg A$ and gives him evidence for A. Indeed, this evidence for A is conclusive so long as he takes his decisions to be causally efficacious. The point is that, *insofar as A and ¬A are concerned, the manager controls the contents of his corpus of certainties and so controls what it is reasonable for him to believe about* A and $\neg A$. Given this, the mere fact that A conflicts with his evidence cannot rule it out as a serious practical possibility. After all, it is practically possible for him to make a decision that will alter his evidence so as to make A epistemically possible. Since (as everyone will agree) A is practically possible if it is epistemically possible, it follows that *making A practically possible* is one of the manager's practically possible options. This means that A is practically possible *tout court* even though the manager is sure he will not change his mind and perform it.[37]

Levi might accuse me of begging the question here. I have been claiming that the manager can simultaneously be certain of $\neg A$ and still take himself to be in a position to change his mind and decide on A. If he is certain of $\neg A$ on the basis of a decision to do $\neg A$, Levi would argue, then he cannot see changing his mind as a practical possibility, and so cannot see himself as controlling his evidence about A. What prevents this? Here is Levi's answer:

> It may be objected that [the manager] can renege on his past decision. If reneging is an option for him and *if he is not certain he will not renege*, the point is well taken. But, given that [he] has chosen to reject the third candidate under the assumption of efficaciousness, he has ruled out reneging as a serious possibility. To be sure, [the manager] may *subsequently* change his mind and conclude that his initial decision is not efficacious after all. But as long as he fails to do so, he remains certain that he will not choose the rejected option. Consequently, in the context of his deliberation at the time, the rejected option is not a feasible option for him (emphasis added).[38]

In addition to making the misleading suggestion that changing one's mind is like "reneging" on a bargain, this passage asserts that agent who is certain of $\neg A$ can *only* change his mind and decide to do A by rejecting *Efficacy*. The claim seems to be that someone with subjective probabilities $\mathbf{P}(\neg A) = \mathbf{P}(\mathrm{d}\neg A) = \mathbf{P}(\neg A/\mathrm{d}\neg A) = 1$ cannot move to a new credal state in which $\mathbf{P}_{\mathrm{NEW}}(A) > 0$ unless $\mathbf{P}_{\mathrm{NEW}}(\neg A/\mathrm{d}\neg A) < 1$. Presumably, the idea is that after changing his mind the person will know that his *prior* decision to do $\neg A$ was not efficacious, and somehow this will force him to adopt *new* beliefs in which he is no longer certain of $\neg A$ conditional on $\mathrm{d}\neg A$.

Once again, Levi's errors by ignoring the role of causal beliefs in decision-making. As we have seen, the manager will only see his decision to perform $\neg A$ as efficacious if he is convinced that it will *cause* $\neg A$. Given this, he might move to a new credal state in which $\mathbf{P}(A) > 0$ in either of *two* ways. He might acquire new evidence that undermines his views about the causal powers of his decision, as the asthma sufferer does. This will indeed undermine Efficacy. On the other hand, he might simply delete $\mathrm{d}\neg A$ from his corpus of certainties *while retaining his view that deciding on* $\neg A$ *will cause* $\neg A$. Only the second sort of belief revision counts as a "change of mind" in the sense relevant to this discussion. These are changes in an agent's opinions about what he will *decide*, that are not accompanied by any ancillary changes in his views about what his decisions will *cause*. They are, more precisely, belief revisions in which the agent moves from a credal state in which $\mathrm{d}\neg A$ and $\neg A$ are elements of his corpus of certainties and $\mathbf{P}(A\backslash \mathrm{d}A) = \mathbf{P}(\neg A\backslash \mathrm{d}\neg A) = 1$ hold to a new credal state in which $\mathrm{d}\neg A$ and $\neg A$ are no longer in his corpus and

$\mathbf{P}_{\text{NEW}}(A\backslash \mathrm{d}A) = \mathbf{P}_{\text{NEW}}(\neg A\backslash \mathrm{d}\neg A) = 1$. The possibility of changing one's mind in this way does not conflict with Efficacy in any way, so long as Efficacy is rightly understood as a principle about what the agent's decisions will *cause*. Hence, Levi has given no reason to think that practical and epistemic possibility will coincide for an agent's own acts.

The point generalizes. Even if *DM* is certain of *H* when $\mathbf{P}(A/H) = 0$ or $\mathbf{P}(\mathrm{d}A/H) = 0$, *A* can still be a serious practical possibility for her as long as she takes herself to be in a position to decide what evidence she has for *H*. Suppose *H* says that *DM* will not choose or perform any irrational act. If *DM* is certain both of *H* and of *A*'s irrationality, then $\mathbf{P}(A/H) = \mathbf{P}(\mathrm{d}A/H) = 0$. Even so, it is clear that *DM* is ultimately in control of her evidence regarding *H*. She knows that by deciding on ¬*A* she will give herself conclusive evidence for thinking that she will act rationally (as she already believes). She also knows that by deciding on *A* she will give herself conclusive evidence for thinking she will act irrationally. And, most important, *the fact that she now believes H does nothing to alter either of these things*. Insofar as she sees herself as free, *DM* sees herself as being able to *decide* whether or not *H* is part of her evidence, and so to control whether or not d*A* or *A* conflict with her evidence. Since such a conflict is, on Levi's view, the *only* thing that prevents *DM* from seeing *A* as an option, her ability to control what she believes about *H* makes doing *A* practically possible for her even when she is certain she will do something else. Levi's arguments for Premise-2 thus fail.

3.3. *A deeper worry?*

Perhaps there is a deeper worry here though. I am portraying the agent who changes her mind as altering her *beliefs* about what she will decide *on the basis of no evidence whatever*. She goes from being certain that she has decided on ¬*A* to being certain that she has decided on *A* without *learning* anything. Can this sort of belief change be rational? By letting agents assign subjective probabilities to their own acts it seems that we are also letting them believe whatever they want about them. This means that act probabilities must be radically unlike other probabilities in that they seem not to be at all constrained by the believer's evidence. So, for example, an agent with a vast store of independent evidence for the conclusion that she will not choose irrationally is portrayed as being able to simply contravene this evidence by an act of her will. This, I suspect, gets us to what is really bothering people about act probabilities.

The issue can be brought into sharper focus by reconsidering the Twin's Dilemma with Symmetry and Uncertainty. Recall that the only probability that Row can assign to her acts in these circumstances is $\mathbf{P}(R) = 1/2$. If we imagine that Row has compelling evidence for Symmetry and Uncertainty, then she seems to be confined in a kind of "epistemic straight jacket" that prevents her from modifying her opinions about what she will do. The evidence she has *prior to* deliberating justifies $\mathbf{P}(C/R) = \mathbf{P}(R/C)$, $\mathbf{P}(C/\neg R) = \mathbf{P}(R/\neg C)$ and $\mathbf{P}(C) = 1/2$, and so *forces* her to set $\mathbf{P}(R) = 1/2$. The only *new* evidence she acquires *during* deliberations concerns which of her acts best serves her interests, i.e., she only learns that she *prefers* $\neg R$ to R. How could learning *this* justify her in violating Symmetry and Uncertainty or in raising her confidence in $\neg R$ to one? There seems to be no way out: Row must remain in her state of indecision, $\mathbf{P}(R) = 1/2$, because altering this belief without acquiring any new evidence appears be irrational. Her situation is worse than that of Buridan's Ass – at least the Ass was caught between *equally desirable* options.

To see the way out, note first that when *DM* sees herself as a free agent in the matter of *A*, Efficacy ensures that all of her evidence about *A* comes by way of *evidence about her decisions*. Her justification for claiming that she will do *A* will always have the form: "here is such-and-such evidence that I will decide on *A*, and (via Efficacy) deciding on it will cause me to do it." This might seem to push the problem back from beliefs about acts to beliefs about decisions, but this is not so. An agent's beliefs about her own decisions have a property that most other beliefs lack: under the right conditions they are *self-fulfilling*, so that if the agent has them then they are true. Understanding this is one of the keys to understanding human agency and freedom.

Self-fulfilling beliefs have long been discussed in connection with doxastic voluntarism. Much of this literature is irrelevant here, as it is a mistake to think of people *choosing* their decisions or their beliefs about their decisions – one chooses acts, not choices. For us, the crucial fact about self-fulfilling beliefs is that they *generate their own evidence*. This point has been made forcefully by David Velleman whose "Epistemic Freedom" is essential reading on this topic.[39] Following Elizabeth Anscombe,[40] Velleman notes that a person can be warranted in adopting a self-fulfilling belief when she lacks evidence for it, or even when she has evidence against it, because (a) she might know the belief is self-fulfilling, and so (b) recognize that she will have evidence for the belief *once she adopts it*. This evidence will consist in knowing that she holds a belief that ensures its own truth. In a way, Descartes saw the point first: no matter how much evidence I might have against my own existence I can always justifiably believe that I exist because the very having of this belief is conclusive evidence for its truth. The same happens with all self-fulfilling beliefs: the fact that one holds them is evidence of their truth (albeit not always conclusive evidence). According to Velleman, this means that the believer has a kind of "epistemic freedom" with respect to self-fulfilling beliefs that she lacks for her other opinions; she can *justifiably* believe whatever she wants about them. If she is sure that believing H will make H true and that believing $\neg H$ will make $\neg H$ true then, *no matter what other evidence she might posses*, she is at liberty to believe either H or $\neg H$ because she knows that *whatever* opinion she adopts will be warranted by the evidence she will acquire *as a result of adopting it*. More generally, any increase or decrease in her confidence in H provides her with evidence in favor of that increase or decrease – the stronger a self-fulfilling belief is, the more evidence one has in its favor.

Velleman takes pains to point out that epistemic freedom should *not* be confused with metaphysical freedom. A person can see herself as epistemically free *in re H* even when she knows that her beliefs about H are determined by facts beyond her control. Being epistemically free is not a matter of choosing what to believe, or even of being able to believe otherwise. It involves being in a position to disregard evidence concerning H because one knows that it will be made moot by the fact of one's belief. Velleman holds, as I do, that agents are epistemically free with respect to their own decisions and intentions, and he hopes to use this to explain the "feeling of freedom" that people have when they act. Whether or not we go along with him on this last point, the idea that agents are epistemically free regarding their own decisions is important and entirely correct.

Beliefs can be self-fulfilling for a variety of reasons. In the most commonly discussed cases, the self-fulfilling belief contributes causally to its own truth (as in William James's famous "crevasse jumper" example). Beliefs about decisions are self-fulfilling for a different reason. They are akin to *performatives*. In much the same way in which *saying* that one promises to give a man a horse can, under the right conditions, *be* a promise to give a man a horse, so *believing* that one has decided to do *A* can, under the right conditions, *be* a decision to do *A*. The "right" conditions are just those of deliberation. If *DM* becomes certain of d*A* during the course of deliberations about whether to do *A*, then she has decided on *A*. This *constitutive* relationship between decisions and beliefs about them ensures that any belief of the form "I decide to do *A*" adopted during deliberation will be self-fulfilling, and that the deliberator is epistemically free with respect to it. This explains how a decision-maker can conform her beliefs about what she will decide to her preferences for acts without engaging in wishful thinking. During the course of her deliberations *DM*'s confidence in "I decide to do *A*" will wax or wane *in response to information about A's desirability relative to her other options* (e.g., information about expected utilities). If *A* and ¬*A* seem equally desirable at some point in the process, then she will be equally confident of d*A* and d¬*A* at that time. If further deliberation leads her to see *A* as the better option, then her confidence in d*A* will increase as her confidence in d¬*A* decreases. These deliberations will ordinarily cease when *DM* is certain of either d*A* or d¬*A*, at which point she will have made her decision about whether or not to perform *A by making up her mind what to believe about* d*A*.[41] Though this process would be nothing more than an exercise in wishful thinking if *DM*'s beliefs about d*A* and d¬*A* were not self-fulfilling, the fact that they are ensures that her subjective probability for each proposition increases or decreases in proportion to the evidence she has in its favor. This explains how *DM*'s beliefs about what she will decide can be both responsive to her preferences and warranted by her evidence at each moment of her deliberations.

If this is the right picture of rational deliberation, as I think it is, then it is no mystery at all how a rational agent can assign intermediate probabilities to acts. When Efficacy holds these beliefs fix a subjective probabilities for A via the rule $\mathbf{P}(A) = \mathbf{P}(\mathrm{d}A)$ (When Efficacy fails $\mathbf{P}(A) = \mathbf{P}(\mathrm{d}A)\mathbf{P}(A/\mathrm{d}A) + \mathbf{P}(\mathrm{d}A)\mathbf{P}(\neg A/\neg \mathrm{d}A)$). Moreover, the act probabilities are well justified in light of *DM*'s evidence because (a) her degree of belief in d*A* is warranted in virtue of being self-fulfilling, and (b) since she is sure d*A* will *cause A*, this self-fulfilling belief provides her with conclusive evidence for *A*. *DM* is never "hemmed in" by her evidence about *A*. She is free to believe whatever she wants about d*A* and, because she is certain that her decision will be efficacious, this "epistemic freedom" carries over to her belief about *A*.

In Twin's Dilemma with Symmetry and Uncertainty, for example, as soon as Row begins to see ¬*R* as the better option her subjective probability for d¬*R* will rise. As it does, Row acquires evidence for ¬*R*, and her subjective probability for ¬*R* increases as well. Of course, this means that Symmetry or Uncertainty have to go. Which one goes depends entirely the character of Row's prior evidence. If she starts out with a great deal of evidence for Similarity, but her only justification for Uncertainty is that she is uncertain, then Uncertainty will go. If she starts out with strong, independent reasons for thinking that *C* and ¬*C* are equally likely, and Symmetry only holds because she is undecided about what to do then Symmetry will go. When the evidence is mixed, both may be jettisoned. No matter what happens, as long as the agent sees herself as free in the matter of *A* the evidence that underlies Symmetry and Uncertainty will not constrain her beliefs about her own decision or actions in any way. She is free to believe what she wants.

This way of looking at matters also lets us assuage Worry-3 above. As Wolfgang Spohn has long argued, there is no reason allow act probabilities in decision theory if we cannot find anything useful for them to do.[42] Given that they play no role in the *evaluation* or *justification* of acts, it would seem that there is nothing useful for them to do. Why not abolish them? We now know the answer. Act probabilities are a kind of epiphenomena in decision theory. Though they do no real explanatory work, they are tied to things that do. We need act probabilities because (i) we need unconditional subjective probabilities for *decisions about acts* to *causally* explain action (though not to rationalize it), and (ii) we need Efficacy to explain what it is for an agent to regard acts as being under her control. Efficacy requires that $\mathbf{P}(A\backslash \mathrm{d}A) = \mathbf{P}(A\backslash \mathrm{d}\neg A) = 1$, and so $\mathbf{P}(A/\mathrm{d}A) = \mathbf{P}(A/\mathrm{d}\neg A) = 1$. One cannot have these latter conditional probabilities and unconditional probabilities for dA and d$\neg A$ without also having unconditional probabilities for A and $\neg A$. Act probabilities are not only coherent, they are *compulsory* if we are to adequately explain rational agency. We cannot outlaw them without jettisoning other subjective probabilities that are essential ingredients in the causal processes that result in deliberate actions. When it comes to beliefs about one own actions, deliberation does not "crowd out" prediction; it mandates it!

NOTES

[1] "Review Essay: *The Foundations of Causal Decision Theory*", *Journal of Philosophy* (2000) 97, 394.

[2] Jeffrey's views are developed in *The Logic of Decision*, 2nd revised edition (1983), Chicago: University of Chicago Press.

[3] Savage, L. (1972): *The Foundations of Statistics*, 2nd revised edition, New York: Dover.

[4] Levi, "Review Essay", p. 387.

[5] I am assuming an idealized agent who assigns sharp numerical probabilities to states and sharp utilities to outcomes. Neither Levi nor I think decision theory applies only to such agents, but the issues between us arise most clearly for them.

[6] This point is not new; it has been part of the doctrine of causal decision theory since its inception. It is made by, among others: Gibbard, A. and Harper, W. (1978): "Counterfactuals and Two Kinds of Expected Utility", in C. Hooker, J. Leach and E. McClennen (eds.), *Foundations and Applications of Decision Theory* (pp. 125–162), Dordrecht: Reidel; and Skyrms, B. (1980): *Causal Necessity*, New Haven: Yale University Press; and Armendt, B. (1986): "A Foundation for Causal Decision Theory", *Topoi* 1, 3–19.

[7] Pearl, J. (2000): *Causality* (p. 108), Cambridge: Cambridge University Press.

[8] See my *Foundations of Causal Decision Theory* (pp. 161–180), for a discussion of the options.

[9] Levi, "Review Essay", p. 393.

[10] Levi, "Review Essay", p. 393.

[11] Levi, "Review Essay", p. 392. It would be interesting to know which causal decision theorists Levi has in mind here. There *is* a sloppy line in my recent book that might suggest that I commit this error. In explaining Newcomb's Problem I write (with changes in notation), "*DM* regards *C* as very unlikely given *R* and

very unlikely given ¬*R*, that is, her subjective probability for the proposition (*R* & *C*) ∨ (¬*R* & ¬*C*) is nearly 1" (*Foundations*, p. 145). I wish I had the "that is" back. The suggestion that the second claim is equivalent to the first is a mistake for the reasons Levi gives; the first entails the second, but not conversely. That said, my treatment of Newcomb problems consistently uses the *first* reading, so the mistake does not infect the discussion. In particular, *at no point do I infer that* **P**(*C*/*R*) *and* **P**(¬*C*/¬*R*) *are both near 1 from the premise that* **P**(*R* & *C*) + **P**(¬*R* & ¬*C*) *is near one*.

[12] Levi, "Review Essay", p. 392.

[13] Levi, "Review Essay", p. 393. Contrary to what he claims, nearly all decision theorists will agree that it is *not* legitimate to employ dominance reasoning to *any* decision problem in which an agent's degrees of belief are indeterminate. It would be a mistake, e.g., to do this in the "flu shot" case. Like SAVAGE, Dominance contains an implicit restriction: it may only be used when states are causally independent of acts from the decision-maker's perspective. Given this, any appeal to Dominance in the Twin's Dilemma relies implicitly on the fundamental insight of causal decision theory.

[14] Levi, "Review Essay", p. 393.

[15] This is one of the main theses defended in my *Foundations*. For discussions of the Jeffrey/Bolker theory see *The Logic of Decision*, Chapter 9; and Bolker, E. (1966): "Functions Resembling Quotients of Measures", *Transactions of the American Mathematical Society* 124, 293–312.

[16] Levi, "Review Essay", pp. 394–395.

[17] Levi, "Review Essay", pp. 395–396.

[18] Gilboa, I. (1999): "Can Free Choice be Known", in C. Bicchieri, R. Jeffrey and B. Skyrms (eds.), *The Logic of Strategy* (pp. 163–174), Oxford: Oxford University Press.

[19] Levi's most explicit pronouncements on these matters are found in "Rationality, Prediction, and Autonomous Choice" and "Consequentialism and Sequential Choice" both of which appear in his collected papers *The Covenant of Reason* (1977) (pp. 19–39 and 70–101 respectively), Cambridge: Cambridge University Press.

[20] "Consequentialism and Sequential Choice", p. 77.

[21] Ramsey, F.P. (1931): "Truth and Probability", in R. Braithwaite (ed.), *The Foundations of Mathematics and Other Logical Essays* (pp. 156–98), London: Kegan Paul. de Finetti, B. (1964): "Foresight: Its Logical Laws, Its Subjective Sources", in H. Kyburg and H. Smokler (eds.), *Studies in Subjective Probability* (pp. 93–158). New York: John Wiley.

[22] See de Finetti (1974): *Theory of Probability* 1, 87–91, New York: John Wiley and Sons. I have not formulated things precisely as de Finetti does, but the differences are immaterial for present purposes.

[23] This is a reconstruction of the argument that Levi offers in footnote 5 of "Consequentialism and Sequential Choice", pp. 76–77. An abbreviated version of the argument appears in footnote 8 of the "Review Essay", p. 395.

[24] One might wonder whether it is even makes sense for *DM* to be certain about

her decision regarding *A* during her deliberations. We will discuss this below, but for the moment let us assume it is possible for purposes of argument. If it is not, then Strategy-2 cannot be rationally pursued at all, so we need not worry about it.

[25] This is clearest the first paragraph of footnote 5 of "Consequentialism and Sequential Choice", p. 76–77.

[26] This is an oversimplification. People do not have unfettered epistemic access to their basic desires or beliefs. Deliberation really involves both a process of self-discovery, in which one discovers or creates one's basic desires and becomes aware of one's beliefs, and a process of act evaluation, in which one uses this information to decide which act to perform.

[27] Levi, "Review Essay", p. 394.

[28] See *The Enterprise of Knowledge* (1980) (pp. 2–5 and 222–223), Cambridge MA: MIT Press.

[29] Levi's most explicit pronouncements on these matters are found in "Rationality, Prediction, and Autonomous Choice", pp. 30–36, and "Consequentialism and Sequential Choice", pp. 77–78.

[30] Levi sometimes seems to suggest that these conditions are sufficient as well. He writes that "as far as decision theory is concerned the epistemic condition *DM* satisfies when we say the truth of act-descriptions is under *DM*'s control is this: The 'expansion' $\underline{K}$ + dA of *DM*'s corpus [of certainties] $\underline{K}$ by adding the information that *DM* chooses *A* contains *A* ... entails the truth of *A* while $\underline{K}$ does not ... This characterization of control of truth-values proceeds along entirely epistemic lines once we are given a list of option descriptions. Causality does not enter into the picture". "Consequentialism and Sequential Choice", p. 75, with minor changes in notation. While I am not sure whether Levi believes in the sufficiency of these principles, I am sure that he is mistaken if he does. There is no difficulty when *DM* is sure that deciding on *A* would cause her to do *A*, but problems to arise for other cases. For example, *DM* might think her ability to carry out her decision is contingent upon some causally independent factor that is positively correlated with her choice. Imagine a cocky surgical resident who knows that the surgeon overseeing her work will not let her remove the patient's gall bladder unless it is medically necessary, but who is supremely confident in her judgments about what is medically necessary. Even as she is figuring out what to do, and is thus not certain of either her decision or her act, she might be certain that she will remove the gall bladder iff she decides to do so. So, all three of the evidential conditions will be satisfied, but the resident clearly does not see the matter as being under her control; she knows that the surgeon's decision will settle the issue.

[31] Though Levi tends to include (b) as an independent clause, it is redundant. If *DM* is certain of *H*, so that $\mathbf{P}(H) = 1$, then $\mathbf{P}(\neg \mathrm{d}A/H) = 1$ iff $\mathbf{P}(\neg \mathrm{d}A) = 1$ and $\mathbf{P}(A/\neg \mathrm{d}A \,\&\, H) < 1$ iff $\mathbf{P}(A/\neg \mathrm{d}A) < 1$.

[32] One type of pseudo-Newcomb problem involves compulsive behavior, as when a brain event causes a person both to want to sneeze and to sneeze by independent mechanisms. Cases of "preestablished" harmony between acts and decisions can also serve as examples.

[33] Levi, "Review Essay", p. 292.
[34] "Rationality, Prediction, and Autonomous Choice", p. 27.
[35] "Rationality, Prediction, and Autonomous Choice", p. 29.
[36] He explicitly states that the manager "may subsequently change his mind". See "Rationality, Prediction, and Autonomous Choice", p. 29.
[37] Here I assume that practical possibility has at least the structure of an S4 modal logic, so that $\Diamond\Diamond A \rightarrow \Diamond A$.
[38] "Rationality, Prediction, and Autonomous Choice", p. 29.
[39] Velleman, J.D. (1989): "Epistemic Freedom", *Pacific Philosophical Quarterly* 70, 73–97.
[40] Anscombe, E. (1963): *Intention*, 2nd edition (pp. 1–4), Ithaca: Cornell University Press.
[41] I cannot profess to know how this process works in detail, but the formal models explored in Skyrms, B. (1990): *The Dynamics of Rational Deliberation*, Cambridge, MA: Harvard University Press, seem to be on the right track. They fit the picture of deliberation described here perfectly. On Skyrms's models *DM*'s deliberations can end in an equilibrium in which neither A or $\neg A$ is definitely preferred, and so neither dA or d$\neg A$ has a probability of one. In these cases *DM* must simply make an unreasoned "pick" of which act she will do. Nevertheless, she will be justified in believing whatever act she happens to pick.
[42] Spohn, W. (1977): "Where Luce and Krantz Do Really Generalize Savage's Decision Model", *Erkenntnis* 11, 115.

Department of Philosophy
University of Michigan
435 South State Street
Ann Arbor, MI 48109-1003
USA
E-mail: jjoyce@umich.edu